Informatica® Big Data Suite
10.2.2

Installation and Configuration Guide

Informatica Big Data Suite Installation and Configuration Guide
10.2.2

Publication Date: 2019-06-28

Table of Contents

Preface

The *Informatica Installation and Configuration Guide* is written for the system administrator who is responsible for installing the Informatica product. This guide assumes you have knowledge of operating systems, relational database concepts, and the database engines, flat files, or mainframe systems in your environment. This guide also assumes you are familiar with the interface requirements for your supporting applications.

Informatica Resources

Informatica provides you with a range of product resources through the Informatica Network and other online portals. Use the resources to get the most from your Informatica products and solutions and to learn from other Informatica users and subject matter experts.

Informatica Network

The Informatica Network is the gateway to many resources, including the Informatica Knowledge Base and Informatica Global Customer Support. To enter the Informatica Network, visit https://network.informatica.com.

As an Informatica Network member, you have the following options:

- Search the Knowledge Base for product resources.
- View product availability information.
- Create and review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

Informatica Knowledge Base

Use the Informatica Knowledge Base to find product resources such as how-to articles, best practices, video tutorials, and answers to frequently asked questions.

To search the Knowledge Base, visit https://search.informatica.com. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at KB_Feedback@informatica.com.

Informatica Documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases. To explore the Documentation Portal, visit https://docs.informatica.com.

Informatica maintains documentation for many products on the Informatica Knowledge Base in addition to the Documentation Portal. If you cannot find documentation for your product or product version on the Documentation Portal, search the Knowledge Base at https://search.informatica.com.

If you have questions, comments, or ideas about the product documentation, contact the Informatica Documentation team at infa_documentation@informatica.com.

Informatica Product Availability Matrices

Product Availability Matrices (PAMs) indicate the versions of the operating systems, databases, and types of data sources and targets that a product release supports. You can browse the Informatica PAMs at https://network.informatica.com/community/informatica-network/product-availability-matrices.

Informatica Velocity

Informatica Velocity is a collection of tips and best practices developed by Informatica Professional Services and based on real-world experiences from hundreds of data management projects. Informatica Velocity represents the collective knowledge of Informatica consultants who work with organizations around the world to plan, develop, deploy, and maintain successful data management solutions.

You can find Informatica Velocity resources at http://velocity.informatica.com. If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Marketplace

The Informatica Marketplace is a forum where you can find solutions that extend and enhance your Informatica implementations. Leverage any of the hundreds of solutions from Informatica developers and partners on the Marketplace to improve your productivity and speed up time to implementation on your projects. You can find the Informatica Marketplace at https://marketplace.informatica.com.

Informatica Global Customer Support

You can contact a Global Support Center by telephone or through the Informatica Network.

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:
https://www.informatica.com/services-and-training/customer-success-services/contact-us.html.

To find online support resources on the Informatica Network, visit https://network.informatica.com and select the eSupport option.

Part I: Installation Getting Started

This part contains the following chapter:

CHAPTER 1

Installation Getting Started

This chapter includes the following topics:

Checklist to Getting Started

This chapter contains high-level concepts and planning information related to installation. Use this checklist to track the completion of preliminary tasks.

❑ Understand high-level concepts:

- The installer description and process.
- Informatica domain terminology and components.

❑ Start high-level planning:

- Installation options. Review the installation options to know the product and options for installation.
- Installation components. Review the description of the installation components and the planning notes.

Installation Overview

Welcome to the Informatica Big Data Suite installation to install Informatica domain services and clients. The Informatica domain services consist of core services to support the domain and of application services to support big data processing in the native and Hadoop environments. The Informatica clients consist of thick client applications and web client applications.

Run the Informatica Big Data Suite installer to install domain components and application services that support the following big data products:

- Big Data Management
- Big Data Parser

- Big Data Quality
- Big Data Streaming
- Enterprise Data Catalog
- Enterprise Data Lake

When you install the Informatica domain services, you are prompted to create a domain or to join a domain. The domain is a collection of nodes that represent the machines on which the application services run. The first time you run the installer, you must create the domain. If you install on a single machine, you create the Informatica domain and a gateway node on the machine. If you install on multiple machines, you create an Informatica domain and a gateway node during the first installation. During the installation on the additional machines, you create gateway or worker nodes that you join to the domain.

When you run the installer, it installs files for services. You can optionally create application services during the installation process, or you can manually create application services when the installation completes.

Installation Process

The installation of the Informatica domain services and Informatica clients consists of multiple phases.

The installation process varies based on the products that you install. Consider the following high-level tasks of the installation process:

Perform pre-installation tasks.

1. Plan the Informatica installation. Determine the big data products that you want to run in your environment. If you are creating a domain, consider the number of nodes in the domain, the application services that will run on each node, the system requirements, and the type of user authentication that the domain will use.
2. Prepare the databases required for repositories, warehouses, and catalogs. Verify the database requirements and set up the databases.
3. Set up the machines to meet Linux requirements to ensure that you can successfully install and run the Informatica services.
4. If you are installing Enterprise Data Catalog, determine whether you want to install on a cluster embedded with the domain or use one that is external to the domain.
5. Determine security requirements for the domain, services, and databases.

Run the installer to install the big data products required for your environment.

When you install version 10.2.2, you can choose the following types of installation:

- Informatica domain services. Installs the domain services and the applications service binaries to support Big Data Management, Big Data Parser, Big Data Quality, and Big Data Streaming.
- Enterprise Data Catalog. Installs the application service binaries to support Enterprise Data Catalog. You can also choose to install the Informatica domain services with Enterprise Data Catalog.
- Enterprise Data Lake. Installs Enterprise Data Lake on an existing domain. You can also choose to install the service binaries to support Enterprise Data Catalog with Enterprise Data Lake.
- All products. Install the domain and all big data products.

You can also install Data Transformation Engine for Big Data Parser.

Complete the configuration.

1. Verify code page compatibility.
2. Configure environment variables.
3. Complete tasks required by the type of user authentication used by the domain.
4. Optionally, configure secure communication for the domain.
5. Create and configure application services.
6. Configure connections required by the application services.
7. Create the users and connections required by the application services.
8. Integrate the domain with the Hadoop environment.

Install the Developer tool.

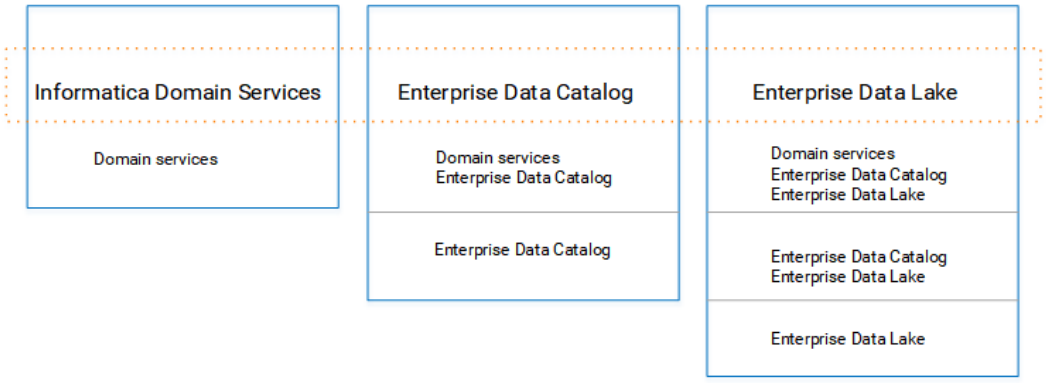
1. Verify the installation and third-party software requirements for the clients.
2. Use the client installer to install on Windows machines.
3. Configure required environment variables, and optionally install additional languages.

Plan the Installation Option

When you run the Big Data Suite installer, you have options to install Informatica domain services, Enterprise Data Catalog, or Enterprise Data Lake. Based on your initial selection, you can also choose to install parent products.

Before you begin the planning and preparation for install, you must determine the type of installation that you want to run.

The following image shows the products that you can install based on the installation options:



Consider the different options available when you run the installer:

Informatica domain services

When you install Informatica domain services, you install the core services that support the domain and the binaries for application services that support the following big data products:

- Big Data Management
- Big Data Parser
- Big Data Quality

- Big Data Streaming

When you install Informatica domain services, you can choose to create a domain or join a domain.

Enterprise Data Catalog

You can install Informatica services and Enterprise Data Catalog, or you can install Enterprise Data Catalog on a node that is running Informatica services.

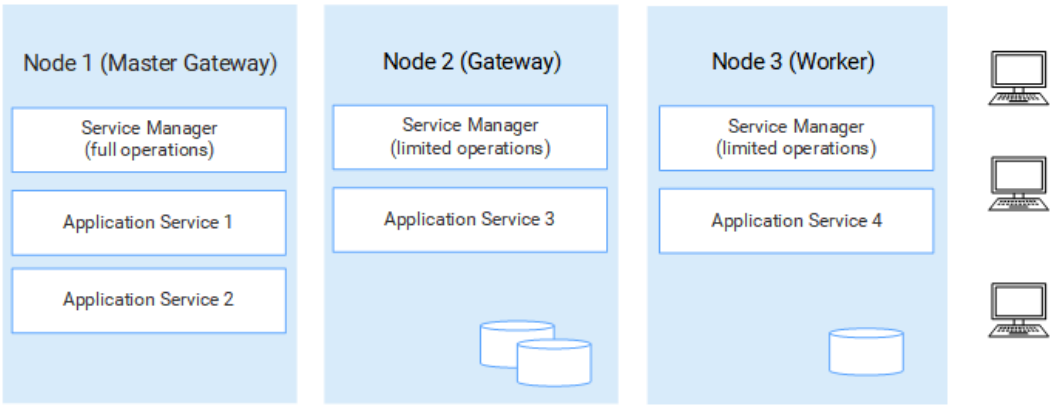
Enterprise Data Lake

You can perform a full installation of Informatica services, Enterprise Data Catalog, and Enterprise Data Lake. If Enterprise Data Catalog is not installed, you can install the binaries for Enterprise Data Catalog and Enterprise Data Lake. If the Informatica services and Enterprise Data Catalog are installed, you can install Enterprise Data Lake.

Plan the Installation Components

An Informatica domain is a collection of nodes and services. A node is the logical representation of a machine in a domain. Services include the Service Manager that manages all domain operations and a set of application services that represent server-based functionality. The domain and some services require databases to write metadata and run-time results.

The following image shows a high-level architecture of a domain on multiple nodes:



Nodes

The first time that you install the domain services, you create the Informatica domain and a gateway node. When you install the domain services on other machines, you create additional nodes that you join to the domain.

The domain has the following types of nodes:

- Gateway node. A gateway node is any node that you configure to serve as a gateway for the domain. A gateway node can run application services and it can serve as a master gateway node. The master gateway node is the entry point to the domain. You can configure more than one node as a gateway node, but only gateway node acts as the master gateway node at any given time.
- Worker node. A worker node is any node that you do not configure to serve as a gateway for the domain. A worker node can run application services, but it cannot serve as a gateway.

When you plan the installation: You need to plan the number and type of nodes that you need based on your service and processing requirements. If you have high availability, you will want to create more than one gateway node for fail-over functionality.

Service Manager

The Service Manager is a service that manages all domain operations. The Service Manager runs on each node in the domain and performs domain functions, such as authentication, logging and application service management. The Service Manager on a gateway node performs more tasks than the Service Manager on a worker node.

When you plan the installation: Note that the Service Manager functionality is associated with the type of node.

Application Services

Application services represent server-based functionality. An application service might be required or optional, and it might require access to a database.

When you run the installer, you can choose to create some services. After you complete the installation, you create other application services based on the license key generated for your organization.

When you plan the installation: When you plan the application services, you must account for the associated services that connect to the application service. You also must plan the relational databases that are required to create the application service.

Databases

Some application services require databases to store metadata and to write run-time results. You need to create databases for the application services in the domain.

When you plan the installation: You need to create databases and database users required by application services.

Hadoop Environment

The big data products process different types of jobs on the Hadoop cluster.

The types of jobs that you run depend on your product licenses and business requirements. When you install Enterprise Data Catalog, you can choose to use an external cluster or to create an embedded cluster. When you create an embedded cluster, the installer creates the cluster on the node where the Informatica services run.

You can use an embedded cluster if you do not also run Big Data Management or Enterprise Data Lake in the same domain.

When you plan the installation: Plan the type of cluster you want to run for Enterprise Data Catalog.

User Authentication

When you run the installer, you can choose the authentication to use for the domain.

The Informatica domain can use the following types of authentication to authenticate uses in the domain:

- Native. Native user accounts are stored in the domain and can only be used within the domain. Native authentication is default.
- LDAP. LDAP user accounts are stored in an LDAP directory service and are shared by applications within the enterprise. You can configure LDAP authentication after you run the installer.
- Kerberos. Kerberos user accounts are stored in an LDAP directory service and are shared by applications within the enterprise. If you enable Kerberos authentication during installation, you must configure the Informatica domain to work with the Kerberos Key Distribution Center (KDC).
- SAML. You can configure Security Assertion Markup Language (SAML) authentication for the Administrator tool, the Analyst tool, and the Monitoring tool. You can configure SAML authentication after you run the installer.

When you plan the installation: You need to plan the type of authentication that you want to use in the domain. If you want the installer to configure Kerberos authentication, you must prepare the network prior to installation. You can also configure Kerberos after installation. Note that you cannot configure both SAML and Kerberos authentication.

Secure Data Storage

Informatica encrypts sensitive data before it stores the data in the Informatica repositories. Informatica uses a keyword to create an encryption key with which to encrypt sensitive data.

When you create a domain, you must specify a keyword for the installer to use to generate the encryption key for the domain. Based on the keyword, the installer generates an encryption key file named siteKey and stores it in a default directory or the directory you specify. All nodes in a domain must use the same encryption key. You must specify a keyword even if you do not enable secure communication for the domain or use Kerberos authentication.

Important: Secure the domain name, the keyword, and the encryption key file location. This information is required when you change the encryption key or move a repository to another domain.

When you plan the installation: Determine if you want to use a custom siteKey or if you want the installer to generate it.

Domain Security

When you create a domain, you can enable options to configure security in the domain.

You can configure secure communication for the following domain components:

- Administrator tool. Configure a secure HTTPS connection for the Administrator tool. During installation, you can provide the keystore file to use for the HTTPS connection.
- Service Manager. Configure a secure connection between the Service Manager and other domain services. During installation, you can provide keystore and truststore files containing SSL certificates that you want to use.
- Domain configuration repository. You can secure the domain configuration repository with SSL protocol. During installation, you can provide the truststore file containing the SSL certificate that you want to use.

When you plan the installation: Determine the level of security that you want to configure for the domain components. If you decide to configure security for the domain, you must know the location and password for the keystore and truststore files. If you decide to use Kerberos authentication for the Informatica domain,

you must work with the Kerberos administrator to set up the user and service principals required by the domain.

Informatica Client Tools

You use Informatica clients to access underlying Informatica functionality in the domain. The clients make requests to the Service Manager and to application services.

The Informatica clients consist of thick client applications and thin or web client applications that you use to access services and repositories in the domain.

The following table describes the clients associated with big data products:

Informatica Client	Description
Informatica Developer (the Developer tool)	A thick client application to create and run data objects, mappings, profiles, and workflows. Used by Big Data Management, Big Data Parser, Big Data Quality, and Big Data Streaming.
Informatica Administrator (the Administrator tool)	A web application to manage the domain and application services. Used by Big Data Management, Big Data Parser, Big Data Quality, Big Data Streaming, Enterprise Data Catalog, and Enterprise Data Lake.
Informatica Analyst (the Analyst tool)	A web application to analyze, cleanse, integrate, and standardize data in an enterprise. Used by Big Data Management, Big Data Quality, Big Data Streaming, and Enterprise Data Catalog.
Informatica Mass Ingestion (the Mass Ingestion tool)	A web application to create, deploy, run, and monitor mass ingestion specifications. Used by Big Data Management, Big Data Parser, Big Data Quality, and Big Data Streaming.
Enterprise Data Lake application	A web client application to search, discover, and prepare big data sets. Used by Enterprise Data Lake.
Enterprise Data Catalog application	A web application to analyze and understand large volumes of catalog metadata in the enterprise. Used by Enterprise Data Catalog and Enterprise Data Lake.
Enterprise Data Catalog Administrator (the Catalog Administrator tool)	A web application to manage and monitor the catalog resources, scanners, schedules, attributes, and connections. Used by Enterprise Data Catalog and Enterprise Data Lake.

When you plan the installation: Determine how many instances of the Developer tool you want to install. You do not need to plan for the web client applications.

Part II: Before You Install the Services

This part contains the following chapters:

CHAPTER 2

Before You Begin

This chapter includes the following topics:

Before You Begin Checklist

This chapter contains preliminary tasks that you must complete. Use this checklist to track preliminary tasks before you prepare for services.

- ❑ Read the Informatica Release Notes for updates to the installation and upgrade process.
- ❑ Verify system requirements:
  - Verify the Hadoop distribution.
  - Verify sizing requirements based upon your processing and concurrency requirements.
  - Review the patch requirements to verify that the machine has the required operating system patches and libraries.
  - Verify that the port numbers to use for application service processes are available on the machines where you install the Informatica services.
  - Verify that the operating system meets the file descriptor limit.
- ❑ Back up the Data Transformation files that were created in a previous installation.
- ❑ Review system environment variables.
- ❑ Create a system user account to run the installer.

- ❑ Set up keystore and truststore files if you want to configure secure communication for the domain and set up a secure connection to web client applications.
- ❑ Extract the installer files.
- ❑ Verify the license key.

Read the Release Notes

Read the Release Notes for updates to the installation and upgrade process. You can also find information about known and fixed limitations for the release.

Verify System Requirements

Verify that your environment meets the minimum system requirements for the installation process, temporary disk space, port availability, databases, and application service hardware.

For more information about product requirements and supported platforms, see the Product Availability Matrix on Informatica Network:
https://network.informatica.com/community/informatica-network/product-availability-matrices

Verify Temporary Disk Space and Permissions

Verify that your environment meets the minimum system requirements for the temporary disk space.

Disk space for the temporary files

The installer writes temporary files to the hard disk. Verify that you have 1 GB disk space on the machine to support the installation. When the installation completes, the installer deletes the temporary files and releases the disk space.

Permissions for the temporary files

Verify that you have read, write, and execute permissions on the `/tmp` directory.

For more information about product requirements and supported platforms, see the Product Availability Matrix on Informatica Network:
https://network.informatica.com/community/informatica-network/product-availability-matrices

Verify the Distributions

You can verify the distributions for the big data products in the Hadoop or in the Azure Databricks environment.

Hadoop Environment

Informatica big data products integrate with the non-native environment. The integration varies by product, as do the requirements at installation.

The following table lists the supported non-native distribution versions for the big data products:

Product	EMR	HDI	CDH	HDP	MapR
Big Data Management	5.16	3.6.x	5.15 5.16	2.6.x	6.0.x MEP 5.0
Big Data Streaming	5.16	3.6.x **Note:** HDI is supported only for ADLS non-Kerberos.	5.15 5.16	2.6.x	Deferred support
Enterprise Data Catalog	N/A	3.6.x **Note:** HDI is supported only for WASB non-Kerberos.	5.15 5.16 **Note:** You can use OpenJDK 1.8.0 only on Enterprise Data Catalog deployed on a CDH 5.16 Hadoop distribution.	2.6.x	N/A
Enterprise Data Lake	5.16	3.6.x **Note:** HDI is supported only for ADLS and WASB non-Kerberos.	5.15 5.16	2.6.x	6.0.x MEP 5.0

Note: Effective in version 10.2.2, Informatica dropped support for the Hive engine.

The following table lists the installer dependencies on the non-native environment for each product:

Product	Installer Dependency on Non-Native Environment
Informatica domain services *	The non-native environment is not required at install time. Integrate the environments after installation.
Enterprise Data Catalog	If you choose to use an external cluster, the non-native environment is required at install time. If you choose to use an embedded cluster, the non-native environment is not required at install time.
Enterprise Data Lake	The non-native environment is required at install time if you want to create and enable the Data Preparation Service and the Enterprise Data Lake Service when you run the installer. You complete the environment integration after installation.
**The Informatica domain services installation includes the following big data products: Big Data Management, Big Data Quality, and Big Data Streaming.*	

In each release, Informatica adds, defers, and drops support for the non-native distribution versions. Informatica might reinstate support for deferred versions in a future release. To see a list of the latest supported versions, see the Product Availability Matrix on the Informatica Customer Portal: https://network.informatica.com/community/informatica-network/product-availability-matrices.

Databricks Environment

Big Data Management can connect to Azure Databricks. Azure Databricks is an analytics cloud platform that is optimized for the Microsoft Azure cloud services. It incorporates the open-source Apache Spark cluster technologies and capabilities.

Informatica supports Big Data Management on Databricks distribution version 5.1. Informatica does not support Big Data Quality, Big Data Streaming, Enterprise Data Catalog, or Enterprise Data Lake on the Databricks environment.

Verify Sizing Requirements

Allocate resources for installation and deployment of services based on the expected deployment type of your environment.

Before you allocate resources, you need to identify the deployment type based on your requirements for the volume of processing and the level of concurrency. Based on the deployment type, you can allocate resources for disk space, cores, and RAM. You can also choose to tune services when you run the installer.

Determine the Installation and Service Deployment Type

The following table describes the environment for the different deployment types:

Deployment Type	Environment Description
Sandbox	Used for proof of concepts or as a sandbox with minimal users.
Basic	Used for low volume processing with low levels of concurrency.
Standard	Used for high volume processing with low levels of concurrency.
Advanced	Used for high volume processing with high levels of concurrency.

Identify Sizing Requirements

The following table provides the minimum sizing requirements:

Deployment Type	Disk Space per Node	Total Virtual Cores	RAM per Node
Sandbox	50 GB *	16	32 GB
Basic	100 GB	24	64 GB
Standard	100 GB	48	64 GB
Advanced	100 GB	96	64 GB
** Enterprise Data Catalog requires 100 GB of disk space for a Sandbox deployment type.*			

The sizing requirements account for the following factors:

- Disk space required to extract the installer
- Temporary disk space to run the installer
- Disk space required to install the services and components
- Disk space required for log directories
- Requirements to run the application services

The sizing numbers do not account for operational data processing and object caching requirements for native mode of execution.

Note: For cloud deployments, choose machines whose configuration is closest to the sizing requirements.

Tune During Installation

When you run the installer, you can choose to tune the services based on the deployment size. If you create a Model Repository Service, a Data Integration Service, or a Content Management Service during installation, the installer can tune the services based on the deployment type that you enter. The installer configures properties such as maximum heap size and execution pool size.

You can tune services at any time after you install the services by using the infacmd autotune command. When you run the command, you can tune properties for other services as well as the Hadoop run-time engine properties.

Review Patch Requirements

Before you install the Informatica services, verify that the machine has the required operating system patches and libraries.

The following table lists the patches and libraries required for installation:

Platform	Operating System	Operating System Patch
Linux-x64	Red Hat Enterprise Linux 6.7	All of the following packages, where <version> is any version of the package: - e2fsprogs-libs-<version>.el6 - keyutils-libs-<version>.el6 - libselinux-<version>.el6 - libsepol-<version>.el6
Linux-x64	Red Hat Enterprise Linux 7.3	All of the following packages, where <version> is any version of the package: - e2fsprogs-libs-<version>.el7 - keyutils-libs-<version>.el7 - libselinux-<version>.el7 - libsepol-<version>.el7
Linux-x64	SUSE Linux Enterprise Server 11	Service Pack 4
Linux-x64	SUSE Linux Enterprise Server 12	Service Pack 2

Verify Port Requirements

The installer sets up the ports for components in the Informatica domain, and it designates a range of dynamic ports to use for some application services.

You can specify the port numbers to use for the components and a range of dynamic port numbers to use for the application services. Or you can use the default port numbers provided by the installer. Verify that the port numbers are available on the machines where you install the Informatica domain services, Enterprise Data Catalog, or Enterprise Data Lake.

The following table describes the port requirements for installation:

Port	Description
Node port	Port number for the node created during installation. Default is 6005.
Service Manager port	Port number used by the Service Manager on the node. The Service Manager listens for incoming connection requests on this port. Client applications use this port to communicate with the services in the domain. The Informatica command line programs use this port to communicate to the domain. This is also the port for the SQL data service JDBC/ODBC driver. Default is 6006.
Service Manager Shutdown port	Port number that controls server shutdown for the domain Service Manager. The Service Manager listens for shutdown commands on this port. Default is 6007.
Informatica Administrator port	Port number used by Informatica Administrator. Default is 6008.
Informatica Administrator HTTPS port	No default port. Enter the required port number when you create the service. Setting this port to 0 disables an HTTPS connection to the Administrator tool.
Informatica Administrator shutdown port	Port number that controls server shutdown for Informatica Administrator. Informatica Administrator listens for shutdown commands on this port. Default is 6009.
Minimum port number	Lowest port number in the range of dynamic port numbers that can be assigned to the application service processes that run on this node. Default is 6014.
Maximum port number	Highest port number in the range of dynamic port numbers that can be assigned to the application service processes that run on this node. Default is 6114.
Range of dynamic ports for application services	Range of port numbers that can be dynamically assigned to application service processes as they start up. When you start an application service that uses a dynamic port, the Service Manager dynamically assigns the first available port in this range to the service process. The number of ports in the range must be at least twice the number of application service processes that run on the node. Default is 6014 to 6114. The Service Manager dynamically assigns port numbers from this range to the Model Repository Service.

Port	Description
HTTPS port for Hadoop distributions	If you deploy Enterprise Data Catalog in an HTTPS-enabled Hadoop distribution, the following are the default port numbers: - Cloudera. 7183 - Hortonworks. 8443 - Azure HDInsight. 8443 Required only if you install Enterprise Data Catalog.
Static ports for application services	Static ports have dedicated port numbers assigned that do not change. When you create the application service, you can accept the default port number, or you can manually assign the port number. The following services use static port numbers: - Catalog Service. Default is 9085 for HTTP. - Content Management Service. Default is 8105 for HTTP. - Data Integration Service. Default is 8095 for HTTP. - Data Preparation Service. Default is 8099 for HTTP. - Enterprise Data Lake Service. Default is 9045 for HTTP. - Informatica Cluster Service. Default is 9075 for HTTP. - Mass Ingestion Service. Default is 9050 for HTTP. - Metadata Access Service. Default is 7080 for HTTP. The Metadata Access Service uses consecutive port numbers to connect to multiple Hadoop distributions.

Note: Services and nodes can fail to start if there is a port conflict.

Guidelines for Port Configuration

The installer validates the port numbers that you specify to ensure that there will be no port conflicts in the domain.

Use the following guidelines to determine the port numbers:

- The port number you specify for the domain and for each component in the domain must be unique.
- The port number for the domain and domain components cannot be within the range of the port numbers that you specify for the application service processes.
- The highest number in the range of port numbers that you specify for the application service processes must be at least three numbers higher than the lowest port number. For example, if the minimum port number in the range is 6400, the maximum port number must be at least 6403.
- The port numbers that you specify cannot be lower than 1025 or higher than 65535.

Verify the File Descriptor Limit

Verify that the operating system meets the file descriptor requirement.

Informatica service processes can use a large number of files. To prevent errors that result from the large number of files and processes, you can change system settings with the limit command if you use a C shell, or the ulimit command if you use a Bash shell.

To get a list of the operating system settings, including the file descriptor limit, run the following command:
C Shell

```
limit
```

Bash Shell

```
ulimit -a
```

Informatica service processes can use a large number of files. Set the file descriptor limit per process to 16,000 or higher. The recommended limit is 32,000 file descriptors per process.

To change system settings, run the limit or ulimit command with the pertinent flag and value. For example, to set the file descriptor limit, run the following command:

C Shell

```
limit -h filesize <value>
```

Bash Shell

```
ulimit -n <value>
```

Informatica services use a large number of user processes. Use the ulimit -u command to adjust the max user processes setting to a level that is high enough to account for all the processes required by Blaze. Depending on the number of mappings and transformations that might run concurrently, set the file descriptor limit per process to 16,000 or higher.

Run the following command to set the max user processes setting:

C Shell

```
limit -u processes <value>
```

Bash Shell

```
ulimit -u <value>
```

Back Up the Data Transformation Files

Before installation, you must back up the Data Transformation files that were created under previous versions. After you complete the installation, copy the files to the new installation directories to get the same repository and custom global components as in the previous version.

The following table lists the files or directories that you must back up:

File or Directory	Default Location
Repository	<Informatica installation directory>\DataTransformation\ServiceDB
Custom Global Components directory (TGP files)	<Informatica installation directory>\DataTransformation\autoInclude\user
Custom Global Components directory (DLL and JAR files)	<Informatica installation directory>\DataTransformation\externLibs\user
Configuration file	<Informatica installation directory>\DataTransformation\CMConfig.xml
License file	<Informatica installation directory>\DataTransformation\CDELicense.cfg

Do not copy the Data Transformation Library files. Instead, install the Data Transformation Libraries again.

Review the Environment Variables

Configure the environment variables to work with the Informatica installation.

The following table describes the environment variables to review on Linux:

Variable	Description
IATEMPDIR	Location of the temporary files created during installation. Informatica requires 1 GB disk space for temporary files. Configure the environment variable if you do not want to create temporary files in the `/tmp` directory. If you want to change the default `/tmp` directory, you must set IATEMPDIR and _JAVA_OPTIONS environment variables to the new directory. For example, set the variable to export IATEMPDIR=/home/user. **Note:** Unset the IATEMPDIR variable after the installation.
_JAVA _OPTIONS	Configure the environment variable to change the temporary directory. If you want to change the default /tmp directory, you must set IATEMPDIR and _JAVA_OPTIONS the environment variables to the new directory. For example, set the variable to export _JAVA_OPTIONS=-Djava.io.tmpdir=/home/user. **Note:** Unset the JAVA _OPTIONS variable after the installation.
LANG and LC_ALL	Change the locale to set the appropriate character encoding for the terminal session. For example, set the encoding to `Latin1` or `ISO-8859-1` for French, `EUC-JP` or `Shift JIS` for Japanese, or `UTF-8` for Chinese or Korean. The character encoding determines the types of characters that appear in the UNIX terminal.
DISPLAY	Unset the DISPLAY environment before you run the installer. Installation might fail if the DISPLAY environment variable has some value.

Note: Make sure that the NOEXEC flag is not set for the file system mounted on the/tmp directory.

Create a System User Account

Create a user account specifically to run the Informatica daemon.

Verify that the user account you use to install Informatica has write permission on the installation directory.

Set Up a Keystore File

When you install the Informatica services, you can configure secure communication for the domain and set up a secure connection to Informatica Administrator (the Administrator tool). If you configure these security options, you must set up keystore and truststore files.

Before you install the Informatica services, set up the files for secure communication within the Informatica domain or for a secure connection to the Administrator tool. To create the required files, you can use the following programs:

keytool

You can use keytool to create an SSL certificate or a Certificate Signing Request (CSR) as well as keystores and truststores in JKS format.

For more information about using keytool, see the documentation on the following web site: http://docs.oracle.com/javase/7/docs/technotes/tools/windows/keytool.html.

OpenSSL

You can use OpenSSL to create an SSL certificate or CSR as well as convert a keystore in JKS format to PEM format.

For more information about OpenSSL, see the documentation on the following website: https://www.openssl.org/docs/

For a higher level of security, send your CSR to a Certificate Authority (CA) to get a signed certificate.

The software available for download at the referenced links belongs to a third party or third parties, not Informatica LLC. The download links are subject to the possibility of errors, omissions or change. Informatica assumes no responsibility for such links and/or such software, disclaims all warranties, either express or implied, including but not limited to, implied warranties of merchantability, fitness for a particular purpose, title and non-infringement, and disclaims all liability relating thereto.

Secure Communication Within the Informatica domain

Before you enable secure communication within the Informatica domain, verify that the following requirements are met:

You created a certificate signing request (CSR) and private key.

You can use keytool or OpenSSL to create the CSR and private key.

If you use RSA encryption, you must use more than 512 bits.

You have a signed SSL certificate.

The certificate can be self-signed or CA signed. Informatica recommends a CA signed certificate.

You imported the certificate into keystores.

You must have a keystore in PEM format named `infa_keystore.pem` and a keystore in JKS format named `infa_keystore.jks`.

The keystore files must contain the root and intermediate SSL certificates.

Note: The password for the keystore in JKS format must be the same as the private key pass phrase used to generate the SSL certificate.

You imported the certificate into truststores.

You must have a truststore in PEM format named `infa_truststore.pem` and a truststore in JKS format named `infa_truststore.jks`.

The truststore files must contain the root, intermediate, and end user SSL certificates.

The keystores and truststores are in the correct directory.

The keystore and truststore must be in a directory that is accessible to the installer.

For more information about how to create a custom keystore and truststore, see the Informatica How-To Library article "How to Create Keystore and Truststore Files for Secure Communication in the Informatica Domain": https://kb.informatica.com/h2l/HowTo%20Library/1/0700-CreateKeystoresAndTruststores-H2L.pdf

Secure Connection to the Administrator tool

Before you secure the connection to the Administrator tool, verify that the following requirements are met:

You created a certificate signing request (CSR) and private key.

You can use keytool or OpenSSL to create the CSR and private key.

If you use RSA encryption, you must use more than 512 bits.

You have a signed SSL certificate.

The certificate can be self-signed or CA signed. Informatica recommends a CA signed certificate.

You imported the certificate into a keystore in JKS format.

A keystore must contain only one certificate. If you use a unique certificate for each web application service, create a separate keystore for each certificate. Alternatively, you can use a shared certificate and keystore.

If you use the installer-generated SSL certificate for the Administrator tool, you do not need to import the certificate into a keystore in JKS format.

The keystore is in the correct directory.

The keystore must be in a directory that is accessible to the installer.

Extract the Installer Files

The installer files are compressed and distributed as a tar file.

Use a native tar or GNU tar utility to extract the installer files to a directory on your machine. The user that runs the installer must have read and write permissions on the installer files directory and execute permissions on install.sh.

You can get the installation file from the FTP link. Download the Informatica installation tar file from the Informatica Electronic Software Download site to a directory on your machine and then extract the installer files.

Note: Make sure that you download the file to a local directory or a shared network drive that is mapped on your machine. You can then extract the installer files. However, you cannot run the installer from a mapped file. Copy the extracted files to a local drive and then run the installer.

Installer Code Signing

You can verify the signature of the Informatica software code.

Informatica uses a certificate based digital signature to sign the Informatica software code. The code signing helps to validate the authenticity of the code and ensures that there has been no changes or corruptions to the code after Informatica signs the code. You can determine whether to trust the software based on whether the code sign is present or not.

You can request a code signing certificate that contains information that fully identifies Informatica LLC and a Certificate Authority (CA) that issues the certificate. The digital certificate binds the identity of Informatica to a public key and to a private key.

Digital signing of software begins with the creation of a cryptographic hash, or a digest. The digest has a one to one correspondence with the original data. Use the digest as there are no hints on how to recreate the original data, and even a small change in the original data results in a change in the hash value. Informatica uses its private key to sign the digest, or generates a signature in the form of a string of bits. Good digital signature algorithms allow a user with the public key to verify the creator of the signature.

To Verify the Signed Code is Authentic

After Informatica signs the software bundle, you can contact Informatica Global Customer Support to access the code signing certificate. Informatica ships the installer along with the signature file that contains the hash of the installer binary encrypted with Informatica's private key. You can validate the integrity of digitally signed binaries using any available tools, such as OpenSSL.

For instance, if you have to verify the package authentication and confirm the code security, enter the following two OpenSSL commands:

```
openssl base64 -d -in $signature -out /tmp/sign.sha256
openssl dgst -sha256 -verify (openssl x509 -in <cert> -pubkey -noout) -signature /tmp/
sign.sha256 <file>
```

Where `<signature>` is the file containing the signature in Base64, `<cert>` is the code signing certificate, and `<file>` is the file to verify.

If the verification is successful, OpenSSL displays a message to validate if it is a successful certificate or not. To verify the 10.2.2 Informatica server installer on Linux, it might take around two minutes.

Verify the License Key

Before you install the software, verify that you have the license key available.

When you download the installation files from the Informatica Electronic Software Download (ESD) site, the license key is in an email message from Informatica. Copy the license key file to a directory accessible to the user account that installs the product.

Contact Informatica Global Customer Support if you do not have a license key or if you have an incremental license key and you want to create a domain.

Run the Pre-Installation (i10Pi) System Check Tool

Run the Pre-installation (i10Pi) System Check Tool to verify whether the machine meets the system requirements for installation or upgrade.

Ensure that you verified the system requirements and prepared the domain configuration repository database.

1. Log in to the machine with a system user account.
2. Close all other applications.
3. On a shell command line, run the install.sh file from the root directory.

The installer displays the message to verify that the locale environment variables are set.

4. If the environment variables are not set, press **n** to exit the installer and set them as required.

   If the environment variables are set, press **y** to continue.
5. Press **1** to install or upgrade Informatica.
6. Press **1** to run the Pre-Installation (i10Pi) System Check Tool that verifies whether the machine meets the system requirements for the installation or upgrade.
7. From the Informatica Pre-Installation (i10Pi) System Check Tool **Welcome** section, press **Enter**.

   The **System Information** section appears.
8. Type the absolute path for the installation directory.

   The directory names in the path must not contain spaces or the following special characters: @|* $ # ! % ( ) { } [ ] , ; '

   Note: Informatica recommends using alphanumeric characters in the installation directory path. If you use a special character such as á or €, unexpected results might occur at run time.
9. Press **Enter**.
10. Enter the starting port number for the node that you will create or upgrade on the machine. The default port number for the node is 6005.
11. Press **Enter**.

    The **Database and Connection Information** section appears.
12. To enter the JDBC connection information using a custom JDBC connection string, press **1**. To enter the JDBC connection information using the JDBC URL information, press **2**.

    To connect to a secure database, you must enter the JDBC connection using a custom JDBC connection string.
13. Enter the JDBC connection information.

    - To enter the connection information using a custom JDBC connection string, type the connection string and specify the connection parameters.
      Use the following syntax in the JDBC connection string:
      IBM DB2

      ```
      jdbc:Informatica:db2://host_name:port_no;DatabaseName=
      ```

      Oracle

      ```
      jdbc:Informatica:oracle://host_name:port_no;ServiceName=
      ```

      Microsoft SQL Server

      ```
      jdbc:Informatica:sqlserver://host_name:port_no;SelectMethod=cursor;DatabaseName=
      ```

      Microsoft Azure SQL Database

      ```
      jdbc:Informatica:sqlserver://
      host_name:port_no;SelectMethod=cursor;DatabaseName=database_name;SnapshotSerializa
      ble=true;EncryptionMethod=SSL;HostNameInCertificate=*.database.windows.net;Validat
      eServerCertificate=false
      ```

      Sybase

      ```
      jdbc:Informatica:sybase://host_name:port_no;DatabaseName=
      ```

      Verify that the connection string contains all the connection parameters required by your database system.

- To enter the connection information using the JDBC URL information, specify the JDBC URL properties.
  The following table describes the connection information:

Prompt	Description
Database type	Type of database for the domain configuration repository. Select from the following database types: - 1 - Oracle - 2 - Microsoft SQL Server - 3 - IBM DB2 - 4 - Sybase ASE
Database user ID	User ID for the database user account for the domain configuration repository.
Database user password	Password for the database user account.
Database host name	Host name for the database server.
Database port number	Port number for the database.
Database service name	Service name for Oracle and IBM DB2 databases or database name for Microsoft SQL Server and Sybase ASE.

- To connect to a secure database, select **1** to use a custom string and type the connection string. You must include the security parameters in addition to the connection parameters. For information about the security parameters you must include in the JDBC connection for a secure database, see "Connection String to a Secure Database" on page 97.

The tool checks the settings of the hard drive, the availability of the ports, and the configuration of the database. After the system check is complete, the **System Check Summary** section displays the results of the system check.

14. Analyze the results of the system check.

    Each requirement is listed, along with one of the following check statuses:

    - [Pass] - The requirement meets the criteria for the Informatica installation or upgrade.
    - [Fail] - The requirement does not meet the criteria for the Informatica installation or upgrade. Resolve the issue before you proceed with the installation or upgrade.
    - [Information] - Verify the information and perform any additional tasks as outlined in the details.

    The results of the system check are saved to the following file: `.../Server/i10Pi/i10Pi/en/i10Pi_summary.txt`

15. Press **Enter** to close the Pre-Installation (i10Pi) System Check Tool.

    You can continue to the Informatica service installation or upgrade immediately or end the system check and continue with the installation or upgrade later. If you continue to the installation or upgrade immediately, you do not have to restart the installer.

16. To continue to the Informatica service installation or upgrade immediately, press **y**.

    To end the system check and continue with the installation or upgrade later, press **n**.

If the Pre-Installation (i10Pi) System Check Tool finishes with failed requirements, resolve the failed requirements and run the Pre-Installation (i10Pi) System Check Tool again.

Note: If the Informatica Pre-Installation (i10Pi) System Check Tool check finishes with failed requirements, you can still perform the Informatica installation or upgrade. However, Informatica highly recommends that you resolve the failed requirements before you proceed.

CHAPTER 3

Prepare for Application Services and Databases

This chapter includes the following topics:

Checklist to Prepare for Application Services

This chapter contains information about application services and databases for the different big data products. Use this checklist to track service planning and database preparation.

- ❑ Identify the application services that you need in your environment.
- ❑ Identify the application services that you want the installer to create.

❑ Prepare databases for the services:

- Create the database.
- Create a user for the database.
- Create environment variables.
- Configure connectivity.

Prepare for Application Services and Databases Overview

When you plan the application services, you must account for the associated services that connect to the application service. You also must plan the relational databases that the application service requires.

The installer prompts you to optionally create some services during the installation. Some service properties require database information. If you want the installer to create a service that requires a database, you must prepare the database before you run the installer. To prepare the databases, verify the data base requirements, set up the database, and set up a user account. The database requirements depend on the application services that you create.

If you do not create services during installation, you can create them manually after you install.

Set Up Database User Accounts

Set up a database and user account for the domain configuration repository and for the repository databases associated with the applications services.

Use the following rules and guidelines when you set up the user accounts:

- The database user account must have permissions to create and drop tables, indexes, and views, and to select, insert, update, and delete data from tables.
- Use 7-bit ASCII to create the password for the account.
- To prevent database errors in one repository from affecting any other repository, create each repository in a separate database schema with a different database user account. Do not create a repository in the same database schema as the domain configuration repository or any other repository in the domain.
- If you create more than one domain, each domain configuration repository must have a separate user account.

Identify Application Services by Product

Each application service provides different functionality within the Informatica domain. You create application services based on the license key generated for your organization.

The following table lists the application services that each product uses:

Product	Application Services
Big Data Management	- Analyst Service - Data Integration Service * - Mass Ingestion Service - Metadata Access Service - Model Repository Service *
Big Data Parser	- Data Integration Service * - Mass Ingestion Service - Metadata Access Service - Model Repository Service *
Big Data Quality	- Analyst Service - Content Management Service - Data Integration Service * - Mass Ingestion Service - Metadata Access Service - Model Repository Service * - Search Service
Big Data Streaming	- Analyst Service - Data Integration Service * - Mass Ingestion Service - Metadata Access Service - Model Repository Service *
Enterprise Data Catalog	- Analyst Service - Catalog Service * - Content Management Service * - Data Integration Service * - Model Repository Service * - Informatica Cluster Service * - Search Service
Enterprise Data Lake	- Catalog Service - Content Management Service - Data Integration Service * - Data Preparation Service * - Enterprise Data Lake Service * - Model Repository Service *
** You can create these services when you install the product.*	

Domain Configuration Repository Database Requirements

Informatica components store metadata in relational database repositories. The domain stores configuration and user information in a domain configuration repository.

You must set up a database and user account for the domain configuration repository before you run the installation. The database must be accessible to all gateway nodes in the Informatica domain.

When you install Informatica, you provide the database and user account information for the domain configuration repository. The Informatica installer uses JDBC to communicate with the domain configuration repository.

The domain configuration repository supports the following database types:

- IBM DB2 UDB
- Microsoft SQL Server
- Microsoft Azure SQL Database
- Oracle
- Sybase ASE

Allow 200 MB of disk space for the database.

IBM DB2 Database Requirements

Use the following guidelines when you set up the repository on IBM DB2:

- If the repository is in an IBM DB2 database, verify that IBM DB2 Version 10.5 is installed.
- On the IBM DB2 instance where you create the database, set the following parameters to ON:
  - DB2_SKIPINSERTED
  - DB2_EVALUNCOMMITTED
  - DB2_SKIPDELETED
  - AUTO_RUNSTATS
- On the database, set the configuration parameters.

  The following table lists the configuration parameters that you must set:

Parameter	Value
logfilsiz	8000
maxlocks	98
locklist	50000
auto_stmt_stats	ON

- Set the tablespace pageSize parameter to 32768 bytes.

  In a single-partition database, specify a tablespace that meets the pageSize requirements. If you do not specify a tablespace, the default tablespace must meet the pageSize requirements.

In a multi-partition database, specify a tablespace that meets the pageSize requirements. Define the tablespace in the catalog partition of the database.

- Set the NPAGES parameter to at least 5000. The NPAGES parameter determines the number of pages in the tablespace.
- Verify that the database user has CREATETAB, CONNECT, and BINDADD privileges.
- Informatica does not support IBM DB2 table aliases for repository tables. Verify that table aliases have not been created for any tables in the database.
- In the DataDirect Connect for JDBC utility, update the DynamicSections parameter to 3000.

  The default value for DynamicSections is too low for the Informatica repositories. Informatica requires a larger DB2 package than the default. When you set up the DB2 database for the domain configuration repository or a Model repository, you must set the DynamicSections parameter to at least 3000. If the DynamicSections parameter is set to a lower number, you can encounter problems when you install or run Informatica services.

  For more information about updating the DynamicSections parameter, see Appendix C, "Updating the DynamicSections Parameter of a DB2 Database" on page 347.

Microsoft SQL Server Database Requirements

Use the following guidelines when you set up the repository:

- Set the allow snapshot isolation and read committed isolation level to ALLOW_SNAPSHOT_ISOLATION and READ_COMMITTED_SNAPSHOT to minimize locking contention.
  To set the isolation level for the database, run the following commands:

  ```
  ALTER DATABASE DatabaseName SET ALLOW_SNAPSHOT_ISOLATION ON
  ```

  ```
  ALTER DATABASE DatabaseName SET READ_COMMITTED_SNAPSHOT ON
  ```

  To verify that the isolation level for the database is correct, run the following commands:

  ```
  SELECT snapshot_isolation_state FROM sys.databases WHERE name=[DatabaseName]
  ```

  ```
  SELECT is_read_committed_snapshot_on FROM sys.databases WHERE name = DatabaseName
  ```

- The database user account must have the CONNECT, CREATE TABLE, and CREATE VIEW privileges.

Oracle Database Requirements

Use the following guidelines when you set up the repository on Oracle:

- Set the open_cursors parameter to 4000 or higher.
- Set the permissions on the view `$parameter` for the database user.
- Set the privileges for the database user to run *show parameter open_cursors* in the Oracle database. When you run the pre-installation (i10Pi) system check tool, i10Pi runs the command against the database to identify the OPEN_CURSORS parameter with the domain database user credentials.

  You can run the following query to determine the open cursors setting for the domain database user account:

  ```
  SELECT VALUE OPEN_CURSORS FROM V$PARAMETER WHERE UPPER(NAME)=UPPER('OPEN_CURSORS')
  ```

- Verify that the database user has the following privileges:

  CREATE SEQUENCE

  CREATE SESSION

  CREATE SYNONYM

CREATE TABLE

CREATE VIEW

- Informatica does not support Oracle public synonyms for repository tables. Verify that public synonyms have not been created for any tables in the database.

Sybase Database Requirements

Use the following guidelines when you set up the repository on Sybase ASE:

- Set the database server page size to 16K or higher. You must set the page size to 16K as this is a one-time configuration and cannot be changed afterwards.
- Set the database locking configuration to use row-level locking.
  The following table describes the database locking configuration that you must set:

Database Configuration	Sybase System Procedure	Value
Lock scheme	sp_configure "lock scheme"	0, datarows

- Set the Sybase database option "ddl in tran" to TRUE.
- Set "allow nulls by default" to TRUE.
- Turn ON the Sybase database option select into/bulkcopy/pllsort.
- Enable the "select" privilege for the sysobjects system table.
- Create the following login script to disable the default VARCHAR truncation:

```
create procedure dbo.sp_string_rtrunc_proc as set string_rtruncation on
sp_modifylogin "user_name", "login script", sp_string_rtrunc_proc
```

  The login script is executed every time the user logs into the Sybase instance. The stored procedure sets the parameter at the session level. The sp_modifylogin system procedure updates "user_name" with the stored procedure as its "login script". The user must have permission to invoke the stored procedure.

- Verify that the database user has CREATE DEFAULT, CREATE PROCEDURE, CREATE RULE, CREATE TABLE, and CREATE VIEW privileges.
- Set the database configurations to the recommended baseline values.
  The following table lists the database memory configuration parameters that you must set:

Database Configuration	Sybase System Procedure	Value
Maximum amount of total physical memory	sp_configure "max memory"	2097151
Procedure cache size	sp_configure "procedure cache size"	500000
Number of open objects	sp_configure "number of open objects"	5000
Number of open indexes	sp_configure "number of open indexes"	5000
Number of open partitions	sp_configure "number of open partitions"	5000
Heap memory per user	sp_configure "heap memory per user"	49152
Number of locks	sp_configure "number of locks"	100000

Analyst Service

The Analyst service runs the Analyst tool. It manages the connections between service components and the user that have access to the Analyst tool. When you create the service, you need to associate other application services with it.

The following table summarizes some dependencies that are associated with the Analyst Service:

Dependency	Summary
Products	The following products use the Analyst Service: - Big Data Management - Big Data Quality - Big Data Streaming - Enterprise Data Catalog
Services	The Analyst Service requires a direct association with the following services: - Data Integration Service - Model Repository Service
Databases	The Analyst Service does not have any associated database.
Installer	You cannot create the Analyst Service during installation.

Catalog Service

The Catalog Service manages the connections between service components and the users that have access to Enterprise Data Catalog search interface and Catalog Administrator.

The following table summarizes the dependencies for products, services, and databases that are associated with the Catalog Service:

Dependency	Summary
Products	The following products use the Catalog Service: - Enterprise Data Catalog - Enterprise Data Lake
Services	The Catalog Service depends on the following services: - Content Management Service - Data Integration Service - Informatica Cluster Service * - Model Repository Service
Databases	The Catalog Service does not have any associated database.
Installer	You can create the Catalog Service when you install Enterprise Data Catalog.
** Required and available only on an embedded cluster.*	

Content Management Service

The Content Management Service manages reference data for data domains that use reference tables. It uses the Data Integration Service to run mappings to transfer data between reference tables and external data sources. When you create the service, you need to associate other application services with it.

The following table summarizes the dependencies for products, services, and databases that are associated with the Content Management Service:

Dependency	Summary
Products	The following products use the Content Management Service: - Big Data Quality - Enterprise Data Catalog - Enterprise Data Lake
Services	The Content Management Service requires a direct association with the following services: - Model Repository Service - Data Integration Service
Databases	The Content Management Service uses the following database: - Reference data warehouse. Stores data values for the reference table objects that you define in the Model repository. When you add data to a reference table, the Content Management Service writes the data values to a table in the reference data warehouse.
Installer	You can create the Content Management Service when you install Enterprise Data Catalog.

Rules and Guidelines

Consider the following rules and guidelines for Content Management Service creation:

- You must create the Content Management Service on the same node as the Data Integration Service.

Reference Data Warehouse Requirements

The reference data warehouse stores the data values for reference table objects that you define in a Model repository. You configure a Content Management Service to identify the reference data warehouse and the Model repository.

You associate a reference data warehouse with a single Model repository. You can select a common reference data warehouse on multiple Content Management Services if the Content Management Services identify a common Model repository. The reference data warehouse must support mixed-case column names.

The reference data warehouse supports the following database types:

- IBM DB2 UDB
- Microsoft SQL Server
- Microsoft Azure SQL Database
- Oracle

Allow 200 MB of disk space for the database.

Note: Ensure that you install the database client on the machine on which you want to run the Content Management Service.

IBM DB2 Database Requirements

Use the following guidelines when you set up the repository on IBM DB2:

- Verify that the database user account has CREATETAB and CONNECT privileges.
- Verify that the database user has SELECT privileges on the SYSCAT.DBAUTH and SYSCAT.DBTABAUTH tables.
- Informatica does not support IBM DB2 table aliases for repository tables. Verify that table aliases have not been created for any tables in the database.
- Set the tablespace pageSize parameter to 32768 bytes.
- Set the NPAGES parameter to at least 5000. The NPAGES parameter determines the number of pages in the tablespace.

Microsoft SQL Server Database Requirements

Use the following guidelines when you set up the repository:

- Verify that the database user account has CONNECT and CREATE TABLE privileges.

Oracle Database Requirements

Use the following guidelines when you set up the repository on Oracle:

- Verify that the database user has the following privileges:

  ALTER SEQUENCE

  ALTER TABLE

  CREATE SEQUENCE

  CREATE SESSION

  CREATE TABLE

  CREATE VIEW

  DROP SEQUENCE

  DROP TABLE
- Informatica does not support Oracle public synonyms for repository tables. Verify that public synonyms have not been created for any tables in the database.

Data Integration Service

The Data Integration Service receives requests from Informatica client tools to run integration, profile, and data preparation jobs. It writes results to different databases, and it writes run-time metadata to the Model repository. When you create the service, you need to associate another application service with it.

The following table lists the dependencies for products, services, and databases that are associated with the Data Integration Service.

Dependency	Summary
Products	The following products use the Data Integration Service: - Big Data Management - Big Data Parsing - Big Data Quality - Big Data Streaming - Enterprise Data Catalog - Enterprise Data Lake
Services	The Data Integration Service requires a direct association with the following service: - Model Repository Service
Databases	The Data Integration Service uses the following databases: - Data object cache. Stores cached logical data objects and virtual tables. - Profiling warehouse. Stores profiling information, such as profile and scorecard results. - Workflow database. Stores run-time metadata for workflows.
Installer	You can create the Data Integration Service when you run the installer.

Rules and Guidelines

Consider the following rules and guidelines for Data Integration Service creation:

- If you plan to use operating system profiles for Big Data Management, you must create a dedicated Data Integration Service for Enterprise Data Catalog. Enterprise Data Catalog does not support operating system profiles.
- Informatica recommends creating a dedicated Data Integration Service for Enterprise Data Lake.

Data Object Cache Database Requirements

The data object cache database stores cached logical data objects and virtual tables for the Data Integration Service. You specify the data object cache database connection when you create the Data Integration Service.

The data object cache database supports the following database types:

- IBM DB2 UDB
- Microsoft SQL Server
- Microsoft Azure SQL Database
- Oracle

Allow 200 MB of disk space for the database.

Note: Ensure that you install the database client on the machine on which you want to run the Data Integration Service.

IBM DB2 Database Requirements

Use the following guidelines when you set up the repository on IBM DB2:

- Verify that the database user account has CREATETAB and CONNECT privileges.
- Informatica does not support IBM DB2 table aliases for repository tables. Verify that table aliases have not been created for any tables in the database.
- Set the tablespace pageSize parameter to 32768 bytes.
- Set the NPAGES parameter to at least 5000. The NPAGES parameter determines the number of pages in the tablespace.

Microsoft SQL Server Database Requirements

Use the following guidelines when you set up the repository:

- Verify that the database user account has CONNECT and CREATE TABLE privileges.

Oracle Database Requirements

Use the following guidelines when you set up the repository on Oracle:

- Verify that the database user has the following privileges:

  CREATE INDEX

  CREATE SESSION

  CREATE SYNONYM

  CREATE TABLE

  CREATE VIEW

  DROP TABLE

  INSERT INTO TABLE

  UPDATE TABLE

- Informatica does not support Oracle public synonyms for repository tables. Verify that public synonyms have not been created for any tables in the database.

Profiling Warehouse Requirements

The profiling warehouse database stores profiling and scorecard results. You specify the profiling warehouse connection when you create the Data Integration Service.

The profiling warehouse supports the following database types:

- IBM DB2 UDB
- Microsoft SQL Server
- Microsoft Azure SQL Database
- Oracle

Allow 10 GB of disk space for the database.

Note: Ensure that you install the database client on the machine on which you want to run the Data Integration Service. You can specify a JDBC connection or Hive connection as a profiling warehouse connection for IBM DB2 UDB, Microsoft SQL Server, and Oracle database types.

For more information about configuring the database, see the documentation for your database system.

IBM DB2 Database Requirements

Use the following guidelines when you set up the repository on IBM DB2:

- The database user account must have the CREATETAB, CONNECT, CREATE VIEW, and CREATE FUNCTION privileges.
- Informatica does not support IBM DB2 table aliases for repository tables. Verify that table aliases have not been created for any tables in the database.
- Set the tablespace pageSize parameter to 32768 bytes.
- Set the NPAGES parameter to at least 5000. The NPAGES parameter determines the number of pages in the tablespace.

Note: Informatica does not support the partitioned database environment for IBM DB2 databases when you use a JDBC connection as the profiling warehouse connection.

Microsoft SQL Server Database Requirements

Use the following guidelines when you set up the repository:

- The database user account must have the CONNECT, CREATE TABLE, CREATE VIEW, and CREATE FUNCTION privileges.

Oracle Database Requirements

Use the following guidelines when you set up the repository on Oracle:

- Verify that the database user has the following privileges:

  ALTER TABLE

  CREATE ANY INDEX

  CREATE PROCEDURE

  CREATE SESSION

  CREATE TABLE

  CREATE VIEW

  DROP TABLE

  UPDATE TABLE
- Informatica does not support Oracle public synonyms for repository tables. Verify that public synonyms have not been created for any tables in the database.
- Set the tablespace parameter.
- Set the following parameters to the Informatica recommended values:

Parameter	Recommended Value
open_cursors	3000
Sessions	1000
Processes	1000

Workflow Database Requirements

The Data Integration Service stores run-time metadata for workflows in the workflow database. Before you create the workflow database, set up a database and database user account for the workflow database.

You specify the workflow database connection when you create the Data Integration Service.

The workflow database supports the following database types:

- IBM DB2 UDB
- Microsoft SQL Server
- Microsoft Azure SQL Database
- Oracle

Allow 200 MB of disk space for the database.

Note: Ensure that you install the database client on the machine on which you want to run the Data Integration Service.

IBM DB2 Database Requirements

Use the following guidelines when you set up the repository on IBM DB2:

- Verify that the database user account has CREATETAB and CONNECT privileges.
- Informatica does not support IBM DB2 table aliases for repository tables. Verify that table aliases have not been created for any tables in the database.
- Set the tablespace pageSize parameter to 32768 bytes.
- Set the NPAGES parameter to at least 5000. The NPAGES parameter determines the number of pages in the tablespace.
- Set the connection pooling parameters.

  The following table lists the connection pooling parameters that you must set:

Parameter	Value
Maximum Connection Pool Size	128
Minimum Connection Pool Size	0
Maximum Idle Time	120 seconds

Microsoft SQL Server Database Requirements

Use the following guidelines when you set up the repository:

- Verify that the database user account has CONNECT and CREATE TABLE privileges.
- Set the connection pooling parameters.

The following table lists the connection pooling parameters that you must set:

Parameter	Value
Maximum Connection Pool Size	128
Minimum Connection Pool Size	0
Maximum Idle Time	120 seconds

Oracle Database Requirements

Use the following guidelines when you set up the repository on Oracle:

- Verify that the database user has the following privileges:

  ALTER TABLE

  ALTER VIEW

  CREATE SEQUENCE

  CREATE SESSION

  CREATE SYNONYM

  CREATE TABLE

  CREATE VIEW

  DROP TABLE

  DROP VIEW

- Informatica does not support Oracle public synonyms for repository tables. Verify that public synonyms have not been created for any tables in the database.
- Set the connection pooling parameters.

  The following table lists the connection pooling parameters that you must set:

Parameter	Value
Maximum Connection Pool Size	128
Minimum Connection Pool Size	0
Maximum Idle Time	120 seconds

Data Preparation Service

The Data Preparation Service manages data preparation within Enterprise Data Lake. When an analyst prepares data in a project, the Data Preparation Service stores worksheet metadata in the Data Preparation repository. When you create the service, you can associate other application services with it.

The following table summarizes the dependencies for products, services, and databases that are associated with the Data Preparation Service:

Dependency	Summary
Products	The following products use the Data Preparation Service: - Enterprise Data Lake
Services	If you plan to use rules during data preparation, you can provide a direct association with the following services: - Model Repository Service - Data Integration Service
Databases	The Data Preparation Service uses the following database: - Data Preparation repository. Stores worksheet metadata created when users prepare data assets for publication.
Installer	You can create the Data Preparation Service when you run the installer.

Rules and Guidelines

Consider the following rules and guidelines for Data Preparation Service creation:

- If you use the installer to create the Enterprise Data Lake and Data Preparation Service, you must create both of the application services on the same node.
- If you configure Enterprise Data Lake to use rules, you must associate a Data Integration Service with the Data Preparation Service. The Data Integration Service is required to run rules during data preparation.
- If you configure Enterprise Data Lake to use rules, you must associate the Model Repository Service that manages the Model repository in which rule objects and metadata are stored with the Data Preparation Service.
- If you want to create and enable the Data Preparation Service when you run the installer, the domain must contain connections associated with the Hadoop environment. For more information, see "Prepare to Create the Enterprise Data Lake Services" on page 61.

Data Preparation Repository Database Requirements

Set up the MySQL database or the Oracle database to use as the Data Preparation repository. The Data Preparation Service stores recipe and mapping metadata in the repository.

Allow 5 GB of disk space for the repository database. Allocate more space based on the amount of metadata you want to store.

MySQL and MariaDB Database Requirements

You can use a MySQL database or a MariaDB database as the Data Preparation repository.

Set the following system variables on the database server:

- For MySQL version 5.6.26 and higher, set lower_case_table_names=1.

- For MySQL version 5.7 and higher, set explicit_defaults_for_timestamp=1.

Set the same system variable values for a MariaDB database.

Ensure that the MySQL or MariaDB database has the following permissions:

- Create tables and views.
- Drop tables and views.
- Insert, update, and delete data.

The MySQL connector .jar file is not included with the installer. You must download the file and copy it to the following directory before you start the installer:

```
$USER_INSTALL_DIR$/services/shared/jars/thirdparty/
```

Make sure the name of the file is in the following format:

```
mysql-connector-java-<versiondetails>.jar
```

You must also set the mysql_connector_jar_path environment variable to the location of the MySQL connector .jar file.

Oracle Database Requirements

You can use an Oracle database as the Data Preparation repository.

Ensure that the database has the following permissions:

- Create tables and views.
- Create sequence, session, and synonyms.
- Drop tables and views.
- Insert, update, and delete data.

Grant Permissions on the Hive Warehouse Directory

The Data Preparation Service uses an HDFS location for data preparation file storage. You must grant access to the absolute HDFS file path of the default database for the Hive warehouse. You must also grant read and write permissions on the Hive warehouse directory.

You can find the location of the warehouse directory in the hive.metastore.warehouse.dir property of the hive-site.xml file. For example, the default might be /user/hive/warehouse or /apps/hive/warehouse.

Grant the appropriate HDFS permissions to the Hadoop impersonation user. HDFS permissions determine what a user can do to files and directories stored in HDFS. To access a file or directory, a user must have permissions or belong to a group that has permissions to the file or directory.

Enterprise Data Lake Service

The Enterprise Data Lake Service runs the Enterprise Data Lake application in the Informatica domain. When you create the service, you need to associate other application services with it.

The following table summarizes the dependencies for products, services, and databases that are associated with the Enterprise Data Lake Service:

Dependency	Summary
Products	The following products use the Enterprise Data Lake Service: - Enterprise Data Lake
Services	The Enterprise Data Lake Service requires a direct association with the following services: - Model Repository Service - Data Integration Service - Data Preparation Service
Databases	The Enterprise Data Lake Service uses the following database: - Model repository. Stores mappings generated for published data assets.
Installer	You can create the Enterprise Data Lake Service when you run the installer.

Rules and Guidelines

Consider the following rules and guidelines for Enterprise Data Lake Service creation:

- If you create the Enterprise Data Lake Service and Data Preparation Service during installation, you must create both of the application services on the same node.
- If you want to create and enable the Enterprise Data Lake Service when you run the installer, the domain must contain connections associated with the Hadoop environment. For more information, see "Prepare to Create the Enterprise Data Lake Services" on page 61.

Informatica Cluster Service

The Informatica Cluster Service runs and manages the embedded Hadoop cluster that runs with Enterprise Data Catalog. It distributes the Hortonworks binaries and launches the required Hadoop services on the hosts where the embedded cluster runs.

The following table summarizes the dependencies for products, services, and databases that are associated with the Informatica Cluster Service:

Dependency	Summary
Products	The following product uses the Informatica Cluster Service: - Enterprise Data Catalog
Services	The Informatica Cluster Service properties do require an association with another application service.

Dependency	Summary
Databases	The Informatica Cluster Services does not have any associated databases.
Installer	You can create the Informatica Cluster Service when you install Enterprise Data Catalog on an embedded cluster.

Mass Ingestion Service

The Mass Ingestion Service manages and validates mass ingestion specifications that ingest data to targets in the Hadoop environment. When you create the service, you need to associate another application service with it.

The following table summarizes the dependencies for products, services, and databases that are associated with the Mass Ingestion Service:

Dependency	Summary
Products	The following products use the Mass Ingestion Service: - Big Data Management - Big Data Parser - Big Data Quality - Big Data Streaming
Services	The Mass Ingestion Service requires a direct association with the following service: - Model Repository Service
Databases	The Mass Ingestion Service does not have any associated database.
Installer	You cannot create the Mass Ingestion Service when you run the installer.

Metadata Access Service

The Metadata Access Service allows the Developer tool to access Hadoop connection information to import and preview metadata from the Hadoop environment. The Metadata Access Service is required for design-time access to the Hadoop environment.

The following table summarizes the dependencies for products, services, and databases that are associated with the Metadata Access Service:

Dependency	Summary
Products	The following products use the Metadata Access Service: - Big Data Management - Big Data Parser - Big Data Quality - Big Data Streaming
Services	The Metadata Access Service properties do require an association with another application service.
Databases	The Metadata Access Service does not have any associated database.
Installer	You cannot create the Metadata Access Service when you run the installer.

Rules and Guidelines

Consider the following rules and guidelines for the Metadata Access Service creation:

- You must create the service after the installation completes.

Model Repository Service

The Model Repository Service manages the Model repository. It receives requests from Informatica clients and application services to store or access metadata in the Model repository.

The following table summarizes the dependencies for products, services, and databases that are associated with the Model Repository Service.

Dependency	Summary
Products	The following products use the Model Repository Service: - Big Data Management - Big Data Parsing - Big Data Quality - Big Data Streaming - Enterprise Data Catalog - Enterprise Data Lake
Services	The Model Repository Service properties do require an association with another application service.

Dependency	Summary
Databases	The Model Repository Service uses the following database: - Model repository. Stores metadata created by Informatica clients and application services.
Installer	You can create the Model Repository Service when you run the installer.

Model Repository Database Requirements

Informatica services and clients store data and metadata in the Model repository. Configure a monitoring Model repository to store statistics for ad hoc jobs, applications, logical data objects, SQL data services, web services, and workflows. Before you create the Model Repository Service, set up a database and database user account for the Model repository. It is recommended that you use different database configuration for Model repository and monitoring Model repository.

The Model repository supports the following database types:

- IBM DB2 UDB
- Microsoft SQL Server
- Microsoft Azure SQL Database
- Oracle

Allow 3 GB of disk space for DB2. Allow 200 MB of disk space for all other database types.

When you configure Microsoft SQL Server, you can choose to configure the Microsoft Azure SQL Database as the Model repository.

For more information about configuring the database, see the documentation for your database system.

IBM DB2 Database Requirements

Use the following guidelines when you set up the repository on IBM DB2:

- Specify the tablespace name when you use IBM DB2 as the Model Repository database.
- If the repository is in an IBM DB2 database, verify that IBM DB2 Version 10.5 is installed.
- On the IBM DB2 instance where you create the database, set the following parameters to ON:
  - DB2_SKIPINSERTED
  - DB2_EVALUNCOMMITTED
  - DB2_SKIPDELETED
  - AUTO_RUNSTATS
- On the database, set the configuration parameters.

  The following table lists the configuration parameters that you must set:

Parameter	Value
logfilsiz	8000
maxlocks	98

Parameter	Value
locklist	50000
auto_stmt_stats	ON

- Set the tablespace pageSize parameter to 32768 bytes.

  In a single-partition database, specify a tablespace that meets the pageSize requirements. If you do not specify a tablespace, the default tablespace must meet the pageSize requirements.

  In a multi-partition database, specify a tablespace that meets the pageSize requirements. Define the tablespace in the catalog partition of the database.
- Set the NPAGES parameter to at least 5000. The NPAGES parameter determines the number of pages in the tablespace.
- Verify that the database user has CREATETAB, CONNECT, and BINDADD privileges.
- Informatica does not support IBM DB2 table aliases for repository tables. Verify that table aliases have not been created for any tables in the database.
- In the DataDirect Connect for JDBC utility, update the DynamicSections parameter to 3000.

  The default value for DynamicSections is too low for the Informatica repositories. Informatica requires a larger DB2 package than the default. When you set up the DB2 database for the domain configuration repository or a Model repository, you must set the DynamicSections parameter to at least 3000. If the DynamicSections parameter is set to a lower number, you can encounter problems when you install or run Informatica services.

  For more information about updating the DynamicSections parameter, see Appendix C, “Updating the DynamicSections Parameter of a DB2 Database” on page 347.

Microsoft SQL Server Database Requirements

Use the following guidelines when you set up the repository:

- Specify the database schema name when you use Microsoft SQL Server as the Model Repository database.
- Set the allow snapshot isolation and read committed isolation level to ALLOW_SNAPSHOT_ISOLATION and READ_COMMITTED_SNAPSHOT to minimize locking contention.
  To set the isolation level for the database, run the following commands:

  ```
  ALTER DATABASE DatabaseName SET ALLOW_SNAPSHOT_ISOLATION ON
  ```

  ```
  ALTER DATABASE DatabaseName SET READ_COMMITTED_SNAPSHOT ON
  ```

  To verify that the isolation level for the database is correct, run the following commands:

  ```
  SELECT snapshot_isolation_state FROM sys.databases WHERE name=[DatabaseName]
  ```

  ```
  SELECT is_read_committed_snapshot_on FROM sys.databases WHERE name = DatabaseName
  ```
- The database user account must have the CONNECT, CREATE TABLE, and CREATE VIEW privileges.

Oracle Database Requirements

Use the following guidelines when you set up the repository on Oracle:

- Set the OPEN_CURSORS parameter to 4000 or higher.
  Verify that the database user has the following privileges:

  CREATE SEQUENCE

  CREATE SESSION

  CREATE SYNONYM

  CREATE TABLE

  CREATE VIEW

- Informatica does not support Oracle public synonyms for repository tables. Verify that public synonyms have not been created for any tables in the database.

Monitoring Model Repository Service

The monitoring Model Repository Service is a Model Repository Service that monitors statistics for Data Integration Service jobs. You configure the monitoring Model Repository Service in the domain properties.

The following table summarizes the dependencies for products, services, and databases that are associated with the monitoring Model Repository Service:

Dependency	Summary
Products	The following products use the monitoring Model Repository Service: - Big Data Management - Big Data Parsing - Big Data Quality - Big Data Streaming
Services	The monitoring Model Repository Service properties do not require an association with another application service.
Databases	The monitoring Model Repository Service uses the following database: - Model repository. Stores run-time monitoring statistics that you can view in the Administrator tool.
Installer	You can create the monitoring Model Repository Service when you run the installer.

Rules and Guidelines

Consider the following rules and guidelines for the monitoring Model Repository Service:

- If you want to generate monitoring statistics, you must create a dedicated Model Repository Service for monitoring. You cannot store run-time monitoring statistics in the same repository where you store object metadata.

Search Service

The Search Service manages searches in the Analyst tool and returns search results from the Model repository. When you create the service, you need to associate another application service with it.

The following table summarizes the dependencies for products, services, and databases that are associated with the Search Service:

Dependency	Summary
Products	The following products use the Search Service: - Big Data Management - Big Data Quality - Big Data Streaming - Enterprise Data Catalog
Services	The Search Service requires a direct association with the following service: - Model Repository Service
Databases	The Search Service is not associated with any database.
Installer	You cannot create the Search Service when you run the installer.

Prepare to Create the Enterprise Data Lake Services

To create the Enterprise Data Lake services, the domain must be integrated with the Hadoop environment through a domain cluster configuration object.

The Enterprise Data Lake services require connections to the Hadoop environment. The connections are associated with the Hadoop environment through a cluster configuration. The process to integrate the environments and create the services can vary based on the type of installation you choose.

Install Enterprise Data Lake with Informatica Domain Services

If you install the Informatica domain services when you install Enterprise Data Lake and you want to create the Enterprise Data Lake services, you must provide the cluster information during the installation. The installer can import the cluster configuration from the Hadoop environment, and create the connections required by the Enterprise Data Lake services.

Before you run the installer, you need the information you need to import the cluster configuration from the Hadoop administrator. The Hadoop administrator can provide the information to you in one of the following formats:

- Cluster authentication information. The Hadoop administrator can provide you with cluster authentication information to connect to the cluster for the import process.
- Archive file. The Hadoop administrator can provide you an archive file that contains properties from *-site.xml files on the cluster. If you are importing from Amazon EMR or MapR, you can import only from an archive file.

Note: When the installation completes, you must fully integrate the domain with the Hadoop environment, including a task to refresh the cluster configuration. If you want to complete all integration tasks at one time, you can skip creating the services during installation and create them manually after you integrate the domain with the Hadoop environment.

Install Enterprise Data Lake and Enterprise Data Catalog on an Existing Node

When you install Enterprise Data Lake and Enterprise Data Catalog on a domain node, the installer installs the service binaries. After you install the binaries, you run the installer to configure the Enterprise Data Lake and Enterprise Data Catalog services.

Before you configure the services, verify that the domain is integrated with the Hadoop environment and that the Hadoop, HDFS, and Hive connections are associated with the cluster configuration. If the cluster configuration does not exist, you can use the Administrator tool to create it, and then create the connections manually after you integrate the domain with the Hadoop environment.

For more information about integrating the domain with the Hadoop environment, see the *Informatica Big Data Management Integration Guide*.

Install Enterprise Data Lake on a Node with Enterprise Data Catalog

When you install Enterprise Data Lake on a node with Enterprise Data Catalog, you can choose to create the Enterprise Data Lake services.

Before you run the installer, verify that the domain is integrated with the Hadoop environment and that the Hadoop, HDFS, and Hive connections are associated with the cluster configuration. If the cluster configuration does not exist, you can use the Administrator tool to create it, and then create the connections manually after you integrate the domain with the Hadoop environment.

For more information about integrating the domain with the Hadoop environment, see the *Informatica Big Data Management Integration Guide*.

Prepare for Archive File Import with a Full Installation

If you want to create the Enterprise Data Lake services when you perform a full installation, you must import properties from the *-site.xml files into the domain. The Hadoop administrator might choose to provide you with a .zip or .tar archive file instead of with direct connection information.

If you are integrating with an Amazon EMR or MapR cluster, you must import the cluster configuration through an archive file.

Get an archive file that contains the following *-site.xml files from the cluster:

- core-site.xml
- hbase-site.xml. Required only if you access HBase sources and targets.
- hdfs-site.xml
- hive-site.xml
- mapred-site.xml or tez-site.xml. Include the mapred-site.xml file or the tez-site.xml file based on the Hive execution type used on the Hadoop cluster.
- yarn-site.xml

Note: Verify that the Hadoop administrator creates an archive file from all the listed *-site.xml files. Big Data Management might require them even though Enterprise Data Lake might not.

After creating the archive file, the Hadoop administrator needs to edit it for the following distributions:

Azure HDInsight

Edit the Hortonworks Data Platform (HDP) version string wherever it appears in the archive file. Search for the string `${hdp.version}` and replace all instances with the HDP version that HDInsight includes in the Hadoop distribution.

Hortonworks HDP

Edit the Hortonworks Data Platform (HDP) version string wherever it appears in the archive file. Search for the string `${hdp.version}` and replace all instances with the HDP version that Hortonworks includes in the Hadoop distribution.

Prepare for Direct Import with a Full Installation

If you want to create the Enterprise Data Lake services when you perform a full installation, you must import properties from the *-site.xml files into the domain. You can get connection information for the cluster or an archive file from the Hadoop administrator to import cluster configuration from the Hadoop cluster.

The following table describes information that you need from the Hadoop administrator to create the cluster configuration directly from the cluster:

Property	Description
Host	IP address of the cluster manager.
Port	Port of the cluster manager.
User ID	Cluster user ID.
Password	Password for the user.
Cluster name	Name of the cluster. Use the display name if the cluster manager manages multiple clusters. If you do not provide a cluster name, the installer imports information based on the default cluster. **Note:** To find the correct Cloudera cluster name when you have multiple clusters, the Hadoop administrator can add the string `/api/v8/clusters` to the URL and provide you with the name that appears in the browser tab.

Configure Native Connectivity on Service Machines

To establish native connectivity between an application service and a database, install the database client software for the database that you want to access.

Native drivers are packaged with the database server and client software. Configure connectivity on the machines that need to access the databases. To ensure compatibility between the application service and the database, install a client software that is compatible with the database version and use the appropriate database client libraries. For more information about configuring connectivity, see Appendix B, “Connecting to Databases” on page 327.

The Data Integration Service uses native connectivity to connect to different databases:

- Source and target databases. Reads data from source databases and writes data to target databases.
- Data object cache database. Stores the data object cache.
- Profiling source databases. Reads from relational source databases to run profiles against the sources.
- Profiling warehouse. Writes the profiling results to the profiling warehouse.
- Reference tables. Runs mappings to transfer data between the reference tables and the external data sources.

When the Data Integration Service runs on a single node or on primary and back-up nodes, install database client software and configure connectivity on the machines where the Data Integration Service runs. When

the Data Integration Service runs on a grid, install database client software and configure connectivity on each machine that represents a node with the compute role or a node with both the service and compute roles.

Install Database Client Software

You must install the database clients on the required machines based on the types of databases that the application services access.

To ensure compatibility between the application service and the database, use the appropriate database client libraries and install a client software that is compatible with the database version.

Install the following database client software based on the type of database that the application service accesses:

IBM DB2 Client Application Enabler (CAE)

Configure connectivity on the required machines by logging in to the machine as the user who starts Informatica services.

Microsoft SQL Server 2012 Native Client

Download the client from the following Microsoft website: http://www.microsoft.com/en-in/download/details.aspx?id=29065.

Oracle client

Install compatible versions of the Oracle client and Oracle database server. You must also install the same version of the Oracle client on all machines that require it. To verify compatibility, contact Oracle.

Sybase Open Client (OCS)

Install an Open Client version that is compatible with the Sybase ASE database server. You must also install the same version of Open Client on the machines hosting the Sybase ASE database and Informatica. To verify compatibility, contact Sybase.

Configure Database Client Environment Variables on Linux

Configure database client environment variables on the machines that run the Data Integration Service processes.

After you configure the database environment variables, you can test the connection to the database from the database client.

The following table lists the database environment variables you need to set:

Database	Environment Variable Name	Database Utility	Value
Oracle	ORACLE_HOME PATH	sqlplus	Set to: <*DatabasePath*> Add: <*DatabasePath*>/bin
IBM DB2	DB2DIR DB2INSTANCE PATH	db2connect	Set to: <*DatabasePath*> Set to: <*DB2InstanceName*> Add: <*DatabasePath*>/bin
Sybase ASE	SYBASE15 SYBASE_ASE SYBASE_OCS PATH	isql	Set to: <*DatabasePath*>/sybase<*version*> Set to: ${SYBASE15}/ASE-<*version*> Set to: ${SYBASE15}/OCS-<*version*> Add: ${SYBASE_ASE}/bin:${SYBASE_OCS}/bin:$PATH

CHAPTER 4

Prepare for Kerberos Authentication

This chapter includes the following topics:

Checklist to Prepare for Kerberos Authentication

This chapter contains tasks to perform if you want the installer to enable Kerberos during installation. Use this checklist to track tasks required to prepare for Kerberos authentication.

- ❑ Set up the Kerberos configuration file.
- ❑ Generate the service principal and keytab name file format.
- ❑ Review the SPN and keytab format text file.
- ❑ Create the SPN and keytab files.

Prepare for Kerberos Authentication Overview

You can configure the Informatica domain to use Kerberos network authentication to authenticate users, services, and nodes.

Kerberos is a network authentication protocol which uses tickets to authenticate access to services and nodes in a network. Kerberos uses a Key Distribution Center (KDC) to validate the identities of users and services and to grant tickets to authenticated user and service accounts. In the Kerberos protocol, users and services are known as principals. The KDC has a database of principals and their associated secret keys that are used as proof of identity. Kerberos can use an LDAP directory service as a principal database.

To use Kerberos authentication, you must install and run the Informatica domain on a network that uses Kerberos network authentication. Informatica can run on a network that uses Kerberos authentication with Microsoft Active Directory service as the principal database.

The Informatica domain requires keytab files to authenticate nodes and services in the domain without transmitting passwords over the network. The keytab files contain the service principal names (SPN) and associated encrypted keys. Create the keytab files before you create nodes and services in the Informatica domain.

Set Up the Kerberos Configuration File

Kerberos stores configuration information in a file named *krb5.conf*. Informatica requires specific properties set in the Kerberos configuration file so that the Informatica domain can use Kerberos authentication correctly. You must set the properties in the krb5.conf configuration file.

The configuration file contains the information about the Kerberos server, including the Kerberos realm and the address of the KDC. You can request the Kerberos administrator to set the properties in the configuration file and send you a copy of the file.

1. Back up the krb5.conf file before you make any changes.
2. Edit the krb5.conf file.
3. In the *libdefaults* section, set or add the properties required by Informatica.

   The following table lists the values to which you must set properties In the libdefaults section:

Parameter	Value
default_realm	Name of the service realm for the Informatica domain.
forwardable	Allows a service to delegate client user credentials to another service. Set this parameter to True. The Informatica domain requires application services to authenticate the client user credentials with other services.
default_tkt_enctypes	Encryption types for the session key in ticket-granting tickets (TGT). Set this parameter only if session keys must use specific encryption types.
udp_preference_limit	Determines the protocol that Kerberos uses when it sends a message to the KDC. Set udp_preference_limit = 1 to always use TCP. The Informatica domain supports only the TCP protocol. If the udp_preference_limit is set to any other value, the Informatica domain can shut down unexpectedly.

4. In the *realms* section, include the port number in the address of the KDC separated by a colon.

   For example, if the KDC address is kerberos.example.com and the port number is 88, set the *kdc* parameter to the following:

   ```
   kdc = kerberos.example.com:88
   ```

5. Save the krb5.conf file.
6. Store the krb5.conf file in a directory that is accessible to the machine where you plan to install the Informatica services.

The following example shows the content of a krb5.conf with the required properties:

```
[libdefaults]
default_realm = AFNIKRB.AFNIDEV.COM
forwardable = true
udp_preference_limit = 1

[realms]
AFNIKRB.AFNIDEV.COM = {
  admin_server = SMPLKERDC01.AFNIKRB.AFNIDEV.COM
  kdc = SMPLKERDC01.AFNIKRB.AFNIDEV.COM:88
 }

[domain_realm]
afnikrb.afnidev.com = AFNIKRB.AFNIDEV.COM
.afnikrb.afnidev.com = AFNIKRB.AFNIDEV.COM
```

For more information about the Kerberos configuration file, see the Kerberos network authentication documentation.

Generate the Service Principal and Keytab File Name Format

If you run the Informatica domain with Kerberos authentication, you must associate Kerberos service principal names (SPN) and keytab files with the nodes and processes in the Informatica domain. Informatica requires keytab files to authenticate services without requests for passwords.

Based on the security requirements for the domain, you can set the service principal level to one of the following levels:

Node Level

If the domain is used for testing or development and does not require a high level of security, you can set the service principal at the node level. You can use one SPN and keytab file for the node and all the service processes on the node. You must also set up a separate SPN and keytab file for the HTTP processes on node.

Process Level

If the domain is used for production and requires a high level of security, you can set the service principal at the process level. Create a unique SPN and keytab file for each node and each process on the node. You must also set up a separate SPN and keytab file for the HTTP processes on node.

The Informatica domain requires the service principal and keytab file names to follow a specific format. To ensure that you follow the correct format for the service principal and keytab file names, use the Informatica Kerberos SPN Format Generator to generate a list of the service principal and keytab file names in the format required by the Informatica domain.

The Informatica Kerberos SPN Format Generator is shipped with the Informatica services installer.

Service Principal Requirements at Node Level

If the Informatica domain does not require a high level of security, the node and service processes can share the same SPNs and keytab files. The domain does not require a separate SPN for each service process in a node.

The Informatica domain requires SPNs and keytab files for the following components at node level:

Principal distinguished name (DN) for the LDAP directory service

Principal name for the bind user DN that is used to search the LDAP directory service. The name of the keytab file must be `infa_ldapuser.keytab`.

Node process

Principal name for the Informatica node that initiates or accepts authentication calls. The same principal name is used to authenticate the services in the node. Each gateway node in the domain requires a separate principal name.

HTTP processes in the domain

Principal name for all web application services in the Informatica domain, including Informatica Administrator. The browser uses this principal name to authenticate with all HTTP processes in the domain. The name of the keytab file must be `webapp_http.keytab`.

Service Principal Requirements at Process Level

If the Informatica domain requires a high level of security, create a separate SPN and keytab file for each node and each service in the node.

The Informatica domain requires SPNs and keytab files for the following components at process level:

Principal distinguished name (DN) for the LDAP directory service

Principal name for the bind user DN that is used to search the LDAP directory service. The name of the keytab file must be `infa_ldapuser.keytab`.

Node process

Principal name for the Informatica node that initiates or accepts authentication calls.

Informatica Administrator service

Principal name for the Informatica Administrator service that authenticates the service with other services in the Informatica domain. The name of the keytab file must be*_AdminConsole.keytab*.

HTTP processes in the domain

Principal name for all web application services in the Informatica domain, including Informatica Administrator. The browser uses this principal name to authenticate with all HTTP processes in the domain. The name of the keytab file must be `webapp_http.keytab`.

Service process

Principal name for the service that runs on a node in the Informatica domain. Each service requires a unique service principal and keytab file name.

You do not need to create the SPNs and keytab files for the services before you run the installer. You can create the SPN and keytab file for a service when you create the service in the domain. The SPN and keytab file for a service must be available when you enable the service.

Running the SPN Format Generator on Linux

You can run the Informatica Kerberos SPN Format Generator to generate a file that shows the correct format for the SPNs and keytab file names required in the Informatica domain.

You can run the SPN Format Generator from the command line or from the Informatica installer. The SPN Format Generator generates a file with the names of the service principal and keytab files based on the parameters you provide.

Note: Verify that the information you provide is correct. The SPN Format Generator does not validate the values you enter.

1. On the machine where you extracted the installation files, go to the following directory: `<Informatica installation files directory>/Server/Kerberos`
2. On a shell command line, run the SPNFormatGenerator.sh file.
3. Press **Enter** to continue.
4. In the **Service Principal Level** section, select the level at which to set the Kerberos service principals for the domain.

   The following table describes the levels you can select:

Level	Description
Process Level	Configures the domain to use a unique service principal name (SPN) and keytab file for each node and each application service on a node. The number of SPNs and keytab files required for each node depends on the number of application service processes that run on the node. Use the process level option for domains that require a high level of security, such as productions domains.
Node Level	Configures the domain to share SPNs and keytab files on a node. This option requires one SPN and keytab file for the node and all application services that run on the node. It also requires a separate SPN and keytab file for all HTTP processes on the node. Use the node level option for domains that do not require a high level of security, such as test and development domains.

5. Enter the domain and node parameters required to generate the SPN format.

   The following table describes the parameters you must specify:

Prompt	Description
Domain Name	Name of the domain. The name must not exceed 128 characters and must be 7-bit ASCII only. It cannot contain a space or any of the following characters: ` % * + ; " ? , < > \ /
Node name	Name of the Informatica node.
Node host name	Fully qualified host name or the IP address of the machine on which to create the node. The node host name cannot contain the underscore (_) character. **Note:** Do not use *localhost*. The host name must explicitly identify the machine.
Service Realm Name	Name of the Kerberos realm for the Informatica domain services. The realm name must be in uppercase.

   If you set the service principal at node level, the prompt **Add Node?** appears. If you set the service principal at process level, the prompt **Add Service?** appears.

6. At the **Add Node?** prompt, enter 1 to generate the SPN format for an additional node. Then enter the node name and node host name.

   To generate the SPN formats for multiple nodes, enter 1 at each **Add Node?** prompt and enter a node name and node host name.

7. At the **Add Service?** prompt, enter 1 to generate the SPN format for a service that will run on the preceding node. Then enter the service name.

   To generate the SPN formats for multiple services, enter 1 at each **Add Service?** prompt and enter a service name.

8. Enter 2 to end the **Add Service?** or **Add Node?** prompts.

   The SPN Format Generator displays the path and file name of the file that contains the list of service principal and keytab file names.

9. Press Enter to exit the SPN Format Generator.

The SPN Format Generator generates a text file that contains the SPN and keytab file names in the format required for the Informatica domain.

Review the SPN and Keytab Format Text File

The Kerberos SPN Format Generator generates a text file named SPNKeytabFormat.txt that lists the format for the service principal and keytab file names required by the Informatica domain. The list includes the SPN and keytab file names based on the service principal level you select.

Review the text file and verify that there are no error messages.

The text file contains the following information:

Entity Name

Identifies the node or service associated with the process.

SPN

Format for the SPN in the Kerberos principal database. The SPN is case sensitive. Each type of SPN has a different format.

An SPN can have one of the following formats:

Keytab type	SPN Format
NODE_SPN	isp/<*NodeName*>/<*DomainName*>@<*REALMNAME*>
NODE_AC_SPN	_AdminConsole/<*NodeName*>/<*DomainName*>@<*REALMNAME*>
NODE_HTTP_SPN	HTTP/<*NodeHostName*>@<*REALMNAME*> **Note:** The Kerberos SPN Format Generator validates the node host name. If the node host name is not valid, the utility does not generate an SPN. Instead, it displays the following message: Unable to resolve host name.
SERVICE_PROCESS_SPN	<*ServiceName*>/<*NodeName*>/<*DomainName*>@<*REALMNAME*>

Keytab File Name

Format for the name of the keytab file to be created for the associated SPN in the Kerberos principal database. The keytab file name is case sensitive.

The keytab file names use the following formats:

Keytab type	Keytab File Name
NODE_SPN	<*NodeName*>.keytab
NODE_AC_SPN	_AdminConsole.keytab
NODE_HTTP_SPN	webapp_http.keytab
SERVICE_PROCESS_SPN	<*ServiceName*>.keytab

Keytab Type

Type of the keytab. The keytab type can be one of the following types:

- NODE_SPN. Keytab file for a node process.
- NODE_AC_SPN. Keytab file for the Informatica Administrator service process.
- NODE_HTTP_SPN. Keytab file for HTTP processes in a node.
- SERVICE_PROCESS_SPN. Keytab file for a service process.

Service Principals at Node Level

The following example shows the contents of the SPNKeytabFormat.txt file generated for service principals at the node level:

```
ENTITY_NAME   SPN                                              KEY_TAB_NAME
KEY_TAB_TYPE
Node01        isp/Node01/InfaDomain@MY.SVCREALM.COM            Node01.keytab
NODE_SPN
Node01        HTTP/NodeHost01.enterprise.com@MY.SVCREALM.COM   webapp_http.keytab
NODE_HTTP_SPN
Node02        isp/Node02/InfaDomain@MY.SVCREALM.COM            Node02.keytab
NODE_SPN
Node02        HTTP/NodeHost02.enterprise.com@MY.SVCREALM.COM   webapp_http.keytab
NODE_HTTP_SPN
Node03        isp/Node03/InfaDomain@MY.SVCREALM.COM            Node03.keytab
NODE_SPN
Node03        HTTP/NodeHost03.enterprise.com@MY.SVCREALM.COM   webapp_http.keytab
NODE_HTTP_SPN
```

Service Principals at Process Level

The following example shows the contents of the SPNKeytabFormat.txt file generated for service principals at the process level:

```
ENTITY_NAME         SPN
KEY_TAB_NAME            KEY_TAB_TYPE
Node01              isp/Node01/InfaDomain@MY.SVCREALM.COM
Node01.keytab           NODE_SPN
Node01              _AdminConsole/Node01/InfaDomain@MY.SVCREALM.COM
_AdminConsole.keytab    NODE_AC_SPN
Node01              HTTP/NodeHost01.enterprise.com@MY.SVCREALM.COM
webapp_http.keytab      NODE_HTTP_SPN
Node02              isp/Node02/InfaDomain@MY.SVCREALM.COM
Node02.keytab           NODE_SPN
Node02              _AdminConsole/Node02/InfaDomain@MY.SVCREALM.COM
_AdminConsole.keytab    NODE_AC_SPN
Node02              HTTP/NodeHost02.enterprise.com@MY.SVCREALM.COM
webapp_http.keytab      NODE_HTTP_SPN
Service10:Node01    Service10/Node01/InfaDomain@MY.SVCREALM.COM
Service10.keytab        SERVICE_PROCESS_SPN
Service100:Node02   Service100/Node02/InfaDomain@MY.SVCREALM.COM
Service100.keytab       SERVICE_PROCESS_SPN
```

```
Service200:Node02  Service200/Node02/InfaDomain@MY.SVCREALM.COM
Service200.keytab      SERVICE_PROCESS_SPN
```

Create the Service Principal Names and Keytab Files

After you generate the list of SPN and keytab file names in Informatica format, send a request to the Kerberos administrator to add the SPNs to the Kerberos principal database and create the keytab files.

Use the following guidelines when you create the SPN and keytab files:

The user principal name (UPN) must be the same as the SPN.

When you create a user account for the service principal, you must set the UPN with the same name as the SPN. The application services in the Informatica domain can act as a service or a client depending on the operation. You must configure the service principal to be identifiable by the same UPN and SPN.

A user account must be associated with only one SPN. Do not set multiple SPNs for one user account.

Enable delegation in Microsoft Active Directory.

You must enable delegation for all user accounts with service principals used in the Informatica domain. In the Microsoft Active Directory Service, set the **Trust this user for delegation to any service (Kerberos only)** option for each user account that you set an SPN.

Delegated authentication happens when a user is authenticated with one service and that service uses the credentials of the authenticated user to connect to another service. Because services in the Informatica domain need to connect to other services to complete an operation, the Informatica domain requires the delegation option to be enabled in Microsoft Active Directory.

Use the ktpass utility to create the service principal keytab files.

Microsoft Active Directory supplies the ktpass utility to create keytab files. Informatica supports Kerberos authentication only on Microsoft Active Directory and has certified only keytab files that are created with ktpass.

The keytab files for a node must be available on the machine that hosts the node. By default, the keytab files are stored in the following directory: `<Informatica installation directory>/isp/config/keys`. During installation, you can specify a directory on the node to store the keytab files.

When you receive the keytab files from the Kerberos administrator, copy the keytab files to a directory that is accessible to the machine where you plan to install the Informatica services. When you run the Informatica installer, specify the location of the keytab files. The Informatica installer copies the keytab files to the directory for keytab files on the Informatica node.

Troubleshooting the Service Principal Names and Keytab Files

You can use Kerberos utilities to verify that the service principal and keytab file names created by the Kerberos administrator match the service principal and keytab file names that you requested. You can also use the utilities to determine the status of the Kerberos key distribution center (KDC).

You can use Kerberos utilities such as *setspn*, *kinit* and *klist* to view and verify the SPNs and keytab files. To use the utilities, ensure that the KRB5_CONFIG environment variable contains the path and file name of the Kerberos configuration file.

Note: The following examples show ways to use the Kerberos utilities to verify that SPNs and keytab files are valid. The examples might be different than the way that the Kerberos administrator uses the utilities to

create the SPNs and keytab files required for the Informatica domain. For more information about running the Kerberos utilities, see the Kerberos documentation.

Use the following utilities to verify the SPNs and keytab files:

klist

You can use *klist* to list the Kerberos principals and keys in a keytab file. To list the keys in the keytab file and the time stamp for the keytab entry, run the following command:

```
klist -k -t <keytab_file>
```

The following output example shows the principals in a keytab file:

```
Keytab name: FILE:int_srvc01.keytab
KVNO Timestamp         Principal
---- ----------------- --------------------------------------------------------
   3 12/31/69 19:00:00 int_srvc01/node01_vMPE/Domn96_vMPE@REALM
   3 12/31/69 19:00:00 int_srvc01/node01_vMPE/Domn96_vMPE@REALM
   3 12/31/69 19:00:00 int_srvc01/node01_vMPE/Domn96_vMPE@REALM
   3 12/31/69 19:00:00 int_srvc01/node01_vMPE/Domn96_vMPE@REALM
   3 12/31/69 19:00:00 int_srvc01/node01_vMPE/Domn96_vMPE@REALM
```

kinit

You can use *kinit* to request a ticket-granting ticket for a user account to verify that the KDC is running and can grant tickets. To request a ticket-granting ticket for a user account, run the following command:

```
kinit <user_account>
```

You can also use *kinit* to request a ticket-granting ticket and verify that the keytab file can be used to establish a Kerberos connection. To request a ticket-granting ticket for an SPN, run the following command:

```
kinit -V -k -t <keytab_file> <SPN>
```

The following output example shows the ticket-granting ticket created in the default cache for a specified keytab file and SPN:

```
Using default cache: /tmp/krb5cc_10000073
Using principal: int_srvc01/node01_vMPE/Domn96_vMPE@REALM
Using keytab: int_srvc01.keytab
Authenticated to Kerberos v5
```

setspn

You can use *setspn* to view, modify, or delete the SPN of an Active Directory service account. On the machine that hosts the Active Directory service, open a command line window and run the command.

To view the SPNs that are associated with a user account, run the following command:

```
setspn -L <user_account>
```

The following output example shows the SPN associated with the user account `is96svc`:

```
Registered ServicePrincipalNames for CN=is96svc,OU=AllSvcAccts,OU=People,
DC=ds,DC=intrac0rp,DC=zec0rp:
       int_srvc01/node02_vMPE/Domn96_vMPE
```

To view the user accounts associated with an SPN, run the following command:

```
setspn -Q <SPN>
```

The following output example shows the user account associated with the SPN `int_srvc01/node02_vMPE/Domn96_vMPE`:

```
Checking domain DC=ds,DC=intrac0rp,DC=zec0rp
CN=is96svc,OU=AllSvcAccts,OU=People,DC=ds,DC=intrac0rp,DC=zec0rp
        int_srvc01/node02_vMPE/Domn96_vMPE

Existing SPN found!
```

To search for duplicate SPNs, run the following command:

```
setspn -X
```

The following output example shows multiple user accounts associated with one SPN:

```
Checking domain DC=ds,DC=intrac0rp,DC=zec0rp
Processing entry 1125
HOST/mtb01.REALM is registered on these accounts:
        CN=Team1svc,OU=AllSvcAccts,OU=People,DC=ds,DC=intrac0rp,DC=zec0rp
        CN=MTB1svc,OU=IIS,OU=WPC960K3,OU=WINServers,DC=ds,DC=intrac0rp,DC=zec0rp
```

Note: Searching for duplicate SPNs can take a long time and a large amount of memory.

kdestroy

You can use *kdestroy* to delete the active Kerberos authorization tickets and the user credentials cache that contains them. If you run *kdestroy* without parameters, you delete the default credentials cache.

CHAPTER 5

Prepare for the Enterprise Data Catalog Cluster

This chapter includes the following topics:

Checklist to Prepare for Enterprise Data Catalog Cluster

This chapter contains tasks to complete based on the type of cluster you want to use with Enterprise Data Catalog. Use this checklist to track tasks associated with an external cluster or an embedded cluster.

❑ Embedded cluster that is installed with the services.

- Complete prerequisites and prepare the environment.
- Understand the Informatica Cluster Service that manages the embedded cluster services.
- Understand embedded cluster node management.

❑ External cluster that resides outside the Informatica domain.

- Complete prerequisites and prepare the environment.
- Set up Kerberos and SSL for the cluster.

Prepare for the Enterprise Data Catalog Cluster Overview

When you prepare for installation of Enterprise Data Catalog, you must prepare for a cluster external to the domain or embedded on the same machine as the domain.

If you plan to use any of the other Big Data Products, you must use an external cluster. If you choose to use an embedded cluster, the installer creates the cluster and the Informatica Cluster Service that manages the Hadoop services on the embedded cluster.

Embedded Cluster Prerequisites

Before you install Enterprise Data Catalog on an embedded Hadoop cluster, you must verify that the system environment meets the prerequisites required to deploy Enterprise Data Catalog.

Operating System Prerequisites

See the following sections for the operating system prerequisites on cluster nodes before you install Enterprise Data Catalog on an embedded cluster:

- "Common Operating System Prerequisites for Red Hat Enterprise Linux and SUSE Linux Enterprise Server " on page 77
- "Operating System Prerequisites for Red Hat Enterprise Linux" on page 79
- "Operating System Prerequisites for SUSE Linux Enterprise Server " on page 80

Common Operating System Prerequisites for Red Hat Enterprise Linux and SUSE Linux Enterprise Server

You can install Enterprise Data Catalog on a machine that runs on Red Hat Enterprise Linux Server or SUSE Linux Enterprise Server.

Verify the following common prerequisites for both SUSE and Red Hat Enterprise Linux Servers:

- Ensure that the operating system is 64-bit.
- Ensure that Bash is the default shell.
- If you use a user account without root privileges and if you want to remove sudo access, ensure that `defaults requiretty` is commented in `/etc/sudoers`.
- If you use a user account without root privileges, make sure that the Hadoop user has sudo privileges.
- Make sure that you disable the password prompt for the Hadoop user.
- Make sure that you install python-devel.
- Make sure that you set UMASK to 022 (0022) or 027 (0027).
- Ensure that you set Fully Qualified Domain Name (FQDN) for hostname -f.
- Make sure that the `/var` location does not have write privilege for everyone.
- Ensure that you configure the Linux base repositories.
- Make sure that you install the netstat command line network utility tool.

- Verify that the root directory (/) has a minimum of 10 GB of free disk space.
- If you plan to install Enterprise Data Catalog on an embedded cluster and you want to mount Informatica Cluster Service on a separate mount location, verify that the mount location has a minimum of 50 GB of free disk space.
- Make sure that the NOEXEC flag is not set for the file system mounted on the/tmp directory.
- You might want to ensure that the /tmp directory has a minimum of 20 GB of free disk space to enhance performance.
- Make sure that you install the scp, curl, unzip, wget and tar utilities.
- Ensure that you configure the home directory with write permission.
- Make sure that the configured /etc/hosts file of all machines include the FQDN for the machines.
- Ensure that the Network Time Protocol (NTP) daemon is synchronized and running.
- Make sure that the /tmp directory has chmod 777 permission configured for the directory.
- Make sure that the / and /var directories do not have chmod 777 permission configured.
- Make sure that the /var directory has a minimum of 2 GB of free disk space.
- Make sure that the /usr directory has a minimum of 2 GB of free disk space.
- Ensure that you disable Selinux.
- Make sure that the /etc/hosts has an entry for the loopback address, 127.0.0.1 localhost localhost.domain.com
- Make sure that you set the core limit to unlimited for a user without root privileges.
- If you configure the workingDir to /, validate if the file system mounted on /tmp and /var directories have the EXEC flag set.
- If the workingDir is not configured to /, validate if the workingDir directory has Read, Write, and Execute permissions configured. Validate if the EXEC flag is also set for the directory.
- Verify that you have the write permission on the /home directory. You can configure the permission in the /etc/default/useradd file.
- Make sure that each machine in the cluster includes the 127.0.0.1 localhost localhost.localdomain entry in the /etc/hosts file.
- Verify that the /etc/hosts file includes the fully-qualified host names for all the cluster nodes. Alternatively, make sure that reverse DNS lookup returns the fully-qualified host names for all the cluster nodes.
- Verify that the Linux repository includes postgresql version 8.14.18, release 1.el6_4 or later versions.
- Ensure that you set the soft limit for max user processes to 32000 or more.
- Ensure that you set the hard limit for max user processes to 32000 or more.
- On each host machine, verify that you have the following tools and applications available:
  - YUM and RPM (RHEL/CentOS/Oracle Linux)
  - Zypper
  - scp, curl, unzip, tar, and wget
  - awk
  - OpenSSL version 1.0.1e-30.el6_6.5.x86_64 or later.

  Note: Make sure that the $PATH variable points to the /usr/bin directory to use the correct version of Linux OpenSSL.

- For Enterprise Data Catalog installed on an embedded cluster, if you have not configured the Linux base repository or if you do not have an Internet connection, install the following packages:
  - The following RPMs on the Ambari Server host:
    - postgresql-libs
    - postgresql-server
    - postgresql

    -The following RPMs on all cluster nodes:
    - nc
    - redhat-lsb
    - psmisc
    - python-devel

Operating System Prerequisites for Red Hat Enterprise Linux

Verify the following prerequisites for a Red Hat Linux Enterprise Server if you plan to install Enterprise Data Catalog on a Red Hat Enterprise Linux Server:

Operating System	Prerequisite
Red Hat Enterprise Linux versions 6 and 7	- For Red Hat Enterprise Linux version 7.0, make sure that you use Sudo version 1.8.16 or later. - Install kernel-headers and kernel-devel. - Install libtirpc-devel. - Install openssl version v1.01 build 16 or later or v1.0.2k. - Install YUM. - Make sure that `/etc/sysconfig/network` directory exists and configure read permission for the directory. - `/etc/sysconfig/network` includes the same entry as the entry configured for hostname -f. - Install Python version 2.6.x or 2.7.x. Applicable for Red Hat Enterprise Linux version 6. - Install Python version 2.7.x. Applicable for Red Hat Enterprise Linux version 7. - Disable SSL certificate validation.

Operating System Prerequisites for SUSE Linux Enterprise Server

Verify the following prerequisites for a SUSE Linux Enterprise Server if you plan to install Enterprise Data Catalog on a SUSE Linux Enterprise Server:

Operating System	Prerequisite

SUSE Linux Enterprise Server versions 11 and 12	- Install netcat-openbsd.
	- Install kernel-default-devel.
	- Make sure that `/etc/HOSTNAME` directory exists and configure read permission for the directory.
	- Make sure that the `/etc/HOSTNAME` directory includes the same entry as the entry configured for hostname -f
	- Install Zypper.
	- Install the following versions of Python:
	  - 2.6.8/2.6.9/2.7.x for SUSE Linux version 11.
	  - 2.7.x for SUSE Linux version 12.
	- For SUSE Enterprise Linux Server 11, update all the hosts to Python version 2.6.8-0.15.1.
	- If you install Enterprise Data Catalog on SUSE Linux Enterprise Server 12, make sure that you install the following RPM Package Manager (RPMs) on all the cluster nodes:
	  - openssl-1.0.1c-2.1.3.x86_64.rpm
	  - libopenssl1_0_0-1.0.1c-2.1.3.x86_64.rpm
	  - libopenssl1_0_0-32bit-1.0.1c-2.1.3.x86_64.rpm
	  - python-devel-2.6.8-0.15.1.x86_64
	- Do not install libsnappy if you install Enterprise Data Catalog on SUSE Linux Enterprise Server.

Host Node Prerequisites

Verify the following prerequisites for host nodes:

- Validate that the passwordless ssh is enabled from the domain machine to Apache Ambari Server.
- Validate that the passwordless ssh is enabled from the Apache Ambari Server to all Apache Ambari agents.
- Validate that the number of Apache Ambari agents configured are one or more than two.

Prerequisites to Deploy Enterprise Data Catalog on Multiple Nodes

Verify the following prerequisites to deploy Enterprise Data Catalog on multiple nodes:

- Ensure that you use the same user credentials to start all nodes.
- Enable passwordless-SSH login for all host nodes to the cluster node.
- Make sure that you use the same version of Apache Ambari binaries on all nodes.
- Verify that you use the same resource binary files on all nodes.
- Ensure that you configure Informatica Cluster Service and Catalog Service on separate nodes.

Cluster Node Prerequisites

Verify that the cluster nodes meet the following requirements:

Node Type	Minimum Requirements
Master node	- The number of CPUs is 4. - Unused memory available for use is 12 GB. - Total memory available for use is 16 GB. - Disk space is 60 GB.
Slave node	- The number of CPUs is 4. - Unused memory available for use is 12 GB. - Total memory available for use is 16 GB. - Disk space is 60 GB.

Apache Ambari Prerequisites

Verify the following prerequisites for Apache Ambari:

- Apache Ambari requires certain ports that are open and available during the installation to communicate with the hosts that Apache Ambari deploys and manages. You need to temporarily disable the iptables to meet this requirement.
- Verify that you meet the memory and package requirements for Apache Ambari. For more information, see the Hortonworks documentation.

Apache Ranger Prerequisites

Before you deploy Enterprise Data Catalog on clusters where Apache Ranger is enabled, make sure that you configure the following permissions for the Informatica domain user:

- Write permission on the HDFS folder.
- Permission to submit applications to the YARN queue.

File Descriptor Limit

Verify that the maximum number of open file descriptors is 10,000 or more. Use the ulimit command to verify the current value and change the value if required.

SSL Prerequisites

If you want to enable SSL protocol for the cluster, verify the following prerequisites:

- If the cluster is enabled for SSL, ensure that you import the Ambari Server certificate to the Informatica domain truststore.
- If the cluster is enabled for SSL, it is recommended to enable SSL for the Informatica domain, the Informatica Cluster Service, and the Catalog Service.
- Ensure that you import the Hadoop cluster certificates to the domain trust store before you create a Catalog Service for a Hadoop cluster that uses SSL protocol.

- If you want to enable SSL authentication for Enterprise Data Catalog deployed on a multi-node Informatica domain, make sure that you complete the following prerequisites:
  - Export the Default.keystore of each node to the infa_truststore.jks on all nodes.
  - Make sure that the Default.keystore is unique for each host node.
  - Copy the Default.keystore to a unique location of each node.
  - If Informatica Cluster Service and Catalog Service are on different nodes, then export the Apache Ambari Server certificate to the infa_truststore.jks on all nodes.

Kerberos Prerequisites

If you want to enable Kerberos authentication for the cluster, verify the following prerequisites:

- Make sure that you merge the user and host keytab files before you enable Kerberos authentication for Informatica Cluster Service.
- Set the `udp_preference_limit` value to `1` in `$INFA_HOME/services/shared/security/krb5.conf`
- Make sure that the KDC certificate is present in the domain node.
- If you want to enable Kerberos authentication for Enterprise Data Catalog deployed on a multi-node Informatica domain, make sure that you complete the following prerequisites:
  - Make sure that all the domain nodes include the krb5.conf file in the following directories:
    - `$INFA_HOME/services/shared/security/`
    - `/etc/`
  - Make sure that the `/etc/hosts` file of all cluster nodes and domain nodes include the krb hosts entry and a host entry for other nodes.
  - Install krb5-workstation in all domain nodes.
  - Make sure that the keytab file is present in a common location on all domain nodes.
- Verify that you install the following prerequisite packages before you enable Kerberos for Enterprise Data Catalog on Red Hat Enterprise Linux:
  - krb5-workstation
  - krb5-libs
- For Enterprise Data Catalog on SUSE Linux Enterprise Server, make sure that you install the following prerequisite packages:
  - krb5-server
  - krb5-client

Informatica Cluster Service

The Informatica Cluster Service is an application service that runs and manages all the Hadoop services, Apache Ambari server, and Apache Ambari agents on an embedded Hadoop cluster. If you choose the embedded cluster deployment mode, you need to create the Informatica Cluster Service before you create the Catalog Service. Then, you can pass the Informatica Cluster Service value to the Catalog Service.

Informatica Cluster Service distributes the Hortonworks binaries and launches the required Hadoop services on the hosts where the embedded cluster runs.

You can deploy Informatica Cluster Service on hosts where Centrify is enabled. Centrify integrates with an existing Active Directory infrastructure to manage user authentication on remote Linux hosts.

Note: Informatica does not integrate with Centrify to manage or generate keytabs.

You can deploy Informatica Cluster Service on hosts that provide access using JSch SSH encryption algorithms.

The following table lists the supported methods and algorithms:

Method	Algorithm
Key exchange	- diffie-hellman-group-exchange-sha1 - diffie-hellman-group1-sha1 - diffie-hellman-group14-sha1 - diffie-hellman-group-exchange-sha256 - ecdh-sha2-nistp256 - ecdh-sha2-nistp384 - ecdh-sha2-nistp521
Cipher	- blowfish-cbc - 3des-cbc - aes128-cbc - aes192-cbc - aes256-cbc - aes128-ctr - aes192-ctr - aes256-ctr - 3des-ctr - arcfour - arcfour128 - arcfour256
MAC	- hmac-md5 - hmac-sha1 - hmac-md5-96 - hmac-sha1-96
Host key type	- ssh-dss - ssh-rsa - ecdsa-sha2-nistp256 - ecdsa-sha2-nistp384 - ecdsa-sha2-nistp521

Informatica Cluster Service Workflow

The Informatica Cluster Service is an ISP service that manages the embedded Hadoop cluster in Enterprise Data Catalog.

After Informatica Cluster Service is created, it performs the following actions:

1. Launches the Apache Ambari server and associated agents.
2. Creates Hadoop services and monitoring systems on Apache Ambari including HDFS, Apache Zookeeper, Yarn, and related monitoring services.
3. Starts the Hadoop services.
4. When you shut down Enterprise Data Catalog, the Informatica Cluster Service stops all the Hadoop services and stops the Apache Ambari server and its agents.

Creating an Informatica Cluster Service

You can choose to generate the Informatica Cluster Service when you install Enterprise Data Catalog or create the application service manually using Informatica Administrator.

If you plan to deploy Enterprise Data Catalog on multiple nodes, ensure that you configure Informatica Cluster Service and Catalog Service on separate nodes.

1. In the Administrator tool, select a domain, and click the **Services and Nodes** tab.
2. On the Actions menu, click **New** > **Informatica Cluster Service**.

   The **New Informatica Cluster Service: Step 1 of 4** dialog box appears.
3. Configure the general properties in the dialog box.

   The following table describes the properties:

Property	Description
Name	Name of the service. The name is not case-sensitive and must be unique within the domain. The name cannot exceed 128 characters or begin with @. The name cannot contain character spaces. The characters in the name must be compatible with the code page of the Model repository that you associate with the Catalog Service. The name cannot contain the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
Description	Description of the service. The description cannot exceed 765 characters.
Location	Domain in which the application service runs.
License	License to assign to the Informatica Cluster Service. Select the license that you installed with Enterprise Data Catalog.
Node	Node in the Informatica domain on which the Informatica Cluster Service runs. If you change the node, you must recycle the Informatica Cluster Service.
Backup Nodes	If your license includes high availability, nodes on which the service can run if the primary node is unavailable.

4. Click **Next**.

   The **New Informatica Cluster Service - Step 2 of 4** dialog box appears.
5. Configure the security properties in the dialog box.

   The following table describes the properties:

Property	Description
HTTP Port	A unique HTTP port number used for each Data Integration Service process. The defaults is 8085.
Enable Transport Layer Security (TLS)	Select the option to enable TLS for the Informatica Cluster Service.
HTTPS Port	Port number for the HTTPS connection. Required if you select **Enable Transport layer Security**.
Keystore File	Path and file name of the keystore file. The keystore file contains the keys and certificates required if you use the SSL security protocol with Catalog Administrator. Required if you select **Enable Transport layer Security**.

Property	Description
Keystore Password	Password for the keystore file. Required if you select **Enable Transport Layer Security**.
SSL Protocol	Secure Sockets Layer protocol to use.

6. Click **Next**.

   The **New Informatica Cluster Service - Step 3 of 4** dialog box appears.

7. Configure the Hadoop cluster properties in the dialog box.

   The following table describes the properties:

Property	Description
Hadoop Gateway Host	Host where Apache Ambari server runs.
Hadoop Gateway Port	Web port for the Apache Ambari server.
Gateway User	User name for the Apache Ambari server.
Hadoop Nodes	Hosts where the Apache Ambari agents run.
Override default password	Select this option if you want to change the default password for the cluster. Provide the new password in the **Ambari Server Admin Password** text box.
Enable Kerberos Authentication	Select the option to enable Kerberos authentication for the cluster.
KDC Type	Select one of the following Kerberos Key Distribution Center (KDC) types if you had selected the **Enable Kerberos Authentication** option: - Active Directory. Select this option if you want to use Active Directory KDC. - MIT KDC. Select this option if you want to use MIT KDC. Specify the following options after you select the **KDC Type** - KDC Host. Name of the KDC host machine. - Administrator Server Host. The name of the administrator server machine that hosts the KDC server. - Realm. Name of the Kerberos realm on the machine that hosts the KDC server. - Administrator Principal. The Kerberos administrator principal. - Administrator Password. The Kerberos administrator password. - LDAP URL. This property applies to Microsoft Active Directory and represents the URL to the LDAP server directory. - Container DN. This property applies to Microsoft Active Directory and represents the Distinguished Name of the container to which the user belongs. - KDC Certificate Path. Path to the KDC certificate on the Informatica domain machine.

8. Click **Next**.

   The **New Informatica Cluster Service - Step 4 of 4** dialog box appears.

9. Configure the domain security options for Informatica Cluster Service.

The following table describes the properties:

Property	Description
Domain is SSL Enabled	Specify if the Informatica domain is enabled for SSL.
Domain Truststore File Location	Location to the domain truststore file.
Domain Truststore Password	Password for the domain truststore file.
Enable Service	Select the option to enable the Informatica Cluster Service immediately after you create the service.

10. Click **Finish**.

    Note: After you update the Informatica Cluster Service security options in Informatica Administrator, restart the Informatica Cluster Service.

Before enabling the Informatica Cluster Service in a Kerberos-enabled cluster, verify the following prerequisites:

- You must configure the Key Distribution Center (KDC) hostname and IP address on all cluster nodes and domain machines in the `/etc/hosts`.
- Make sure that the krb5.conf file is located in all cluster nodes and domain machines under the `/etc` directory.
- For an SSL-enabled cluster or a Kerberos-enabled cluster, ensure that the domain truststore file is configured and copied to a common location accessible to all the cluster nodes.
- If the Solr keystore and password are different from the keystore and password of Informatica Cluster Service, you must export the public certificate of Solr to all the cluster nodes and import the certificate to the YARN truststore and domain truststore.

Preparing the Embedded Hadoop Cluster Environment

You need to perform multiple validation checks before you can before you install Enterprise Data Catalog on an embedded Hadoop cluster.

Perform the following steps before you install Enterprise Data Catalog on an embedded Hadoop cluster environment:

- Configure the `/etc/hosts` file on each machine so that you have fully qualified domain names. Informatica recommends the following host name format in lowercase: `<machine ipaddress> <fully qualified name> <alias>`.

  Note: To verify the configured host name, run the #hostname -f command.
- Set up passwordless Secure Shell (SSH) connections between the following components:

  - From Informatica Cluster Service to Hadoop Gateway.

  - From the Hadoop Gateway to Apache Hadoop nodes.
- Make sure that the `/etc/hosts` file on the machine that hosts Informatica domain includes entries for all Hadoop hosts.

Embedded Cluster Node Management

A Hadoop cluster has a set of machines that is configured to run Hadoop applications and services. A typical Hadoop cluster includes a master node and multiple slave or worker nodes. The master node runs the master daemons JobTracker and NameNode. A slave node runs the DataNode and TaskTracker daemons. In small clusters, the master node might also run the slave daemons.

Cluster with High Availability

You can use the highly availability option for the HDFS, HBase, YARN, and Solr components of the embedded Hadoop cluster environment. If you set up Informatica Cluster Service on a multi-node and highly available cluster, you need a minimum of three nodes for Enterprise Data Catalog to function successfully. If you have already set up Informatica Cluster Service on a single node, you cannot make the cluster highly available by adding more nodes to the cluster.

If the embedded cluster contains only three nodes, Enterprise Data Catalog distributes all master and slave services on all the three nodes. If the embedded cluster contains more than three nodes, Enterprise Data Catalog automatically chooses top three nodes with the highest system configuration as master nodes. The remaining nodes serve as slave nodes. When you add nodes to the embedded cluster, the newly added nodes serve as slave nodes. The nodes that you add to the cluster must meet the minimum configuration requirements for slave nodes.

Cluster without High Availability

You can set up Informatica Cluster Service on a single node that is not highly available. In such cases, the master and worker nodes remain on the same node. You cannot bring up Informatica Cluster Service if you add a single node to an existing single-node cluster or try to set up Informatica Cluster Service with two nodes.

Delete Nodes

You can delete nodes from the embedded cluster provided they meet the following conditions:

- You cannot delete a master node.
- You cannot delete a node if the number of live data nodes in the cluster becomes less than three on deleting the node.

Existing Cluster Prerequisites

Before you install Enterprise Data Catalog on an existing Hadoop cluster, you must verify that the system environment meets the prerequisites required to deploy Enterprise Data Catalog.

Host Node Prerequisites

On each host machine, verify that you have the zip and unzip utilities available.

Cluster Node Prerequisites

Verify the following prerequisites on all cluster nodes:

- OpenSSL version on the cluster nodes is openssl-1.0.1e-30.el6_6.5.x86_64 or later or v1.0.2k.
- Ensure that you install Java Development Kit (JDK) 1.8 on all cluster nodes and configure the JAVA_HOME environment variable on the cluster nodes to point to the JDK.

- Make sure that the max processes limit is set to 32000 or more.

Apache Ranger Prerequisites

Before you deploy Enterprise Data Catalog on clusters where Apache Ranger is enabled, make sure that the Informatica domain user has the required permission to submit applications to the YARN queue.

File Descriptor Limit

Verify that the maximum number of open file descriptors is 10,000 or more. Use the ulimit command to verify the current value and change the value if required.

SSL Prerequisites

If you want to enable SSL protocol for the cluster, verify the following prerequisites:

- When you create the Catalog Service that connects to an SSL-enabled existing cluster, verify that you configure the following properties:
  - A keytab file that contains all the users in LDAP.
  - Kerberos domain name.
  - HDFS namenode and YARN Resource Manager service principals
  - Path to Solr keystore file and password.
  - Import the Hadoop cluster certificates to the Informatica domain truststore.
- If the cluster is enabled for SSL, make sure that you enable SSL for the Informatica domain and the Catalog Service.
- If you create a new Solr keystore, make sure that you export the Solr public certificate to all cluster nodes.
- Make sure that you import the Informatica domain public certificates to all YARN truststores.
- If you want to enable SSL authentication for Enterprise Data Catalog deployed on a multi-node Informatica domain, make sure that you complete the following prerequisites:
  - Export the Default.keystore of each node to the infa_truststore.jks on all nodes.
  - Make sure that the Default.keystore is unique for each host node.
  - Copy the Default.keystore to a unique location of each node.
  - If Informatica Cluster Service and Catalog Service are on different nodes, then export the server certificate from your Hadoop distribution to the infa_truststore.jks on all nodes.

Kerberos Prerequisites

If you want to enable Kerberos for the cluster, verify the following prerequisites:

- Verify that you install the following prerequisite packages before you enable Kerberos:
  - krb5-workstation
  - krb5-libs
- If you want to collect the log files in a common location, create the `service-logs` directory under `/Informatica/LDM/<service cluster name>/` and assign the ownership of the directory to the service cluster user if the cluster is enabled for Kerberos.

Note: If the cluster is not enabled for Kerberos, create the `service-logs` directory under `/Informatica/LDM/<service cluster name>/` and assign the ownership of the directory to the domain user.

- If the cluster is not enabled for Kerberos. create the directory `<domain user name>` under `/user` and assign the ownership of directory to the domain user.

  Note: If the cluster is enabled for Kerberos, create the directory `<service cluster name>` under `/user` and assign the ownership of the directory to the service cluster user. If the cluster is not enabled for Kerberos, the assign the ownership of the directory to the domain user.

- If you want to enable Kerberos authentication for Enterprise Data Catalog deployed on a multi-node Informatica domain, make sure that you complete the following prerequisites:
  - Make sure that all the domain nodes include the `krb5.conf` file in the following directories:
    - `$INFA_HOME/services/shared/security/`
    - `/etc/`
  - Make sure that the `/etc/hosts` file of all cluster nodes and domain nodes include the krb hosts entry and a host entry for other nodes.
  - Install `krb5-workstation` in all domain nodes.
  - Make sure that the keytab file is present in a common location on all domain nodes.
  - Make sure that the service cluster user is configured on all cluster nodes.

Informatica Domain Prerequisites

Ensure that you do not create the Informatica domain on a node in the existing Hadoop cluster.

User Permissions

Configure the following permissions for user accounts if you plan to deploy Enterprise Data Catalog on an existing cluster:

- Make sure that you configure the Read, Write, and Execute permissions for owners, groups, and others on HDFS directories.
- If you plan to deploy Enterprise Data Catalog on a Cloudera Hadoop distribution, make sure that you assign the following roles in Cloudera Manager for the Cloudera user account that you use to start the Catalog Service:
  - Operator
  - Configurator
  - Cluster Administrator
  - Navigator Administrator
  - Full Administrator

Existing Hadoop Cluster Deployment

You can deploy Enterprise Data Catalog on a Hadoop cluster that you have set up on Cloudera, Hortonworks, or Azure HDInsight. If you have enabled Kerberos authentication in your enterprise to authenticate users and services on a network, you can configure the Informatica domain to use Kerberos network authentication.

You need to configure Zookeeper, HDFS, and Yarn specifications when you install Enterprise Data Catalog on an existing Hadoop cluster in your enterprise. The Catalog Service uses the following specifications and launches the following services and components on the Hadoop cluster as YARN application:

- Solr version 5.2.1
- HBase version 0.98
- Scanner components

Preparing the Existing Hadoop Cluster Environment

You need to perform multiple validation checks before you install Enterprise Data Catalog on an existing Hadoop cluster.

Perform the following steps before you install Enterprise Data Catalog to use an existing cluster.:

- Create the following directories in HDFS before you create the Catalog Service:

  - `/Informatica/LDM/<service cluster name>`

  - `/user/<user name>`

  Where `<service cluster name>` is the name of the service cluster that you need to enter when you create the Catalog Service and `<user name>` is the username of the Informatica domain user.
- Make `<user name>`who is the Informatica domain user is the owner of the `/Informatica/LDM/<service cluster name>` and `/user/<user name>` directories.

Kerberos and SSL Setup for an Existing Cluster

You can install Enterprise Data Catalog on an existing cluster that uses Kerberos network authentication to authenticate users and services on a network. Enterprise Data Catalog also supports SSL authentication for secure communication in the cluster.

Kerberos is a network authentication protocol which uses tickets to authenticate access to services and nodes in a network. Kerberos uses a Key Distribution Center (KDC) to validate the identities of users and services and to grant tickets to authenticated user and service accounts. In the Kerberos protocol, users and services are known as principals. The KDC has a database of principals and their associated secret keys that are used as proof of identity. Kerberos can use an LDAP directory service as a principal database.

Informatica does not support cross or multi-realm Kerberos authentication. The server host, client machines, and Kerberos authentication server must be in the same realm.

The Informatica domain requires keytab files to authenticate nodes and services in the domain without transmitting passwords over the network. The keytab files contain the service principal names (SPN) and

associated encrypted keys. Create the keytab files before you create nodes and services in the Informatica domain.

Prerequisites for SSL Authentication

Verify that the existing cluster meets the following requirements before you can enable SSL authentication in the cluster:

- Informatica domain is configured in the SSL mode.
- The cluster and YARN REST endpoints are Kerberos-enabled.
- Create a keystore file for the Apache Solr application on all nodes in the cluster. Import public certificates of Apache Solr keystore files on all the hosts into all the truststore files configured for HDFS and YARN. This step is required for Apache Spark and scanner jobs to connect to the Apache Solr application.
- Import the public certificates of Apache Solr and YARN applications into the truststore file of the Informatica domain. This step is required for Catalog Service to connect to YARN and Solr applications.
- Import the public certificates of Informatica domain and the Catalog Service into the YARN truststore.
- Import the public certificate of the Catalog Service into the Informatica domain truststore.
- If you plan to deploy Enterprise Data Catalog on an existing Hortonworks version 2.5 cluster that does not support SSL authentication, perform the following steps:
  1. Configure the following properties in the `/etc/hadoop/conf/ssl-client.xml` file: `ssl.client.truststore.location` and `ssl.client.truststore.password`.
  2. Ensure that the `ssl.client.truststore.location` value is set to `/opt` directory and not `/etc` directory. Verify that you configure the full path to the truststore file for the `ssl.client.truststore.location` property. For example, you can set the value similar to `/opt/truststore/infa_truststore.jks`.
  3. Export the keystore certificate used in the Informatica domain.
  4. Import the keystore certificate into the Informatica domain truststore file.
  5. Place the domain truststore file in all the Hadoop nodes in the `/opt` directory. For example, `/opt/truststore/infa_truststore.jks`.
  6. Open the `/etc/hadoop/conf/ssl-client.xml` file.
  7. Modify the `ssl.client.truststore.location` and `ssl.client.truststore.password` properties.

Prerequisites for Kerberos Authentication

Perform the following steps before you enable the Kerberos authentication for the existing cluster:

- Create the following users in the LDAP security domain where <user name> is the service cluster name.
  - <user name>@KERBEROSDOMAIN.COM
  - <user name>/<hostname>@KERBEROSDOMAIN.COM

    Note: Create the user ID for all the hosts in the cluster.
  - HTTP/<host name>@KERBEROSDOMAIN.COM

    Note: Create the user ID for all the hosts in the cluster.
  - Create a keytab file with credentials for all these users created in LDAP. You can create keytab files for each one of the users in KDC server and merge them using the ktutil command to create single keytab file.
  - Create the following folders in HDFS that Enterprise Data Catalog uses as data directories for the Catalog Service: `/Informatica/LDM/<user name>` and `/user/<user name>`.
  - Change the owner of these two folders to <user name>.

- Create a local user with username as <user name> on all the hosts in the cluster. This step is required to launch the application on YARN as the user configured for the Catalog Service.

- Set up the `udp_preference_limit` parameter in the `krb5.conf` Kerberos configuration file to 1. This parameter determines the protocol that Kerberos uses when it sends a message to the KDC. Set `udp_preference_limit = 1` to always use TCP. The Informatica domain supports only the TCP protocol. If the `udp_preference_limit` parameter is set to any other value, the Informatica domain might shut down unexpectedly.

Note: Enterprise Data Catalog does not support deployment on a Hortonworks version 2.6 cluster where Kerberos is enabled.

CHAPTER 6

Record Information for Installer Prompts

This chapter includes the following topics:

Checklist to Record Installer Prompts

This chapter contains information that you need to enter when you run the installer. Use this checklist to track the recording tasks before you run the installer.

- ❑ Record the names of nodes that you want to create and the services that you want to create on each node.
- ❑ Record basic database information for each database associated with a service that you are creating.
- ❑ If the domain configuration and Model repository databases are secure, record the JDBC connection string with required security parameters.
- ❑ Record the keyword for the installer to generate an encryption key for the domain.
- ❑ If you want to enable Kerberos authentication when you run the installer, record Kerberos information for each node in the domain.
- ❑ If you install Enterprise Data Catalog, record cluster information.

Record Information for Installer Prompts Overview

When you install the Informatica services, you need to know information about the domain, nodes, application services, and databases that you plan to create. If you plan to install the Informatica services on a network that uses Kerberos authentication, you also need to know information about the Kerberos authentication server.

This section lists information that you need to provide when you run the installer. Informatica recommends recording installer prompts before you start the installation process. For example, you might want to create a text file of information so you can copy into the installer.

Domain Object Naming Conventions

You cannot change domain, node, and application service names. Use names that continue to work if you migrate a node to another machine or if you add additional nodes and services to the domain. In addition, use names that convey how the domain object is used. Naming conventions are provided in applicable topics.

For more information about domain object naming conventions, see the following Informatica Velocity Best Practice article available on the Informatica Network: http://velocity.informatica.com/index.php/best-practices-all/139-configuration-management-and-security/708-infa-nam-conv.

Domain

When you create a domain, you must provide a domain name and gateway node name.

The following table describes the domain information that you need to enter during the installation process:

Domain Information	Description
Domain name	Name of the domain that you plan to create. The name must not exceed 128 characters and must be 7-bit ASCII only. It cannot contain a space or any of the following characters: ` % * + ; " ? , < > \ / Consider one of the following naming conventions: DMN, DOM, DOMAIN, _<ORG>_<ENV>
Master gateway node host name	Fully qualified host name of the machine on which to create the master gateway node. If the machine has a single network name, use the default host name. The node host name cannot contain the underscore (_) character. If the machine has multiple network names, you can modify the default host name to use an alternate network name. If the machine has a single network name, use the default host name. **Note:** Do not use localhost. The host name must explicitly identify the machine.
Master gateway node name	Name of the master gateway node that you plan to create on this machine. The node name is not the host name for the machine. Consider the following naming convention: Node<node##>_<ORG>_<optional distinguisher>_<ENV>

Nodes

When you install the Informatica services, you add the installation machine to the domain as a node. You can add multiple nodes to a domain.

The following table describes the node information that you need to enter when you join a domain.

Node Information	Description
Node host name	Fully qualified host name of the machine on which to create nodes. If the machine has a single network name, use the default host name. The node host name cannot contain the underscore (_) character. If the machine has multiple network names, you can modify the default host name to use an alternate network name. If the machine has a single network name, use the default host name. **Note:** Do not use localhost. The host name must explicitly identify the machine.
Node name	Name of the nodes that you plan to create on this machine. The node name is not the host name for the machine. Consider the following naming convention: Node<node##>_<ORG>_<optional distinguisher>_<ENV>

Application Services

Record the application service names and the nodes where you want to create them.

The following table lists the application services that you can create when you run the installer:

Application Service	Naming Convention
Catalog Service	CS_<ORG>_<ENV>
Content Management	CMS_<ORG>_<ENV>
Data Integration Service	DIS_<ORG>_<ENV>
Data Preparation Service	DPS_<ORG>_<ENV>
Enterprise Data Lake Service	EDLS_<ORG>_<ENV>
Informatica Cluster Service	ICS_<ORG>_<ENV>
Model Repository Service	MRS_<ORG>_<ENV>
monitoring Model Repository Service	mMRS_<ORG>_<ENV>

For more information about all service naming conventions, see the following Informatica Velocity Best Practice article available on the Informatica Network: http://velocity.informatica.com/index.php/best-practices-all/139-configuration-management-and-security/708-infa-nam-conv.

Important: If you plan to use Kerberos authentication, you must know the application service and node name before you create the keytab files.

Databases

When you plan the Informatica domain, you also need to plan the required relational databases. The domain requires a database to store configuration information and user account privileges and permissions. Some application services require databases to store information processed by the application service.

Domain

The following table describes the information that you need to enter during the installation process:

Database Information	Description
Domain configuration database type	Database type for the domain configuration repository. The domain configuration repository supports IBM DB2 UDB, Microsoft SQL Server, Oracle, or Sybase ASE.
Domain configuration database host name	The name of the machine hosting the database.

Content Management Service

The following table describes the information that you need to enter during the installation process:

Database Information	Description
Reference data warehouse database type	Database type for the reference data warehouse. The reference data warehouse supports IBM DB2 UDB, Microsoft SQL Server, or Oracle.
Reference data warehouse database host name	The name of the machine hosting the database.

Data Integration Service

The following table describes the information that you need to enter during the installation process:

Database Information	Description
Data object cache database type	Database type for the data object cache database. The data object cache database supports IBM DB2 UDB, Microsoft SQL Server, or Oracle.
Data object cache database host name	The name of the machine hosting the database.
Profiling warehouse database type	Database type for the profiling warehouse. The profiling warehouse supports IBM DB2 UDB, Microsoft SQL Server, or Oracle.
Profiling warehouse database host name	The name of the machine hosting the database.

Database Information	Description
Workflow database type	Database type for the workflow database. The workflow database supports IBM DB2 UDB, Microsoft SQL Server, or Oracle.
Workflow database host name	The name of the machine hosting the database.

Data Preparation Service

The following table describes the information that you need to enter during the installation process:

Database Information	Description
Data Preparation repository database type	Database type for the Data Preparation repository. The Data Preparation repository supports MySQL, MariaDB, or Oracle.
Data Preparation repository database host name	The name of the machine hosting the database.

Model Repository Service

The following table describes the information that you need to enter during the installation process:

Database Information	Description
Model repository database type	Database type for the Model repository. The Model repository supports IBM DB2 UDB, Microsoft SQL Server, or Oracle.
Model repository database host name	The name of the machine hosting the database.

Connection String to a Secure Database

If you create a repository on a secure database, you must provide the truststore information for the database and a JDBC connection string that includes the security parameters for the database.

During installation, you can create the domain configuration repository in a secure database. You can also create the Model repository in a secure database.

You can configure a secure connection to the following databases:

- IBM DB2 UDB
- Microsoft SQL Server
- Microsoft Azure SQL Database
- Oracle

Note: You cannot configure a secure connection to a Sybase database.

When you configure the connection to the secure database, you must specify the connection information in a JDBC connection string. In addition to the host name and port number for the database server, the connection string must include security parameters.

The following table describes the security parameters that you must include in the JDBC connection string:

Parameter	Description
EncryptionMethod	Required. Indicates whether data is encrypted when transmitted over the network. This parameter must be set to `SSL`.
ValidateServerCertificate	Optional. Indicates whether Informatica validates the certificate that is sent by the database server. If this parameter is set to True, Informatica validates the certificate that is sent by the database server. If you specify the HostNameInCertificate parameter, Informatica also validates the host name in the certificate. If this parameter is set to false, Informatica does not validate the certificate that is sent by the database server. Informatica ignores any truststore information that you specify.
HostNameInCertificate	Optional. Host name of the machine that hosts the secure database. If you specify a host name, Informatica validates the host name included in the connection string against the host name in the SSL certificate. If SSL encryption and validation is enabled and this property is not specified, the driver uses the server name specified in the connection URL or data source of the connection to validate the certificate.
cryptoProtocolVersion	Required. Specifies the cryptographic protocol to use to connect to a secure database. You can set the parameter to `cryptoProtocolVersion=TLSv1.1` or `cryptoProtocolVersion=TLSv1.2` based on the cryptographic protocol used by the database server.

You can use the following syntax in the JDBC connection string to connect to a secure database:

IBM DB2

```
jdbc:Informatica:db2://
host_name:port_no;DatabaseName=database_name;EncryptionMethod=SSL;HostNameInCertificate=D
B_host_name;ValidateServerCertificate=true_or_false
```

Oracle

```
jdbc:Informatica:oracle://
host_name:port_no;ServiceName=service_name;EncryptionMethod=SSL;HostNameInCertificate=DB_
host_name;ValidateServerCertificate=true_or_false
```

Microsoft SQL Server

```
jdbc:Informatica:sqlserver://
host_name:port_no;SelectMethod=cursor;DatabaseName=database_name;EncryptionMethod=SSL;Hos
tNameInCertificate=DB_host_name;ValidateServerCertificate=true_or_false
```

Microsoft Azure SQL Database

```
jdbc:Informatica:sqlserver://
host_name:port_no;SelectMethod=cursor;DatabaseName=database_name;SnapshotSerializable=tru
e;EncryptionMethod=SSL;HostNameInCertificate=*.database.windows.net;ValidateServerCertifi
cate=false
```

Note: The installer does not validate the connection string. Verify that the connection string contains all the connection parameters and security parameters required by your database.

Secure Data Storage

When you install the Informatica services, you must provide a keyword for the installer to use to generate the encryption key for the domain.

Use the following table to record the information that you need to configure secure data storage:

Property	Description
Keyword	Keyword to use to create a custom encryption key to secure sensitive data in the domain. The keyword must meet the following criteria: - From 8 to 20 characters long - Includes at least one uppercase letter - Includes at least one lowercase letter - Includes at least one number - Does not contain spaces The encryption key is created based on the keyword that you provide when you create the Informatica domain.
Encryption key directory	Directory in which to store the encryption key for the domain. By default, the encryption key is created in the following directory: `<Informatica installation directory>/isp/config/keys`.

Kerberos

When you install the Informatica application services, you can enable options in the Informatica domain to configure security for the domain, services and databases.

If you want to enable Kerberos authentication and you do not want to use the default file, you need to provide information such as keystore and truststore directories. Each node needs to contain a keystore and truststore that is used by all services on that node.

The following table describes security information to provide during installation:

Security Information	Description
Service realm name	Name of the Kerberos realm to which the Informatica domain services belong. The realm name must be in uppercase. The service realm name and the user realm name must be the same.
User realm name	Name of the Kerberos realm to which the Informatica domain users belong. The realm name must be in uppercase. The service realm name and the user realm name must be the same.
Location of the Kerberos configuration file	Directory where the Kerberos configuration file named *krb5.conf* is stored. Informatica requires specific properties to be set in the configuration file. If you do not have permission to copy or update the Kerberos configuration file, you might have to request the Kerberos administrator to update the file.
Keystore file directory	Directory that contains the keystore files. The directory must contain files named infa_keystore.jks and infa_keystore.pem.

Security Information	Description
Keystore password	A plain-text password for the keystore infa_keystore.jks.
Truststore file directory	Directory that contains the truststore files. The directory must contain files named infa_truststore.jks and infa_truststore.pem.
Truststore password	Password for the infa_truststore.jks file.

Cluster Information for Enterprise Data Catalog

When you install Enterprise Data Catalog, you must provide information about the cluster that Enterprise Data Catalog uses.

The following table describes cluster information to provide during installation:

Domain Information	Description
SSH username	User name for the password-less Secure Shell (SSH) connection.
Ambari server host	Host information for the Ambari server. Ambari is a web-based tool for provisioning, managing, and monitoring Apache Hadoop clusters, which includes support for Hadoop HDFS, Hadoop MapReduce, Hive, HBase and ZooKeeper.
Comma-separated Ambari agent hosts	Applies to high availability. If you use multiple Ambari agent hosts, specify the comma-separated values of multiple Ambari agent host names.
Ambari web port	Port number where the Ambari server needs to run.
Keytab Location	Applies to a Kerberos-enabled cluster. Location of the merged user and host keytab file.
Kerberos configuration file	Applies to a Kerberos-enabled cluster. Location of the Kerberos configuration.
YARN resource manager URI	The service within Hadoop that submits the MapReduce tasks to specific nodes in the cluster. Use the following format: <host name>:<port> - <name node> is the host name or IP address of YARN resource manager. - <port> is the port that the YARN resource manager for Remote Procedure Calls (RPC).
YARN resource manager http URI	The http URI value for the YARN resource manager.
YARN resource manager scheduler URI	Scheduler URI value for the YARN resource manager.
ZooKeeper URI	The URI for the ZooKeeper service, which is a high-performance coordination service for distributed applications.

Domain Information	Description
HDFS namenode URI	The URI to access HDFS. Use the following format to specify the NameNode URI in the Cloudera distribution: hdfs://<name node>:<port> - <name node> is the host name or IP address of the NameNode. - <port> is the port that the NameNode listens for Remote Procedure Calls (RPC).
Service cluster name	Name of the service cluster. Ensure that you have a directory `/Informatica/LDM/<ServiceClusterName>` in HDFS before the installation is complete. **Note:** If you do not specify a service cluster name, Enterprise Data Catalog considers DomainName_CatalogServiceName as the default value. You must then have the /Informatica/LDM/<DomainName>_<CatalogServiceName>
History Server HTTP URI	HTTP URI to access the history server.

Part III: Run the Big Data Suite Installer

This part contains the following chapters:

CHAPTER 7

Introduction to the Big Data Suite Installer

This chapter includes the following topics:

Big Data Suite Installer Tasks

The installer performs install tasks based on the product or products that you install.

The installer can perform the following tasks:

1. Perform pre-install validation and system check.
2. Create a domain or join a node to an existing domain.
3. Install binaries for service support.
4. Tune the services based on the Informatica deployment size.
5. Create application services.
6. Perform some integration tasks with the Hadoop cluster.
7. Configure security between the domain and services.
8. Start the domain and application services that you created.
9. Write message to the log file.

System Check Tool (i10Pi) and SPN Format Generator

Informatica provides utilities to facilitate the Informatica services installation process. You can use the Informatica installer to run the utilities.

Run the following utilities before you install Informatica services:

Pre-Installation (i10Pi) System Check Tool

The Pre-Installation (i10Pi) System Check Tool verifies whether a machine meets the system requirements for the Informatica installation. Informatica recommends that you verify the minimum system requirements before you start the installation. When you run the system check tool before you perform the installation, the installer sets fields for certain fields, such as the database connection and domain port numbers, based on the information that you enter during the system check.

Informatica Kerberos SPN Format Generator

The Informatica Kerberos SPN Format Generator generates a list of Kerberos service principal names (SPN) and keytab file names in the format required by Informatica. If you install Informatica on a network that uses Kerberos authentication, run this utility to generate the service principal and keytab file names in the informatica format. Then request the Kerberos administrator to add the SPNs to the Kerberos principal database and create the keytab files before you start the installation.

Secure Files and Directories

When you install or upgrade Informatica, the installer creates directories to store Informatica files that require restricted access, such as the domain encryption key file and the nodemeta.xml. The installer assigns different permissions for the directories and the files in the directories.

By default, the installer creates the following directories within the Informatica installation directory:

<Informatica installation directory>/isp/config

Contains the nodemeta.xml file. Also contains the /keys directory where the encryption key file is stored. If you configure the domain to use Kerberos authentication, the /keys directory also contains the Kerberos keytab files. You can specify a different directory in which to store the files. The installer assigns the same permissions to the specified directory as the default directory.

<Informatica installation directory>/services/shared/security

If you enable secure communication for the domain, the /security directory contains the keystore and truststore files for the default SSL certificates.

To maintain the security of the directories and files, the installer restricts access to the directories and the files in the directories. The installer assigns specific permissions to the group and user account that own the directories and files.

For more information about permissions assigned to the directories and files, see the *Informatica Security Guide*.

Resume the Installer

When the installation process stops midway, you can resume the installation from the point of failure or exit.

When you run the server installer and the installation process fails, you can still resume from the previous service configuration and recover the last entered details for that service installation.

The install process might fail for reasons such as network outage, when you exit the installation before completing the entire installation process, or because of incorrect information entered.

If a service fails or if the installation process fails during a service creation, you can resume the installation process with the server installer. To resume the installation process, ensure that at least one of the services is created and that the domain is up and running from the installation log. For example, if you want to check whether the Model Repository Service is created, check if you have a service creation success text in the server log in the following format:

```
SUCCESS: MRS Service [mrs_name] is created. Command ran successfully.
```

To resume the installation, run the installer again.

Note: You cannot resume the installer if you are running it to configure services after the services have been created. When you run the service configuration wizard, you cannot resume the installer for Big Data, Enterprise Data Lake, or Enterprise Data Catalog. When you join the domain, you also cannot resume the installer.

When you resume the installer while creating a service, the installer retains all the service and database specific information, such as the create service status, service name, service enabled or disabled status. You can confirm and use the previously entered values or specify new values for the service and resume the installation process.

Resuming the Installer

When the installation process stops midway, you can resume the installation from the point of failure or exit.

Before you can resume the installer, complete the following prerequisites:

- In the installation log file present in the installation directory, verify that at least the domain and one service is created. The installer log file name appears in the following syntax: Informatica_<Version>_Services_<timestamp>.log
- Ensure that you do not delete the installInst.obj object file present in the tools folder of the user installation directory.
- For silent installer, ensure that RESUME_INSTALLATION is set to true in the SilentInput.properties file.

1. Open a command prompt and navigate to the location of the installation files.
2. Run the Installer.

   On Linux, run silentInstall.sh to resume the silent installer. To resume the regular installer, run the ./install.sh command.

3. When the regular installer runs, you might get a prompt confirming whether you want to resume previous installer or not.
   - If you do not want to resume installation, enter 1 for No. Default is 1.
   - If you want to resume installation, enter 2 for Yes.

   Before you can resume the installation, the services get validated.

CHAPTER 8

Create a Domain and Install All Big Data Products

This chapter includes the following topics:

Begin the Install

This task includes installer prompts to begin the installation. You will provide basic information such as acceptance of terms, installation option, and the installation directory.

When you complete the preliminary tasks, you will continue with the installer prompts and will provide information to configure the domain.

Run the Installer

Perform the following steps to run the installer:

1. Log in to the machine with a system user account.
2. Close all other applications.
3. On a shell command line, run the `install.sh` file from the root directory.

   The installer displays the message to verify that the locale environment variables are set.
4. If the environment variables are not set, press **n** to exit the installer and set them as required.

   If the environment variables are set, press **y** to continue.
5. Press **1** to install Informatica Big Data Suite Products.

   The installer displays different options based on the platform you are installing on.

The following options appear:

a. Press **1** to run the Pre-Installation System Check Tool.

   For more information about running the Pre-Installation (i10Pi) System Check Tool, see "Run the Pre-Installation (i10Pi) System Check Tool" on page 35.

b. Press **2** to run the Informatica Kerberos SPN Format Generator.

   For more information about running the Informatica Kerberos SPN Format Generator, see "Running the SPN Format Generator on Linux" on page 69.

c. Press **3** to run the installer.

The **Welcome** section appears.

Accept Terms and Conditions

1. Read the terms and conditions for Informatica installation and the product usage toolkit and select **I agree to the terms and conditions**.
   a. Press **1** if you do not want to accept the terms and conditions.
   b. Press **2** to accept the terms and conditions.
2. Version 10.2.2 is for big data products only, such as Big Data Management and Big Data Quality. This version does not support non-big data products, such as PowerCenter or Informatica Data Quality.
   a. Press **1** and type **quit** to quit the installation.
   b. Press **2** to continue the installation.

   If you choose to not accept the terms and condition, the installer prompts you to accept the terms and conditions.

The **Component Selection** sections appears.

Choose the Installation Option

After you accept terms and conditions, you can install Informatica domain services supporting Big Data Management services, Enterprise Data Catalog, and Enterprise Data Lake.

1. Press **3** to install Informatica Enterprise Data Lake.

   When you select this option, you can choose to install Only Enterprise Data Lake, Enterprise Data Catalog and Enterprise Data Lake, and Informatica domain services, supporting Big Data Management services, Enterprise Data Catalog, and Enterprise Data Lake.
2. Select whether current version of the Informatica domain services is installed on the node.
   a. Press **1** if current version of the Informatica domain services is not installed on the node.
   b. Press **2** if current version of the Informatica domain services is installed on the node.
3. Select the Hadoop cluster type for Enterprise Data Catalog.
   a. Press **1** to select **External** cluster type.
   b. Press **2** to select **Embedded** cluster type.

      If you choose to run on an internal cluster, the installer creates a cluster to run the services.
4. Select whether you have read and accepted terms and conditions to use Java SE Development Kit software.
   a. Press **1** to not accept the terms and conditions to use Java SE Development Kit software.
   b. Press **2** to accept the terms and conditions to use Java SE Development Kit software.

The **Installation Prerequisites** section displays the installation requirements. Verify that all requirements are met before you continue the installation.

Tune the Application Service

After you review the installation prerequisites, you can choose to tune the application services for better performance based on the deployment type in your environment. If you do not tune now, you can tune the services later through infacmd.

1. Select if you want the installer to tune the application services.

Prompt	Description
Select if you want the installer to tune the application services.	Select if you want to tune the services. 1 - No 2 - Yes Select no if you do not want to tune the services. Select yes if you want to tune the services.

If you are joining the node to existing domain, ensure the deployment type you select here is same deployment type as the gateway nodes.

2. Select the deployment type associated with the Informatica environment.

Prompt	Description
1. Sandbox	Choose this option if the environment is used for proof of concepts or as a sandbox environment with minimal users. Sandbox environments are typically configured with 16 cores, 32 GB RAM, and about 50 GB disk space.
2. Basic	Choose this option if the environment is used for low volume processing environments with low levels of concurrency. Basic environments are typically single- or multi-node setups configured with 24 cores, 64 GB RAM, and about 100 GB disk space.
3. Standard	Choose this option if the environment is used for high volume processing but with low levels of concurrency. Standard environments are typically multi-node setups configured with 64 GB RAM, more than 100 GB disk space per node, and 48 cores across nodes.
4. Advanced	Choose this option if the environment is used for high volume processing with high levels of concurrency. Advanced environments are typically multi-node setups configured with 128 GB RAM, more than 100 GB disk space per node, and 96 cores across nodes.

3. Select whether you want to change the deployment type or continue with the current deployment selection.
   a. Press **1** to change the deployment type.
   b. Press **2** to continue with the current deployment selection.

The **License and Installation Directory** section appears.

Specify the Installation Directory

After you verify the installation prerequisites, you can specify the installation directory.

1. Enter the path to the license key file and press **Enter**.
2. Enter the installation directory.

   The directory names in the path must not contain spaces or the following special characters: @|* $ # ! % ( ) { } [ ] , ; ' Default is /home/toolinst.

   Note: Informatica recommends using alphanumeric characters in the installation directory path. If you use a special character such as á or €, unexpected results might occur at run time.
3. Select whether you want to run the pre-validation utility.
   a. Press **1** to skip the pre-validation utility.
   b. Press **2** to run the pre-validation utility.

   The utility helps you validate the prerequisites to install Enterprise Data Catalog in an embedded cluster. The utility also validates the Informatica domain, cluster hosts, and the Hadoop cluster services configuration.

If you choose to skip the pre-validation utility, the **Pre-Installation Summary** section appears. Review the installation summary.

If choose to run the pre-validation utility, the **Embedded Hadoop Cluster Pre-Validaion** section appears.

Prepare Pre-validation for the External Cluster

If you chose to use external Hadoop cluster, you can configure the external Hadoop cluster.

1. Select if there is passwordless Secure Shell connection between the Informatica domain and the cluster nodes.
   a. Press **1** to secure the connection between the Informatica domain and the cluster nodes.
   b. Press **2** if you do not want to secure the connection between the Informatica domain and the cluster nodes.
2. If you chose to secure the connection between the Informatica domain and the cluster nodes, enter the secure connection information.

   The table describes the information you can enter to secure the connection:

Property	Description
Host SSH user	User to connect to other hosts from the domain host.
SSH host names	Cluster nodes that you want to connect using passwordless Secure shell.

3. Select whether the cluster uses Kerberos authentication.
   a. Press **1** if the Hadoop cluster uses Kerberos authentication.
   b. Press **2** if the Hadoop cluster does not use Kerberos authentication.

If you chose that the cluster uses Kerberos authentication, enter the Kerberos information.

Property	Description
Keytab Location	Location of the merged user and host keytab file.
KDC Domain Name	Name of the KDC domain.

4. Select if you want to validate the cluster configuration.
   a. Press **1** to validate the cluster configuration.
   b. Press **2** if you do not want to validate the cluster configuration.
5. Select if you want the installer to automatically configure the cluster.
   a. Press **1** if you do not want the installer to automatically configure the cluster. If you do not choose this option, you cannot validate the cluster configuration.
   b. Press **2** if you want the installer to automatically configure the cluster.
6. If you chose to validate the cluster configuration, you can select the cluster type to configure.

   The following table describes the options you can select the cluster type to configure:

Option	Description
Cloudera	Select to create a cluster configuration for a Cloudera cluster.
Hortonworks	Select to create a cluster configuration for a Hortonworks cluster.
Azure HDInsight	Select to create a cluster configuration for a Azure HDInsight cluster.

7. Enter the information to configure the cluster.

   The following table describes the properties you need to set for configuring the cluster.

Option	Description
Cluster Hadoop distribution URL	URL to access the Hadoop cluster.
Cluster Hadoop distribution URL user	User name to access the Hadoop cluster.
Cluster Hadoop distribution URL password	Password to access the Hadoop cluster.
YARN Queue Name	The YARN scheduler queue name used by the Blaze engine that specifies available resources on a cluster.
Service cluster name	Name of the service cluster.

Perform Pre-validation for the Embedded Hadoop Cluster

If you chose to use embedded Hadoop cluster, you can configure the embedded Hadoop cluster.

1. Select whether you want to enable Kerberos authentication for the cluster.
   a. Press **1** to configure the Hadoop cluster to run on a network that does not use Kerberos authentication.
   b. Press **2** to configure the Hadoop cluster to run on a network that uses Kerberos authentication.

      If you chose to run on a network that uses Kerberos authentication, enter the Kerberos information.

Property	Description
Keytab Location	Location of the merged user and host keytab file.
Kerberos Configuration File	Location of the Kerberos configuration file.

2. Enter the gateway user name and press **Enter**. Default is **root**.
3. Enter the Informatica Hadoop cluster gateway hostname in the following format: `<hostname>.<FQDN>` and press **Enter**.
4. Enter the list of comma-separated Informatica Hadoop cluster nodes as shown in the following format: `<hostname>.<FQDN>, <hostname1>.<FQDN>, <hostname2>.<FQDN>` and press **Enter**.
5. Enter the Informatica Hadoop cluster gateway port and press **Enter**. Default is **8080**.

   Make sure that you do not configure Oracle with port 8080 on the same machine where Informatica Cluster Service runs.
6. Enter the path to the working directory and press **Enter**. The path indicates the location where you want to mount the Informatica Cluster Service.

   The installer starts the pre-validation utility.

The **Pre-Installation Summary** section appears. Review the installation summary.

Configure the Domain

This task includes installer prompts to configure the domain. You will provide information to create a domain, configure the domain security, domain repository, and application services.

Configure the Domain Options

After you review the Pre-Installation summary, you can enter the domain information.

1. Press **1** to create a domain.

   When you create a domain, the node that you create becomes a gateway node in the domain. The gateway node contains a Service Manager that manages all domain operations.
2. Select whether you want to enable secure communication for services in the domain.
   a. Press **1** to disable secure communication for the domain.
   b. Press **2** to enable secure communication for the domain.

By default, if you enable secure communication for the domain, the installer sets up an HTTPS connection for the Informatica Administrator. You can also create a domain configuration repository on a secure database.

3. Specify the connection details for Informatica Administrator.
   a. If you do not enable secure communication for the domain, you can specify whether to set up a secure HTTPS connection for the Informatica Administrator.

      The following table describes the options available to enable or disable a secure connection to Informatica Administrator:

Option	Description
Enable HTTPS for Informatica Administrator	Set up a secure connection to Informatica Administrator.
Disable HTTPS	Do not set up a secure connection to Informatica Administrator.

   b. If you enable secure communication for the domain or if you enable HTTPS connection for the Informatica Administrator, enter the keystore file and port number for the HTTPS connection to Informatica Administrator.

      The following table describes the connection information you must enter if you enable HTTPS:

Option	Description
Port	Port number for the HTTPS connection.
Keystore file	Select whether to use a keystore file generated by the installer or a keystore file you create. You can use a keystore file with a self-signed certificate or a certificate signed by a certification authority. 1 - Use a keystore generated by the installer 2 - Specify a keystore file and password If you select to use a keystore file generated by the installer, the installer creates a self-signed keystore file named Default.keystore in the following location: `<Informatica installation directory>/tomcat/conf/`

   c. If you specify the keystore, enter the password and location of the keystore file.
   d. If you enabled secure communication for the domain, the **Domain Security - Secure Communication** section appears.
   e. If you did not enable secure communication for the domain, the **Domain Configuration Repository** section appears. Skip to "Configure Domain Repository Details" on page 114.

4. Select whether to enable SAML authentication to configure Security Assertion Markup Language (SAML)-based single sign-on (SSO) support for web-based Informatica applications in an Informatica domain.

The following table describes the information you must enter to enable SAML authentication:

Prompt	Description
Enable SAML authentication	Select whether to enable SAML authentication: 1 - No If you select No, skip to "Configure Domain Security" on page 113. 2 - Yes If you select Yes, configure the SAML authentication.

5. Enter the Identity Provider URL for the domain.
6. Enter the identity provider assertion signing certificate alias name.
7. Select whether to use the default Informatica SSL certificates or to use your SSL certificates to enable secure communication in the domain.

   The following table describes the SSL certificate options for SAML authentication:

Option	Description
Use the default Informatica SSL certificate file.	Select to use the default Informatica truststore file for SAML authentication.
Enter the location of the SSL certificate file.	Select to use a custom truststore file for SAML authentication. Specify the directory containing the custom truststore file on gateway nodes within the domain. Specify the directory only, not the full path to the file.

8. If you provide the security certificates, specify the location and passwords of the keystore and truststore files.

   The following table describes the location and password of the truststore file:

Property	Description
Truststore Directory	Specify the directory containing the custom truststore file on gateway nodes within the domain. Specify the directory only, not the full path to the file.
Truststore Password	The password for the custom truststore file.

The **Configure Domain Security** appears.

Configure Domain Security

After you configure the domain, you can configure domain security.

- In the Domain Security - Secure Communication section, specify whether to use the default Informatica SSL certificates or to use your SSL certificates to secure domain communication.
  a. Select the type of SSL certificates to use.

The following table describes the options for the SSL certificates that you can use to secure the Informatica domain:

Option	Description
Use the default Informatica SSL certificates	Use the default SSL certificates contained in the default keystore and truststore. **Note:** If you do not provide an SSL certificate, Informatica uses the same default private key for all Informatica installations. If you use the default Informatica keystore and truststore files, the security of your domain could be compromised. To ensure a high level of security for the domain, select the option to specify the location of the SSL certificate files.
Use custom SSL certificates	Specify the path for the keystore and truststore files that contain the SSL certificates. You must also specify the keystore and truststore passwords. You can provide a self-signed certificate or a certificate issued by a certificate authority (CA). You must provide SSL certificates in PEM format and in Java Keystore (JKS) files. Informatica requires specific names for the SSL certificate files for the Informatica domain. You must use the same SSL certificates for all nodes in the domain. Store the truststore and keystore files in a directory accessible to all the nodes in the domain and specify the same keystore file directory and truststore file directory for all nodes in the same domain.

b. If you provide the SSL certificate, specify the location and passwords of the keystore and truststore files.

The following table describes the parameters that you must enter for the SSL certificate files:

Property	Description
Keystore file directory	Directory that contains the keystore files. The directory must contain files named infa_keystore.jks.
Keystore password	Password for the keystore infa_keystore.jks.
Truststore file directory	Directory that contains the truststore files. The directory must contain files named infa_truststore.jks and infa_truststore.pem.
Truststore password	Password for the infa_truststore.jks file.

The **Domain Configuration Repository** section appears.

Configure Domain Repository Details

After you configure domain security, you can configure domain repository details.

1. Select the database to use for the domain configuration repository details.

The following table lists the databases you can use for the domain configuration repository:

Prompt	Description
Database type	Type of database for the domain configuration repository. Select from the following options: 1 - Oracle 2 - Microsoft SQL Server 3 - IBM DB2 4 - Sybase ASE

The Informatica domain configuration repository stores metadata for domain operations and user authentication. The domain configuration repository must be accessible to all gateway nodes in the domain.

2. Enter the properties for the database user account.

   The following table lists the properties for the database user account:

Property	Description
Database user ID	Name for the domain configuration database user account.
User password	Password for the domain configuration database user account.

3. Select whether to create a secure domain configuration repository.

   You can create a domain configuration repository in a database secured with the SSL protocol. To create a domain configuration repository in a secure database, press 1 and skip to step 5on page 117.

   To create a domain configuration repository in an unsecure database, press 2.

4. If you do not create a secure domain configuration repository, enter the parameters for the database.

   a. If you select IBM DB2, select whether to configure a tablespace and enter the tablespace name.

      The following table describes the properties that you must configure for the IBM DB2 database:

Property	Description
Configure tablespace	Select whether to specify a tablespace: 1 - No 2 - Yes In a single-partition database, if you select No, the installer creates the tables in the default tablespace. In a multi-partition database, you must select Yes.
Tablespace	Name of the tablespace in which to create the tables. Specify a tablespace that meets the pageSize requirement of 32768 bytes. In a single-partition database, if you select Yes to configure the tablespace, enter the name of the tablespace in which to create the tables. In a multi-partition database, specify the name of the tablespace that resides in the catalog partition of the database.

b. If you select Microsoft SQL Server, enter the schema name for the database.

The following table describes the properties that you must configure for the Microsoft SQL Server database:

Property	Description
Schema name	Name of the schema that will contain domain configuration tables. If this parameter is blank, the installer creates the tables in the default schema.

c. To enter the JDBC connection information using the JDBC URL information, press **1**. To enter the JDBC connection information using a custom JDBC connection string, press **2**.

d. Enter the JDBC connection information.

- To enter the connection information using the JDBC URL information, specify the JDBC URL properties.
  The following table describes the database connection information:

Prompt	Description
Database host name	Host name for the database.
Database port number	Port number for the database.
Database service name	Service or database name : - Oracle: Enter the service name. - Microsoft SQL Server: Enter the database name. - IBM DB2: Enter the service name. - Sybase ASE: Enter the database name.
Configure JDBC Parameters	Select whether to add additional JDBC parameters to the connection string: 1 - Yes 2 - No If you select Yes, enter the parameters or press Enter to accept the default. If you select No, the installer creates the JDBC connection string without parameters.

- To enter the connection information using a custom JDBC connection string, type the connection string.
  Use the following syntax in the JDBC connection string:
  IBM DB2

```
jdbc:Informatica:db2://host_name:port_no;DatabaseName=
```

  Oracle

```
jdbc:Informatica:oracle://host_name:port_no;ServiceName=
```

  Microsoft SQL Server

```
jdbc:Informatica:sqlserver://
host_name:port_no;SelectMethod=cursor;DatabaseName=
```

Microsoft Azure SQL Database

```
jdbc:Informatica:sqlserver://
host_name:port_no;SelectMethod=cursor;DatabaseName=database_name;SnapshotSeria
lizable=true;EncryptionMethod=SSL;HostNameInCertificate=*.database.windows.net
;ValidateServerCertificate=false
```

Sybase

```
jdbc:Informatica:sybase://host_name:port_no;DatabaseName=
```

Verify that the connection string contains all the connection parameters required by your database system.

5. If you create a secure domain configuration repository, enter the parameters for the secure database.

   If you create the domain configuration repository on a secure database, you must provide the truststore information for the database. You must also provide a JDBC connection string that includes the security parameters for the database.

   The following table describes the options available to create a secure domain configuration repository database:

Property	Description
Database truststore file	Path and file name of the truststore file for the secure database.
Database truststore password	Password for the truststore file.
Custom JDBC Connection String	JDBC connection string to connect to the secure database, including the host name and port number and the security parameters for the database.

   In addition to the host name and port number for the database server, you must include the following secure database parameters: You can use the following syntax for the connection strings:

   EncryptionMethod

   Required. Indicates whether data is encrypted when transmitted over the network. This parameter must be set to`SSL`.

   ValidateServerCertificate

   Optional. Indicates whether Informatica validates the certificate that the database server sends.

   If this parameter is set to True, Informatica validates the certificate that the database server sends. If you specify the HostNameInCertificate parameter, Informatica also validates the host name in the certificate.

   If this parameter is set to False, Informatica does not validate the certificate that the database server sends. Informatica ignores any truststore information that you specify

   Default is True.

   HostNameInCertificate

   Optional. Host name of the machine that hosts the secure database. If you specify a host name, Informatica validates the host name included in the connection string against the host name in the SSL certificate.

cryptoProtocolVersion

Required. Specifies the cryptographic protocol to use to connect to a secure database. You can set the parameter to `cryptoProtocolVersion=TLSv1.1` or `cryptoProtocolVersion=TLSv1.2` based on the cryptographic protocol used by the database server:

- Oracle: `jdbc:Informatica:oracle://host_name:port_no;ServiceName=service_name;EncryptionMethod=SSL;HostNameInCertificate=DB_host_name;ValidateServerCertificate=true_or_false`
- IBM DB2: `jdbc:Informatica:db2://host_name:port_no;DatabaseName=database_name;EncryptionMethod=SSL;HostNameInCertificate=DB_host_name;ValidateServerCertificate=true_or_false`
- Microsoft SQL Server: `jdbc:Informatica:sqlserver://host_name:port_no;SelectMethod=cursor;DatabaseName=database_name;EncryptionMethod=SSL;HostNameInCertificate=DB_host_name;ValidateServerCertificate=true_or_false`

Note: The installer does not validate the connection string. Verify that the connection string contains all the connection parameters and security parameters required by your database.

6. If the database contains a domain configuration repository for a previous domain, select to overwrite the data or set up another database.

   The following table describes the options of overwriting the data or setting up another database when you create a domain configuration repository for a previous domain:

Option	Description
1 - OK	Enter the connection information for a new database.
2 - Continue	The installer overwrites the data in the database with new domain configuration.

The **Domain Security - Encryption Key** section appears.

Configure the Encryption Key

After you configure domain repository, you can configure encryption key.

▶ In the **Domain Security - Encryption Key** section, enter the keyword and directory for the encryption key for the Informatica domain.

The following table describes the encryption key parameters that you must specify when you create a domain:

Property	Description
Keyword	Keyword to use to create a custom encryption key to secure sensitive data in the domain. The keyword must meet the following criteria: - From 8 to 20 characters long - Includes at least one uppercase letter - Includes at least one lowercase letter - Includes at least one number - Does not contain spaces The encryption key is created based on the keyword that you provide when you create the Informatica domain.
Encryption key directory	Directory in which to store the encryption key for the domain. By default, the encryption key is created in the following directory: `<Informatica installation directory>/isp/config/keys`.

The installer sets different permissions to the directory and the files in the directory. For more information about the permissions for the encryption key file and directory, see "Secure Files and Directories" on page 104.

The **Domain and Node Configuration** section appears.

Configure the Domain and Node

After you configure the encryption key, you can configure the domain and node.

1. Enter the information for the domain and the node that you want to create.

   The following table describes the properties that you set for the domain and gateway node.

Property	Description
Domain name	Name of the Informatica domain to create. The default domain name is Domain_<MachineName>. The name must not exceed 128 characters and must be 7-bit ASCII only. The name cannot contain a space or any of the following characters: `` ` % * + ; " ? , < > \ / ``
Node name	Name of the node to create.
Node host name	Host name or IP address of the machine on which to create the node. If the machine has a single network name, use the default host name. If the a machine has multiple network names, you can modify the default host name to use an alternate network name. **Note:** The node host name cannot contain the underscore (_) character. Do not use localhost. The host name must explicitly identify the machine.

Property	Description
Node port number	Port number for the node. The default port number for the node is 6005. If the port number is not available on the machine, the installer displays the next available port number.
Domain user name	User name for the domain administrator. You can use this user name to initially log in to Informatica Administrator. Use the following guidelines: - The name is not case sensitive and cannot exceed 128 characters. - The name cannot include a tab, newline character, or the following special characters: % * + / ? ; < >\n- The name can include an ASCII space character except for the first and last character. Other space characters are not allowed.

2. Select whether you want to enable password complexity to secure sensitive data in the domain.

   The following table describes the password complexity:

Prompt	Description
Password complexity	Select whether you want to enable password complexity. 1 - Yes 2 - No If you select Yes, the password must meet the following requirements: It must be at least eight characters long and contain at least one alpha character, one numeric character, and one special character.
Domain password	Password for the domain administrator. The password must be more than 2 characters and must not exceed 16 characters. Not available if you configure the Informatica domain to run on a network with Kerberos authentication.
Confirm password	Enter the password again to confirm. Not available if you configure the Informatica domain to run on a network with Kerberos authentication.

3. Select whether to display the default ports for the domain and node components assigned by the installer.

   The following table describes the advanced port configuration page:

Prompt	Description
Display advanced port configuration page	Select whether to display the port numbers for the domain and node components assigned by the installer: 1 - No 2 - Yes If you select Yes, the installer displays the default port numbers assigned to the domain components. You can specify the port numbers to use for the domain and node components. You can also specify a range of port numbers to use for the service process that will run on the node. You can use the default port numbers or specify new port numbers. Verify that the port numbers you enter are not used by other applications.

4. If you display the port configuration page, enter new port numbers at the prompt or press Enter to use the default port numbers.

   The following table describes the ports that you can set:

Port	Description
Service Manager port	Port number used by the Service Manager on the node. The Service Manager listens for incoming connection requests on this port. Client applications use this port to communicate with the services in the domain. The Informatica command line programs use this port to communicate to the domain. This is also the port for the SQL data service JDBC/ODBC driver. Default is 6006.
Service Manager Shutdown port	Port number that controls server shutdown for the domain Service Manager. The Service Manager listens for shutdown commands on this port. Default is 6007.
Informatica Administrator port	Port number used by Informatica Administrator. Default is 6008.
Informatica Administrator HTTPS port	No default port. Enter the required port number when you create the service. Setting this port to 0 disables an HTTPS connection to the Administrator tool.
Informatica Administrator shutdown port	Port number that controls server shutdown for Informatica Administrator. Informatica Administrator listens for shutdown commands on this port. Default is 6009.
Minimum port number	Lowest port number in the range of dynamic port numbers that can be assigned to the application service processes that run on this node. Default is 6014.
Maximum port number	Highest port number in the range of dynamic port numbers that can be assigned to the application service processes that run on this node. Default is 6114.

5. Select if you want to create **Enterprise Data Catalog Application Services**.

   The following table describes the option to configure the application services:

Prompt	Description
Create Enterprise Data Catalog Application Services	Select whether you want to configure the Model Repository Service and Data Integration Service. 1 - Yes 2 - No If you select Yes, you can create the application services. If you select No, you can create the application services from the Administrator tool.

6. Select if you want to create a **monitoring Model Repository Service** to monitor domain statistics.

The following table describes the options to configure monitoring Model Repository :

Prompt	Description
Create a monitoring Model Repository Service	Select whether you want to create a monitoring Model Repository Service. 1 - Yes 2 - No If you select Yes, you can create a monitoring Model Repository Service. If you select No, you can create a monitoring Model Repository Service from the Administrator tool.

If you choose to create Enterprise Data Catalog Application Services, the **Model Repository Database** section appears. If you choose not to create Enterprise Data Catalog Application Services, the **Post-Installation Summary** section indicates whether the installation completed successfully. The summary also shows the status of the installed components and their configuration.

Configure the Model Repository Database

After you configure the domain and the node, you can configure the Model repository database properties.

1. Enter the Model Repository Service name.

   Enter the name of the service. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters:

   ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [

   You cannot change the name of the service after you create it.

   If you selected process level SPN, specify the Model Repository Service keytab file. The keytab file for the Model Repository Service process. The keytab file must have the following name: .keytab

2. Select the database to configure Model repository database.

   The following table lists the database type for the Model repository:

Prompt	Description
Database type	Type of database for the Model repository. Select from the following options: 1 - Oracle 2 - Microsoft SQL Server 3 - IBM DB2

3. Enter the properties for the database user account.

   The following table lists the properties for the database user account:

Property	Description
Database user ID	Name for the Model repository database user account.
User password	Password for the Model repository user account.

4. Select whether to create a secure Model repository database.

You can create a model repository service in a database secured with the SSL protocol. To create a model repository service in a secure database, press 1 and skip to step to enter the JDBC connection information.

To create a Model repository in an unsecure database, press 2.

5. If you do not create a secure Model repository, enter the parameters for the database.
   a. If you select IBM DB2, select whether to configure a tablespace and enter the tablespace name.

      The following table describes the properties that you must configure for the IBM DB2 database:

Property	Description
Configure tablespace	Select whether to specify a tablespace: 1 - No 2 - Yes In a single-partition database, if you select No, the installer creates the tables in the default tablespace. In a multi-partition database, you must select Yes.
Tablespace	Name of the tablespace in which to create the tables. Specify a tablespace that meets the pageSize requirement of 32768 bytes. In a single-partition database, if you select Yes to configure the tablespace, enter the name of the tablespace in which to create the tables. In a multi-partition database, specify the name of the tablespace that resides in the catalog partition of the database.

   b. If you select Microsoft SQL Server, enter the schema name for the database.

      The following table describes the properties that you must configure for the Microsoft SQL Server database:

Property	Description
Schema name	Name of the schema that will contain domain configuration tables. If this parameter is blank, the installer creates the tables in the default schema.

   c. To enter the JDBC connection information using the JDBC URL information, press **1**. To enter the JDBC connection information using a custom JDBC connection string, press **2**.

   d. Enter the JDBC connection information.

      - To enter the connection information using the JDBC URL information, specify the JDBC URL properties.
        The following table describes the database connection information:

Prompt	Description
Database host name	Host name for the database.
Database port number	Port number for the database.

Prompt	Description
Database service name	Service or database name : - Oracle: Enter the service name. - Microsoft SQL Server: Enter the database name. - IBM DB2: Enter the service name.
Configure JDBC Parameters	Select whether to add additional JDBC parameters to the connection string: 1 - Yes 2 - No If you select Yes, enter the parameters or press Enter to accept the default. If you select No, the installer creates the JDBC connection string without parameters.

- To enter the connection information using a custom JDBC connection string, type the connection string.
  Use the following syntax in the JDBC connection string:
  IBM DB2

  ```
  jdbc:Informatica:db2://host_name:port_no;DatabaseName=
  ```

  Oracle

  ```
  jdbc:Informatica:oracle://host_name:port_no;ServiceName=
  ```

  Microsoft SQL Server

  ```
  jdbc:Informatica:sqlserver://
  host_name:port_no;SelectMethod=cursor;DatabaseName=
  ```

  Microsoft Azure SQL Database

  ```
  jdbc:Informatica:sqlserver://
  host_name:port_no;SelectMethod=cursor;DatabaseName=database_name;SnapshotSeria
  lizable=true;EncryptionMethod=SSL;HostNameInCertificate=*.database.windows.net
  ;ValidateServerCertificate=false
  ```

  Sybase

  ```
  jdbc:Informatica:sybase://host_name:port_no;DatabaseName=
  ```

  Verify that the connection string contains all the connection parameters required by your database system.

The **Service Parameters** section appears.

Configure the Monitoring Model Repository Database

If you chose to create a monitoring Model Repository Service, you can provide connection information about the repository database.

1. Enter the monitoring Model Repository Service name.

   Enter the name of the service. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters:

   ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [

   You cannot change the name of the service after you create it.

   If you selected process level SPN, specify the monitoring Model Repository Service keytab file. The keytab file for the monitoring Model Repository Service process. The keytab file must have the following name: .keytab

2. Select the database for the monitoring Model repository.

   The following table lists the database type for monitoring Model repository:

Prompt	Description
Database type	Type of database type for monitoring Model repository. Select from the following options: 1 - Oracle 2 - Microsoft SQL Server 3 - IBM DB2

3. Enter the properties for the database user account.

   The following table lists the properties for the database user account:

Property	Description
Database user ID	Name for the monitoring Model repository database user account.
User password	Password for the Monitoring model repository database user account.

4. Select whether to create a secure monitoring Model repository.

   You can create a monitoring Model repository in a database secured with the SSL protocol. To create a monitoring Model repository in a secure database, press 1 and skip to step to enter the JDBC connection information.

   To create a monitoring Model repository in an unsecured database, press 2.

5. If you do not create a secure monitoring Model repository, enter the parameters for the database.

   a. If you select IBM DB2, select whether to configure a tablespace and enter the tablespace name.

The following table describes the properties that you must configure for the IBM DB2 database:

Property	Description
Configure tablespace	Select whether to specify a tablespace: 1 - No 2 - Yes In a single-partition database, if you select No, the installer creates the tables in the default tablespace. In a multi-partition database, you must select Yes.
Tablespace	Name of the tablespace in which to create the tables. Specify a tablespace that meets the pageSize requirement of 32768 bytes. In a single-partition database, if you select Yes to configure the tablespace, enter the name of the tablespace in which to create the tables. In a multi-partition database, specify the name of the tablespace that resides in the catalog partition of the database.

b. If you select Microsoft SQL Server, enter the schema name for the database.

The following table describes the properties that you must configure for the Microsoft SQL Server database:

Property	Description
Schema name	Name of the schema that will contain domain configuration tables. If this parameter is blank, the installer creates the tables in the default schema.

c. To enter the JDBC connection information using the JDBC URL information, press **1**. To enter the JDBC connection information using a custom JDBC connection string, press **2**.

d. Enter the JDBC connection information.

- To enter the connection information using the JDBC URL information, specify the JDBC URL properties.
  The following table describes the database connection information:

Prompt	Description
Database host name	Host name for the database.
Database port number	Port number for the database.

Prompt	Description
Database service name	Service or database name : - Oracle: Enter the service name. - Microsoft SQL Server: Enter the database name. - IBM DB2: Enter the service name.
Configure JDBC Parameters	Select whether to add additional JDBC parameters to the connection string: 1 - Yes 2 - No If you select Yes, enter the parameters or press Enter to accept the default. If you select No, the installer creates the JDBC connection string without parameters.

- To enter the connection information using a custom JDBC connection string, type the connection string.
  Use the following syntax in the JDBC connection string:
  IBM DB2

  ```
  jdbc:Informatica:db2://host_name:port_no;DatabaseName=
  ```

  Oracle

  ```
  jdbc:Informatica:oracle://host_name:port_no;ServiceName=
  ```

  Microsoft SQL Server

  ```
  jdbc:Informatica:sqlserver://
  host_name:port_no;SelectMethod=cursor;DatabaseName=
  ```

  Microsoft Azure SQL Database

  ```
  jdbc:Informatica:sqlserver://
  host_name:port_no;SelectMethod=cursor;DatabaseName=database_name;SnapshotSeria
  lizable=true;EncryptionMethod=SSL;HostNameInCertificate=*.database.windows.net
  ;ValidateServerCertificate=false
  ```

  Sybase

  ```
  jdbc:Informatica:sybase://host_name:port_no;DatabaseName=
  ```

  Verify that the connection string contains all the connection parameters required by your database system.

The **Service Parameters** section appears.

Configure the Application Service Parameters

After you configure the Model Repository database, you can configure the service parameters for the application services.

1. Enter the following service parameter information:

Port	Description
Data Integration Service name	Name of the Data Integration Service to create in the Informatica domain.
HTTP protocol type	Type of connection to the Data Integration Service. Select one of the following options: - HTTP. Requests to the service uses an HTTP connection. - HTTPS. Requests to the service uses a secure HTTP connection. - HTTP&HTTPS. Requests to the service can use either an HTTP or HTTPS connection.
HTTP port	Port number to used for the Data Integration Service. Default is 9085.
HTTPS port	Port number to used for the Data Integration Service. Default is 9085.

2. Select the SSL certificates contained to secure the Data Integration Service.

Option	Description
Use the default Informatica SSL certificate files	Use the default Informatica SSL certificates contained in the default keystore and truststore. **Note:** If you do not provide an SSL certificate, Informatica uses the same default private key for all Informatica installations. If you use the default Informatica keystore and truststore files, the security of your domain could be compromised. To ensure a high level of security for the domain, select the option to specify the location of the SSL certificate files.
Use custom SSL certificates	Use custom SSL certificates. You must specify the location of the keystore and truststore files. You can provide a self-signed certificate or a certificate issued by a certificate authority (CA). You must provide SSL certificates in PEM format and in Java Keystore (JKS) files. Informatica requires specific names for the SSL certificate files for the Informatica domain. You must use the same SSL certificates for all nodes in the domain. Store the truststore and keystore files in a directory accessible to all the nodes in the domain and specify the same keystore file directory and truststore file directory for all nodes in the same domain.

If you choose to use custom SSL certificates, enter the following information.

Property	Description
Keystore file directory	Directory that contains the keystore files. The directory must contain files named infa_keystore.jks and infa_keystore.pem.
Keystore password	Password for the keystore infa_keystore.jks.

Property	Description
Truststore file directory	Directory that contains the truststore files. The directory must contain files named infa_truststore.jks and infa_truststore.pem.
Truststore password	Password for the infa_truststore.jks file.

3. Do you want to enable big data job recovery for the Data Integration Service?
   - Yes
   - No

   If you choose Yes, you can recover mapping jobs that the Data Integration Service pushes to the Spark engine for processing. Default is No.
4. Do you want to create a cluster configuration?

   The cluster configuration enables the Data Integration Service to push mapping logic to the cluster. If you are integrating with the Hadoop environment, you can create a cluster configuration.

   Press 1 if you want to create a cluster configuration.

   Press 2 if you do not want to create a cluster configuration. Default is 1.

   Note: To create a cluster configuration for the Databricks environment, use the Administrator tool after you complete installation.

   You must create a cluster configuration if you plan to use Enterprise Data Lake. If you want to create Enterprise Data Lake Services during the installation, you must create a cluster configuration. After installation, refer to the *Big Data Management Integration Guide* to fully integrate the domain with the Hadoop environment.
5. Select whether you want to configure profiling warehouse connection.
   a. Press **1** to configure the profiling warehouse connection.
   b. Press **2** to skip configuring the profiling warehouse connection.
   c. If you choose to configure the profiling warehouse connection, the **Profiling Warehouse Connection Database** section appears.
   d. If you choose to skip the profiling warehouse connection, the **Post-Installation Summary** section indicates whether the installation completed successfully. The summary also shows the status of the installed components and their configuration.
6. Select whether you want to configure the Content Management Service for data domain discovery.
   a. Press **1** to configure the Content Management Service for data domain discovery.
   b. Press **2** to skip configuring the Content Management Service for data domain discovery.
   c. If you choose to configure the Content Management Service for data domain discovery, the **Content Management Service Parameters and Database** section appears.
   d. If you choose to skip configure the Content Management Service for data domain discovery, the **Post-Installation Summary** section indicates whether the installation completed successfully. The summary also shows the status of the installed components and their configuration.

Create the Cluster Configuration

Create the cluster configuration, which contains configuration information about the Hadoop cluster. The cluster configuration enables the Data Integration Service to push jobs to the Hadoop environment. You must create a cluster configuration if you plan to use Enterprise Data Lake.

You import configuration properties from the Hadoop cluster to create a cluster configuration. You can import the properties from an archive file that the Hadoop administrator creates, or you can import the properties directly from the cluster.

When you create the cluster configuration, you can also choose to create Hadoop, Hive, HBase, and HDFS connections to the Hadoop environment. The installer appends the connection type to the cluster configuration name to create each connection name.

1. Enter the name of the cluster configuration to create.
2. Specify the Hadoop distribution for the cluster.

   The following table describes the options you can specify:

Option	Description
1	Select to create a cluster configuration for a Cloudera cluster.
2	Select to create a cluster configuration for a Hortonworks cluster.
3	Select to create a cluster configuration for a Azure HDInsight cluster.
4	Select to create a cluster configuration for a MapR cluster. You must import the MapR cluster configuration properties from an archive file.
5	Select to create a cluster configuration for an Amazon EMR cluster. You must import the Amazon EMR cluster configuration properties from an archive file.

3. Import configuration properties from the Hadoop cluster to create the cluster configuration.
   - To import the properties from an archive file, press **1**. If you create a cluster configuration for an Amazon EMR cluster or for a MapR cluster, you must import the properties from an archive file.
   - To import the properties directly from the cluster, press **2**.
4. If you choose to import the properties directly from the cluster, specify the connection properties.

   The following table describes the properties you specify:

Property	Description
Host	The host name or IP address of the cluster manager.
Port	Port of the cluster manager.
User ID	Cluster user name.

Property	Description
Password	Password for the cluster user.
Cluster Name	Name of the cluster. Use the display name if the cluster manager manages multiple clusters. If you do not provide a cluster name, the wizard imports information based on the default cluster.

5. To create the Hadoop, Hive, HDFS, and HBase connections to the cluster, press **1**.

   The installer appends the connection type to the cluster configuration name to create each connection name.

Configure Enterprise Data Catalog

This task includes installer prompts to configure Enterprise Data Catalog. You will provide basic information for configuring the application services and Hadoop cluster.

When you complete the preliminary tasks, you will continue with the installer prompts to configure Enterprise Data Lake.

Configure Profiling Warehouse Database Details

If you chose to configure the service parameters, you can provide warehouse information.

1. Select the database type for the profiling warehouse.

   The following table lists the database type for the profiling warehouse.

Prompt	Description
Database type	Type of database for the profiling warehouse connection. Select from the following options: 1 - Oracle 2 - Microsoft SQL Server 3 - IBM DB2

2. Enter the properties for the database user account.

   The following table lists the properties for the database user account:

Property	Description
Database user ID	Name for the profiling warehouse database user account.
User password	Password for the profiling warehouse database user account.

3. Based on the database type selected, enter the parameters for the database.
   a. If you select IBM DB2, select whether to configure a tablespace and enter the tablespace name.

      The following table describes the properties that you must configure for the IBM DB2 database:

Property	Description
Configure tablespace	Select whether to specify a tablespace: 1 - No 2 - Yes In a single-partition database, if you select No, the installer creates the tables in the default tablespace. In a multi-partition database, you must select Yes.
Tablespace	Name of the tablespace in which to create the tables. Specify a tablespace that meets the pageSize requirement of 32768 bytes. In a single-partition database, if you select Yes to configure the tablespace, enter the name of the tablespace in which to create the tables. In a multi-partition database, specify the name of the tablespace that resides in the catalog partition of the database.

   b. If you select Microsoft SQL Server, enter the schema name for the database.

      The following table describes the properties that you must configure for the Microsoft SQL Server database:

Property	Description
Schema name	Name of the schema that will contain domain configuration tables. If this parameter is blank, the installer creates the tables in the default schema.

   c. To enter the JDBC connection information using the JDBC URL information, press **1**. To enter the JDBC connection information using a custom JDBC connection string, press **2**.

   d. Enter the JDBC connection information.

      - To enter the connection information using the JDBC URL information, specify the JDBC URL properties.

The following table describes the database connection information:

Prompt	Description
Database address	Host name and port number for the database.
Database service name	Service or database name : - Oracle: Enter the service name. - Microsoft SQL Server: Enter the database name. - IBM DB2: Enter the service name.
Configure JDBC Parameters	Select whether to add additional JDBC parameters to the connection string: 1 - Yes 2 - No If you select Yes, enter the parameters or press Enter to accept the default. If you select No, the installer creates the JDBC connection string without parameters.

- To enter the connection information using a custom JDBC connection string, type the connection string.
  Use the following syntax in the JDBC connection string:
  IBM DB2

  ```
  jdbc:Informatica:db2://host_name:port_no;DatabaseName=
  ```

  Oracle

  ```
  jdbc:Informatica:oracle://host_name:port_no;ServiceName=
  ```

  Microsoft SQL Server

  ```
  jdbc:Informatica:sqlserver://
  host_name:port_no;SelectMethod=cursor;DatabaseName=
  ```

  Microsoft Azure SQL Database

  ```
  jdbc:Informatica:sqlserver://
  host_name:port_no;SelectMethod=cursor;DatabaseName=database_name;SnapshotSeria
  lizable=true;EncryptionMethod=SSL;HostNameInCertificate=*.database.windows.net
  ;ValidateServerCertificate=false
  ```

  Sybase

  ```
  jdbc:Informatica:sybase://host_name:port_no;DatabaseName=
  ```

  Verify that the connection string contains all the connection parameters required by your database system.

e. Enter the data access connection string.

The **Content Management Service Parameters and Database** section appears.

Configure the Content Management Service Parameters and Database

After you configure the profiling warehouse, you can configure the content management service parameters and database properties.

1. Enter configuration parameters for the Content Management Service.

   The following table lists the parameters for the Content Management Service:

Prompt	Description
Content Management Service name	Name of the Content Management Service.

2. Enter the following service parameter information:

Port	Description
HTTP protocol type	Type of connection to the Data Integration Service. Select one of the following options: - HTTP. Requests to the service uses an HTTP connection. - HTTPS. Requests to the service uses a secure HTTP connection.
HTTP port	Port number to used for the Data Integration Service. Default is 9085.
HTTPS port	Port number to used for the Data Integration Service. Default is 9085.

3. Enter database information for the reference data warehouse.

   The following table lists the database information for the reference data warehouse.

Prompt	Description
Database type	Type of database for reference data warehouse. Select from the following options: 1 - Oracle 2 - Microsoft SQL Server 3 - IBM DB2

4. Enter the properties for the database user account.

   The following table lists the properties for the database user account:

Property	Description
Database user ID	Name for the reference data warehouse database user account.
User password	Password for the profiling warehouse database user account.

5. Based on the database type selected, enter the parameters for the database .
   a. If you select IBM DB2, select whether to configure a tablespace and enter the tablespace name.

The following table describes the properties that you must configure for the IBM DB2 database:

Property	Description
Configure tablespace	Select whether to specify a tablespace: 1 - No 2 - Yes In a single-partition database, if you select No, the installer creates the tables in the default tablespace. In a multi-partition database, you must select Yes.
Tablespace	Name of the tablespace in which to create the tables. Specify a tablespace that meets the pageSize requirement of 32768 bytes. In a single-partition database, if you select Yes to configure the tablespace, enter the name of the tablespace in which to create the tables. In a multi-partition database, specify the name of the tablespace that resides in the catalog partition of the database.

b. If you select Microsoft SQL Server, enter the schema name for the database.

The following table describes the properties that you must configure for the Microsoft SQL Server database:

Property	Description
Schema name	Name of the schema that will contain domain configuration tables. If this parameter is blank, the installer creates the tables in the default schema.

c. To enter the JDBC connection information using the JDBC URL information, press **1**. To enter the JDBC connection information using a custom JDBC connection string, press **2**.

d. Enter the JDBC connection information.

- To enter the connection information using the JDBC URL information, specify the JDBC URL properties.

The following table describes the database connection information:

Prompt	Description
Database address	Host name and port number for the database.
Database service name	Service or database name : - Oracle: Enter the service name. - Microsoft SQL Server: Enter the database name. - IBM DB2: Enter the service name.
Configure JDBC Parameters	Select whether to add additional JDBC parameters to the connection string: 1 - Yes 2 - No If you select Yes, enter the parameters or press Enter to accept the default. If you select No, the installer creates the JDBC connection string without parameters.

- To enter the connection information using a custom JDBC connection string, type the connection string.
  Use the following syntax in the JDBC connection string:
  IBM DB2

  ```
  jdbc:Informatica:db2://host_name:port_no;DatabaseName=
  ```

  Oracle

  ```
  jdbc:Informatica:oracle://host_name:port_no;ServiceName=
  ```

  Microsoft SQL Server

  ```
  jdbc:Informatica:sqlserver://
  host_name:port_no;SelectMethod=cursor;DatabaseName=
  ```

  Microsoft Azure SQL Database

  ```
  jdbc:Informatica:sqlserver://
  host_name:port_no;SelectMethod=cursor;DatabaseName=database_name;SnapshotSeria
  lizable=true;EncryptionMethod=SSL;HostNameInCertificate=*.database.windows.net
  ;ValidateServerCertificate=false
  ```

  Sybase

  ```
  jdbc:Informatica:sybase://host_name:port_no;DatabaseName=
  ```

  Verify that the connection string contains all the connection parameters required by your database system.

e. Enter the data access connection string.

The **Cluster and Application Service Options** section appears.

Configure External Cluster Details

After you configure the parameters for the Content Management Service, you can configure the cluster and application service options.

1. Select the cluster type to configure.

   The following table describes the options you can select:

Option	Description
Cloudera	Select to create a cluster configuration for a Cloudera cluster.
Hortonworks	Select to create a cluster configuration for a Hortonworks cluster.
Azure HDInsight	Select to create a cluster configuration for a Azure HDInsight cluster.

2. Select an option to confirm if the cluster uses Kerberos authentication
   a. Press **1** if the cluster does not use Kerberos authentication.
   b. Press **2** if the cluster uses Kerberos authentication.

The **Catalog Service Parameters for the Existing Cluster** section appears.

Configure the Catalog Service for the External Cluster

After you can configure the external cluster, you can configure the catalog service parameters for the existing cluster.

1. Enter the information to configure the Catalog Service parameters for the existing cluster, if the cluster uses Kerberos authentication.

   The following table describes the properties you need to set for configuring the Catalog Service parameters for the existing cluster.

Option	Description
Catalog Service name	Name of the Catalog Service.
Catalog Service port	Port number of the Catalog Service.
Cluster Hadoop distribution URL	URL to access the Hadoop cluster.
Cluster Hadoop distribution URL user	User name to access the Hadoop cluster.
Cluster Hadoop distribution URL password	Password to access the Hadoop cluster.

Option	Description
Service cluster name	Name of the service cluster.
KDC domain name	Domain name of the Kerberos Key Distribution Center.
Keytab location	Location of the Kerberos Key Distribution Center (KDC).
Fully qualified path to the Kerberos configuration file	Location of the fully qualifies path to the Kerberos configuration file.
YARN Queue Name	The YARN scheduler queue name used by the Blaze engine that specifies available resources on a cluster.

2. If you chose cluster type as Others, enter the information to configure the Catalog Service parameters for the existing cluster.

   The following table describes the properties you need to set for configuring the Catalog Service parameters for the existing cluster.

Option	Description
Catalog Service name	Name of the Catalog Service.
Catalog Service port	Port number of the Catalog Service.
Yarn resource manager URI	Applies to external cluster. The service within Hadoop that submits the MapReduce tasks to specific nodes in the cluster. Use the following format:<Hostname>:<Port> Where <host name> is the name or IP address of the Yarn resource manager.- <port number> is the port number on which Yarn resource manager listens for Remote Procedure Calls (RPC).
Yarn resource manager HTTPS or HTTP URI	Applies to external cluster. https or http URI value for the Yarn resource manager.
Yarn resource manager scheduler URII	Applies to external cluster. Scheduler URI value for the Yarn resource manager.
Zookeeper Addresses	Multiple ZooKeeper addresses in a comma-separated list.

Option	Description
HDFS Nodename URI	Applies to external cluster. The URI to access HDFS. Use the following format to specify the NameNode URI in the Cloudera distribution:<Hostname>:<Port> Where - <host name> is the host name or IP address of the NameNode - <port number> is the port number that the NameNode listens for Remote Procedure Calls (RPC).
History Server HTTP URI	Applies to external cluster. Specify a value to generate YARN allocation log files for scanners. Catalog Administrator displays the log URL as part of task monitoring.
Service cluster name	Name of the service cluster.
HDFS Service Name for High Availability	Applies to highly available external cluster. Specify the HDFS service namApplies to both internal and external clusters. Name of the service cluster. Ensure that you have a directory `/Informatica/LDM/<ServiceClusterName>` in HDFS. **Note:** If you do not specify a service cluster name, Enterprise Data Catalog considers DomainName_CatalogServiceName as the default value. You must then have the `/Informatica/LDM/<DomainName>_<CatalogServiceName>` directory in HDFS. Otherwise, Catalog Service might fail.
HDFS Service Principal Name	Applies to Kerberos authentication. Principal name for the HDFS Service.
YARN Service Principal Name	Applies to Kerberos authentication. Principal name for the YARN Service.
KDC domain name	The domain name of the Kerberos Key Distribution Center (KDC).
Keytab location	The location of the Kerberos Key Distribution Center (KDC).
Fully qualified path to the Kerberos configuration file	Location of the fully qualifies path to the Kerberos configuration file.
YARN Queue Name	The YARN scheduler queue name used by the Blaze engine that specifies available resources on a cluster.

3. Select the load type.

   The following table describes the options you can choose.

Option	Description
Demo	Represents single datastore. Used for demo purpose.
Low	Represents one million assets or 30-40 datastores.

Option	Description
Medium	Represents 20 million assets or 200-400 datastores.
High	Represents 50 million assets or 500-100 datastores.

The **Model Repository Database** section appears.

Configure the Catalog Service for the Embedded Cluster

If you chose to use the Embedded cluster, you can configure the Catalog Service for the embedded cluster.

- Configure the Hadoop cluster properties in the dialog box.

  The following table describes the properties:

Property	Description
Gateway User	User name for the Apache Ambari server.
Informatica Cluster Service Name	Name of the Informatica Cluster Service for the internal cluster.
Informatica Cluster Service Port	Port number for the Informatica Cluster Service.
Informatica Hadoop Gateway Host	Host where Apache Ambari server runs.
Informatica Hadoop Nodes	Hosts where the Apache Ambari agents run.
Informatica Hadoop Gateway Port	Web port for the Apache Ambari server.
Override default password	Select this option if you want to change the default password for the cluster.
New Hadoop Ambari Password	Password for the Ambari Hadoop cluster.

Property	Description
Confirm Hadoop Ambari Password	Confirm the password for the Ambari Hadoop cluster.
KDC Type	Select one of the following Kerberos Key Distribution Center (KDC) types if you had selected the **Enable Kerberos Authentication** option: - MIT KDC. Select this option if you want to use MIT KDC. - Active Directory. Select this option if you want to use Active Directory KDC. Specify the following options after you select the **KDC Type** - KDC Host. Name of the KDC host machine. - Administrator Server Host. The name of the administrator server machine that hosts the KDC server. - Realm. Name of the Kerberos realm on the machine that hosts the KDC server. - Administrator Principal. The Kerberos administrator principal. - Administrator Password. The Kerberos administrator password. - LDAP URL. This property applies to Microsoft Active Directory and represents the URL to the LDAP server directory. - Container DN. This property applies to Microsoft Active Directory and represents the Distinguished Name of the container to which the user belongs. - KDC Certificate Path. Path to the KDC certificate on the Informatica domain machine.

Configure Enterprise Data Lake

This task includes installer prompts to configure Enterprise Data Lake. You will provide basic information for configuring the application services, Hadoop cluster, and creating the Enterprise Data Lake service .

When you complete the tasks, you will complete the installation.

Configure the Model Repository Service and Model Repository Database Details

Choose whether to create a Model Repository Service and a Data Integration Service to associate with the Enterprise Data Lake Service, or to associate existing application services with Enterprise Data Lake. If you create a Model Repository Service, specify the connection details for the Model repository database.

If you choose to create a Model Repository Service, specify the connection details for the Model repository database.

1. Choose to either create a Model Repository Service and a Data Integration Service to associate with the Enterprise Data Lake Service, or to associate existing application services with Enterprise Data Lake.
   - To create to create a Model Repository Service and a Data Integration Service to associate with the Enterprise Data Lake Service on the node, press **1**.
   - To associate an existing Model Repository Service and an existing Data Integration Service with the Enterprise Data Lake Service, press **2**, and then enter the name of each service.
2. Enter the name of the Model Repository Service.

   The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [

3. Specify the connection details for the Model repository database.

   The following table describes the parameters you set:

Property	Description
Database Type	Database for the Model repository managed by the Model Repository Service.
Database User ID	User name of the database user account to use to log in to the Model repository database.
User Password	Password for the Model repository database user account.
Tablespace	Configure for a IBM DB2 database. Name of the tablespace in which to create the tables. The tablespace must be defined on a single node, and the page size must be 32K. This option is required for a multi-partition database. If this option is not selected for a single-partition database, the installer creates the tables in the default tablespace.
Schema Name	Configure for a Microsoft SQL Server database. Name of the schema that will contain domain configuration tables. If not selected, the installer creates the tables in the default schema.

4. Specify the truststore details required to access a secure Model repository database.

   The following table describes the properties you set:

Property	Description
Database truststore file	Path and file name of the truststore file for the secure database.
Database truststore password	Password for the truststore.

5. Choose whether to configure the database connection using a JDBC URL or a custom JDBC connection string.
   - Press **1** to configure the database connection using a JDBC URL.

The following table describes the properties you set:

Property	Description
Database address	Host name and port number for the database in the format <host name>:<port>.
Database service name	Service name for Oracle and IBM DB2 databases, or database name for Microsoft SQL Server.
JDBC parameters	Optional parameters to include in the database connection string. Use the parameters to optimize database operations for the Model repository. You can use the default parameters or add or modify the parameters based on your database requirements. Verify that the parameter string is valid. The installer does not validate the parameter string before it adds the string to the JDBC URL. If not selected, the installer creates the JDBC URL without additional parameters. The following examples show the default connection strings for each database: - Oracle. jdbc:Informatica:oracle://host_name:port_no ;ServiceName= - IBM DB2. jdbc:Informatica:db2://host_name:port_no ;DatabaseName= - Microsoft SQL Server. jdbc:Informatica:sqlserver://host_name:port_no ;SelectMethod=cursor;DatabaseName= - Azure SQL Server. jdbc:informatica:sqlserver://host_name:port_number ;DatabaseName=<database_name>;SnapshotSerializable=true;SnapshotSerializable=true;EncryptionMethod=SSL;HostNameInCertificate=*.<hostnameincertificate>;ValidateServerCertificate=true

- Press **2** to configure the database connection using a custom JDBC connection string.
  The following table describes the properties you set:

Property	Description
EncryptionMethod	Indicates whether data is encrypted when transmitted over the network. This parameter must be set to SSL.
ValidateServerCertificate	Indicates whether Informatica validates the certificate that is sent by the database server. If this parameter is set to True, Informatica validates the certificate that is sent by the database server. If you specify the HostNameInCertificate parameter, Informatica also validates the host name in the certificate. If this parameter is set to false, Informatica does not validate the certificate that is sent by the database server. Informatica ignores any truststore information that you specify.
HostNameInCertificate	Host name of the machine that hosts the secure database. If you specify a host name, Informatica validates the host name included in the connection string against the host name in the SSL certificate.

The Service Parameters section appears.

Configure the Application Service Properties

If you create a Data Integration Service to associate with Enterprise Data Lake during installation, specify the properties required to create the application service.

1. Specify the name of the Data Integration Service.

   The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
2. Specify the HTTP protocol type for the Data Integration Service, and then enter the port for each protocol you select.
   - To select HTTP only, press **1**.
   - To select HTTPS only, press **2**.
   - To select both HTTP and HTTPS, press **3**.
3. If you select HTTPS or both HTTP and HTTPS, select the SSL certificate to use.
   - To use the default Informatica SSL certificate contained in the default keystore and the default truststore, press **1**.
   - To use a custom SSL certificate contained in custom keystore and truststore files, press **2**, and then enter the path and file name for the keystore and truststore files. You must also enter the keystore and truststore passwords.

The Data Preparation Repository Database section appears.

Configure the Data Preparation Repository Database Details

Specify the Data Preparation repository database connection details. You can choose to use an Oracle database or a MySQL database for the Data Preparation repository database.

If you do not have the database details, you can enter placeholder values, and then create the Data Preparation Service. If you continue without specifying the database connection details, you cannot enable the Data Preparation Service.

Oracle

1. To use an Oracle database for the Data Preparation repository, press **1**.
2. Enter the connection properties for the database.

   The following table describes the connection properties:

Property	Description
Database Host Name	Host name of the machine that hosts the database.
Database Port Number	Port number for the database.
JDBC Parameters	Parameters required to connect to the database.
Custom JDBC Connection String	JDBC connection string to connect to the database. Format the string as follows: `jdbc:informatica:oracle://<database host name>:<port>;ServiceName=<service name>`

3. To connect to a secure database, press **2**, and then enter the secure connection properties.

The following table describes the secure connection properties:

Property	Description
Truststore File	Path and file name for the database truststore file.
Truststore Password	Password for the database truststore file.
Secure JDBC Parameters	List of secure database parameters to connect to the database. Format the parameters as follows: `EncryptionMethod=SSL;HostNameInCertificate=<secure database host name>;ValidateServerCertificate=true`

4. Press **2** to continue.

The Data Preparation Service Details section appears.

MySQL

1. To use a MySQL database or a MariaDB database for the Data Preparation repository, press **2**.
2. Enter the connection properties for the database.
   The following table describes the connection properties:

Property	Description
Database Host Name	Host name of the machine that hosts the Data Preparation repository database.
Database User Name	Database user account to use to connect to the Data Preparation repository.
Database User Password	Password for the Data Preparation repository database user account.
Database Port Number	Port number for the database.
Database Name	Schema or database name of the Data Preparation repository database.

3. To connect to a secure database, press **2**, and then enter the secure connection properties.

   The following table describes the secure connection properties:

Property	Description

Custom JDBC Connection String	Connection string to connect to the database.
	To connect to a non-secure database, format the string as follows:
	`jdbc:mysql://<database host name>:<port>`
	The connection string is optional if you connect to a non-secure database.
	To connect to an SSL-enabled database, format the string as follows:
	`verifyServerCertificate=true&useSSL=true&requireSSL=true`

Secure JDBC Parameters	String containing the path and file name for the database truststore file, and the truststore password. Format the string as follows: `trustCertificateKeyStoreUrl=file://<truststore path/truststore file name>&trustCertificatekeyStorePassword=<truststore password>`

4. Press **Enter** to continue.

The Data Preparation Service Details section appears.

Create the Data Preparation Service

When you install Enterprise Data Lake on a gateway node, you can create the Data Preparation Service and the Enterprise Data Lake Service during installation.

If you do not create the Enterprise Data Lake Service and the Data Preparation Service during installation, or if you install Enterprise Data Lake on another node in the domain, you can use the Administrator tool to create the services after you install the Enterprise Data Lake binaries.

1. Specify the name of the Data Preparation Service.

   The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
2. If you plan to use rules, you must associate a Model Repository Service and a Data Integration Service with the Data Preparation Service.
   - To skip associating a Model Repository Service and a Data Integration Service with the Enterprise Data Lake Service, press **1**.
   - To associate a Model Repository Service and a Data Integration Service with the Data Preparation Service, press **2**, and then enter the service names.
3. To create the Data Preparation Service during installation, enter the name of the current node.

   If you do not want to create the service during installation, do not enter a value. You can use the Administrator tool to create the service after installation.
4. Choose whether to enable secure communication for the service.
   - To enable secure communication, press **1**.
   - To disable secure communication, press **2**.
5. If you enable secure communication for the service, select the SSL certificate to use.
   - To use the default Informatica SSL certificate contained in the default keystore and the default truststore, press **1**.
   - To use a custom SSL certificate contained in a custom keystore and truststore, press **2**, and then enter the path and file name for the keystore and truststore files. You must also enter the keystore and truststore passwords.
6. If you enable secure communication for the service, enter the port number for the HTTPS connection. If you enable non-secure communication for the service, enter the port number for the HTTP connection.
7. Select the Hadoop authentication mode.
   - To select the non-secure authentication mode, press **1**.
   - To select Kerberos authentication, press **2**.
8. If you select Kerberos, enter the authentication parameters.

The following table describes the authentication parameters that you must set if you select Kerberos:

Property	Description
HDFS Principal Name	Service Principal Name (SPN) for the data preparation Hadoop cluster. Specify the service principal name in the following format: user/_HOST@REALM.
Hadoop Impersonation User Name	User name to use in Hadoop impersonation as shown in the Impersonation User Name property for the Hadoop connection in the Administrator tool. If the Hadoop cluster uses Kerberos authentication, the Hadoop impersonation user must have read, write, and execute permissions on the HDFS storage location folder.
Kerberos Keytab File	Path and file name of the SPN keytab file for the user account to impersonate when connecting to the Hadoop cluster. The keytab file must be in a directory on the machine where the Data Preparation Service runs.

9. Specify the HDFS storage location, HDFS connection, local storage location, and Solr port number details.

   The following table describes the properties you must set:

Property	Description
HDFS Storage Location	HDFS location for data preparation file storage. If the Hadoop cluster uses Kerberos authentication, the Hadoop impersonation user must have read, write, and execute permissions on the HDFS storage location folder.
HDFS Connection	HDFS connection for data preparation file storage.
Local Storage Location	Directory for data preparation file storage on the node on which the Data Preparation Service runs. If the connection to the local storage fails, the Data Preparation Service recovers data preparation files from the HDFS storage location.
Solr port	Solr port number for the Apache Solr server used to provide data preparation recommendations.

10. Choose whether to enable the Data Preparation Service.
    - To enable the service at a later time using the Administrator tool, press **1**.
    - To enable the service after you complete the installation process, press **2**.

The Enterprise Data Lake Service section appears.

Create the Enterprise Data Lake Service

When you install Enterprise Data Lake on the master gateway node for the domain, you can create the Enterprise Data Lake Service and the Data Preparation Service during installation.

If you do not create the Enterprise Data Lake Service and the Data Preparation Service during installation, or if you install Enterprise Data Lake on another gateway node in the domain, you can use the Administrator tool to create the services after you install the Enterprise Data Lake binaries.

1. Specify the details for the Enterprise Data Lake Service.

The following table describes the properties that you set:

Property	Description
Enterprise Data Lake Service Name	Name of the Enterprise Data Lake Service. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
Data Preparation Service Name	Name of the Data Preparation Service to associate with the Enterprise Data Lake Service.
Model Repository Service Name	Name of the Model Repository Service to associate with the Enterprise Data Lake Service.
Data Integration Service Name	Name of the Data Integration Service associated with the Enterprise Data Lake Service.
Node Name	To create the Enterprise Data Lake Service during installation, enter the name of the current node. If you do not want to create the service during installation, do not enter a value. You can use the Administrator tool to create the service after installation. If you create the Enterprise Data Lake Service and the Data Preparation Service during installation, you must create both services on the same node.

2. Choose whether to enable secure communication for the service.
   - To enable secure communication, press **1**.
   - To disable secure communication, press **2**.
3. If you enable secure communication for the service, select the SSL certificate to use.
   - To use the default Informatica SSL certificate contained in the default keystore and the default truststore, press **1**.
   - To use a custom SSL certificate contained in a custom keystore and truststore, press **2**, and then enter the path and file name for the keystore and truststore files. You must also enter the keystore and truststore passwords.
4. If you enable secure communication for the service, enter the port number for the HTTPS connection. If you enable non-secure communication for the service, enter the port number for the HTTP connection.
5. Specify the data lake connection properties.

   The following table describes the properties that you set for the data lake connections:

Property	Description
HDFS Connection	HDFS connection for the data lake. If you selected the option to create the connection when creating the cluster configuration, the installer sets this value to the name created for the connection.
HDFS Working Directory	HDFS directory where the Enterprise Data Lake Service copies temporary data and files necessary for the service to run.

Property	Description
Hadoop Connection	Hadoop connection for the data lake. If you selected the option to create the connection when creating the cluster configuration, the installer sets this value to the name created for the connection.
Hive Connection	Hive connection for the data lake. If you selected the option to create the connection when creating the cluster configuration, the installer sets this value to the name created for the connection.
Hive Table Storage Format	Data storage format for the Hive tables.
Local System Directory	Local directory that contains the files downloaded from Enterprise Data Lake application, such as .csv and .tde files.

6. Choose whether to enable logging of user activity events.
   - To disable logging of user activity events, press **1**.
   - To enable logging of user activity events, press **2**.
7. Select the Hadoop authentication mode.
   - To select the non-secure authentication mode, press **1**.
   - To select Kerberos authentication, press **2**.
8. If you select Kerberos, enter the authentication parameters.
9. The following table describes the authentication parameters that you set if you select Kerberos:

   The following table describes the authentication properties that you set if you select Kerberos:

Property	Description
Kerberos Principal	If the Hadoop cluster uses Kerberos authentication, specify the Service Principal Name (SPN) of the user account to impersonate when connecting to the data lake Hadoop cluster.
Kerberos KeyTab File	If the Hadoop cluster uses Kerberos authentication, specify the path and file name of the SPN keytab file for the user account to impersonate when connecting to the Hadoop cluster. The keytab file must be in a directory on the machine where the Enterprise Data Lake Service runs.

   - To enable the service at a later time using the Administrator tool, press **1**.
   - To enable the service immediately after you create the service, press **2**.

Resume the Installer

When the installation process stops midway, you can resume the installation from the point of failure or exit.

When you run the server installer and the installation process fails, you can still resume from the previous service configuration and recover the last entered details for that service installation.

The install process might fail for reasons such as network outage, when you exit the installation before completing the entire installation process, or because of incorrect information entered.

If a service fails or if the installation process fails during a service creation, you can resume the installation process with the server installer. To resume the installation process, ensure that at least one of the services is created and that the domain is up and running from the installation log. For example, if you want to check whether the Model Repository Service is created, check if you have a service creation success text in the server log in the following format:

```
SUCCESS: MRS Service [mrs_name] is created. Command ran successfully.
```

To resume the installation, run the installer again.

Note: You cannot resume the installer if you are running it to configure services after the services have been created. When you run the service configuration wizard, you cannot resume the installer for Big Data, Enterprise Data Lake, or Enterprise Data Catalog. When you join the domain, you also cannot resume the installer.

When you resume the installer while creating a service, the installer retains all the service and database specific information, such as the create service status, service name, service enabled or disabled status. You can confirm and use the previously entered values or specify new values for the service and resume the installation process.

Resuming the Installer

When the installation process stops midway, you can resume the installation from the point of failure or exit.

Before you can resume the installer, complete the following prerequisites:

- In the installation log file present in the installation directory, verify that at least the domain and one service is created. The installer log file name appears in the following syntax: Informatica_<Version>_Services_<timestamp>.log
- Ensure that you do not delete the installInst.obj object file present in the tools folder of the user installation directory.
- For silent installer, ensure that RESUME_INSTALLATION is set to true in the SilentInput.properties file.

1. Open a command prompt and navigate to the location of the installation files.
2. Run the Installer.

   On Linux, run silentInstall.sh to resume the silent installer. To resume the regular installer, run the ./install.sh command.
3. When the regular installer runs, you might get a prompt confirming whether you want to resume previous installer or not.
   - If you do not want to resume installation, enter 1 for No. Default is 1.
   - If you want to resume installation, enter 2 for Yes.

   Before you can resume the installation, the services get validated.

CHAPTER 9

Join a Domain and Install All Big Data Products

This chapter includes the following topics:

Begin the Installation

This task includes installer prompts to begin the installation. You will provide basic information such as acceptance of terms, installation option, and the installation directory.

When you complete the preliminary tasks, you will continue with the installer prompts and will provide information to configure the domain.

Run the Installer

Perform the following steps to run the installer:

1. Log in to the machine with a system user account.
2. Close all other applications.
3. On a shell command line, run the `install.sh` file from the root directory.

   The installer displays the message to verify that the locale environment variables are set.
4. If the environment variables are not set, press **n** to exit the installer and set them as required.

   If the environment variables are set, press **y** to continue.
5. Press **1** to install Informatica Big Data Suite Products.

   The installer displays different options based on the platform you are installing on.
   The following options appear:

   a. Press **1** to run the Pre-Installation System Check Tool.

      For more information about running the Pre-Installation (i10Pi) System Check Tool, see "Run the Pre-Installation (i10Pi) System Check Tool" on page 35.

   b. Press **2** to run the Informatica Kerberos SPN Format Generator.

For more information about running the Informatica Kerberos SPN Format Generator, see "Running the SPN Format Generator on Linux" on page 69.

c. Press **3** to run the installer.

The **Welcome** section appears.

Accept Terms and Conditions

1. Read the terms and conditions for Informatica installation and the product usage toolkit and select **I agree to the terms and conditions**.
   a. Press **1** if you do not want to accept the terms and conditions.
   b. Press **2** to accept the terms and conditions.
2. Version 10.2.2 is for big data products only, such as Big Data Management and Big Data Quality. This version does not support non-big data products, such as PowerCenter or Informatica Data Quality.
   a. Press **1** and type **quit** to quit the installation.
   b. Press **2** to continue the installation.

   If you choose to not accept the terms and condition, the installer prompts you to accept the terms and conditions.

The **Component Selection** sections appears.

Choose the Installation Option

After you accept terms and conditions, you can install Informatica domain services supporting Big Data Management services, Enterprise Data Catalog, and Enterprise Data Lake.

1. Press **3** to install Informatica Enterprise Data Lake.

   When you select this option, you can choose to install Only Enterprise Data Lake, Enterprise Data Catalog and Enterprise Data Lake, and Informatica domain services, supporting Big Data Management services, Enterprise Data Catalog, and Enterprise Data Lake.
2. Select whether current version of the Informatica domain services is installed on the node.
   a. Press **1** if current version of the Informatica domain services is not installed on the node.
   b. Press **2** if current version of the Informatica domain services is installed on the node.
3. Select the Hadoop cluster type for Enterprise Data Catalog.
   a. Press **1** to select **External** cluster type.
   b. Press **2** to select **Embedded** cluster type.

      If you choose to run on an internal cluster, the installer creates a cluster to run the services.
4. Select whether you have read and accepted terms and conditions to use Java SE Development Kit software.
   a. Press **1** to not accept the terms and conditions to use Java SE Development Kit software.
   b. Press **2** to accept the terms and conditions to use Java SE Development Kit software.

The **Installation Prerequisites** section displays the installation requirements. Verify that all requirements are met before you continue the installation.

Tune the Application Service

After you review the installation prerequisites, you can choose to tune the application services for better performance based on the deployment type in your environment. If you do not tune now, you can tune the services later through infacmd.

1. Select if you want the installer to tune the application services.

Prompt	Description
Select if you want the installer to tune the application services.	Select if you want to tune the services. 1 - No 2 - Yes Select no if you do not want to tune the services. Select yes if you want to tune the services.

   If you are joining the node to existing domain, ensure the deployment type you select here is same deployment type as the gateway nodes.

2. Select the deployment type associated with the Informatica environment.

Prompt	Description
1. Sandbox	Choose this option if the environment is used for proof of concepts or as a sandbox environment with minimal users. Sandbox environments are typically configured with 16 cores, 32 GB RAM, and about 50 GB disk space.
2. Basic	Choose this option if the environment is used for low volume processing environments with low levels of concurrency. Basic environments are typically single- or multi-node setups configured with 24 cores, 64 GB RAM, and about 100 GB disk space.
3. Standard	Choose this option if the environment is used for high volume processing but with low levels of concurrency. Standard environments are typically multi-node setups configured with 64 GB RAM, more than 100 GB disk space per node, and 48 cores across nodes.
4. Advanced	Choose this option if the environment is used for high volume processing with high levels of concurrency. Advanced environments are typically multi-node setups configured with 128 GB RAM, more than 100 GB disk space per node, and 96 cores across nodes.

3. Select whether you want to change the deployment type or continue with the current deployment selection.
   a. Press **1** to change the deployment type.
   b. Press **2** to continue with the current deployment selection.

The **License and Installation Directory** section appears.

Specify the Installation Directory

After you verify the installation prerequisites, you can specify the installation directory.

1. Enter the path to the license key file and press **Enter**.

2. Enter the installation directory.

   The directory names in the path must not contain spaces or the following special characters: @|* $ # ! % ( ) { } [ ] , ; ' Default is /home/toolinst.

   Note: Informatica recommends using alphanumeric characters in the installation directory path. If you use a special character such as á or €, unexpected results might occur at run time.

3. Select whether you want to run the pre-validation utility.

   a. Press **1** to skip the pre-validation utility.

   b. Press **2** to run the pre-validation utility.

   The utility helps you validate the prerequisites to install Enterprise Data Catalog in an embedded cluster. The utility also validates the Informatica domain, cluster hosts, and the Hadoop cluster services configuration.

If you choose to skip the pre-validation utility, the **Pre-Installation Summary** section appears. Review the installation summary.

If choose to run the pre-validation utility, the **Embedded Hadoop Cluster Pre-Validaion** section appears.

Perform Pre-validation for the Embedded Hadoop Cluster

If you chose to use embedded Hadoop cluster, you can configure the embedded Hadoop cluster.

1. Select whether you want to enable Kerberos authentication for the cluster.

   a. Press **1** to configure the Hadoop cluster to run on a network that does not use Kerberos authentication.

   b. Press **2** to configure the Hadoop cluster to run on a network that uses Kerberos authentication.

   If you chose to run on a network that uses Kerberos authentication, enter the Kerberos information.

Property	Description
Keytab Location	Location of the merged user and host keytab file.
Kerberos Configuration File	Location of the Kerberos configuration file.

2. Enter the gateway user name and press **Enter**. Default is **root**.
3. Enter the Informatica Hadoop cluster gateway hostname in the following format: `<hostname>.<FQDN>` and press **Enter**.
4. Enter the list of comma-separated Informatica Hadoop cluster nodes as shown in the following format: `<hostname>.<FQDN>, <hostname1>.<FQDN>, <hostname2>.<FQDN>` and press **Enter**.
5. Enter the Informatica Hadoop cluster gateway port and press **Enter**. Default is **8080**.

   Make sure that you do not configure Oracle with port 8080 on the same machine where Informatica Cluster Service runs.

6. Enter the path to the working directory and press **Enter**. The path indicates the location where you want to mount the Informatica Cluster Service.

   The installer starts the pre-validation utility.

The **Pre-Installation Summary** section appears. Review the installation summary.

Configure the Domain

This task includes installer prompts to configure the domain. You will provide information to join a domain, configure the domain security, domain repository, and the encryption key for the domain.

When you complete the tasks, you will complete the installation.

Configure the Domain

After you review the Pre-Installation summary, you can enter the domain information.

1. Press **2** to join a domain.

   The installer joins a node on the machine where you install.
2. Specify whether the domain you want to join has the secure communication option enabled.

   Press 1 to join an unsecure domain or press 2 to join a secure domain.
3. Select the type of node you want to create.

   The following table describes that types of nodes that you can create:

Property	Description
Configure this node as a gateway	Select whether to configure the node as a gateway or worker node. 1 - Yes 2 - No Select 1 to configure a gateway node or 2 to configure a worker node.

   If you configure the node as a gateway, you can enable a secure HTTPS connection to the Informatica Administrator.

4. Specify the connection details to Informatica Administrator.
   a. Specify whether to set up a secure HTTPS connection to the Informatica Administrator.

Option	Description
1 - Enable HTTPS for Informatica Administrator	Set up a secure connection to Informatica Administrator.
2 - Disable HTTPS	Do not set up a secure connection to Informatica Administrator.

b. If you enable HTTPS connection for the Informatica Administrator, enter the keystore file and port number to use to secure the connection.

Option	Description
Port	Port number for the HTTPS connection.
Keystore file	Select whether to use a keystore file generated by the installer or a keystore file you create. You can use a keystore file with a self-signed certificate or a certificate signed by a certification authority. 1 - Use a keystore generated by the installer 2 - Specify a keystore file and password If you select to use a keystore file generated by the installer, the installer creates a self-signed keystore file named Default.keystore in the following location: `<Informatica installation directory>/tomcat/conf/`

c. If you specify the keystore, enter the password and location of the keystore file.

d. If you enabled secure communication for the domain, the **Domain Security - Secure Communication** section appears.

e. If you did not enable secure communication for the domain, the **Domain Configuration Repository** section appears. Skip to "Configure the Domain Repository" on page 158.

5. Select if SAML authentication is enabled to configure Security Assertion Markup Language (SAML)-based single sign-on (SSO) support for web-based Informatica applications in an Informatica domain.

   The following table describes the information you must enter to enable SAML authentication:

Prompt	Description
Does the domain use SAML authentication?	Select if the domain uses SAML authentication: 1 - No If you select No, skip to "Domain Security" on page 157 2 - Yes If you select Yes, configure the SAML authentication.

6. Enter the Identity Provider URL for the domain.
7. Enter the identity provider assertion signing certificate alias name.
8. Select whether to use the default Informatica SSL certificates or to use your SSL certificates to enable SAML authentication in the domain.

   The following table describes the SSL certificate options for SAML authentication:

Option	Description
Use the default Informatica SSL certificate file	Select to use the default Informatica truststore file for SAML authentication.
Enter the location of the SSL certificate file	Select to use a custom truststore file for SAML authentication. Specify the directory containing the custom truststore file on gateway nodes within the domain. Specify the directory only, not the full path to the file.

9. If you provide the security certificates, specify the location and passwords of the keystore and truststore files.

   The following table describes the location and password of the truststore file:

Property	Description
Truststore Directory	Specify the directory containing the custom truststore file on gateway nodes within the domain. Specify the directory only, not the full path to the file.
Truststore Password	The password for the custom truststore file.

The **Domain Security - Secure Communication** appears.

Domain Security

After you configure the domain, you can configure domain security.

- In the Domain Security - Secure Communication section, specify whether to use the default Informatica SSL certificates or to use your SSL certificates to secure domain communication.

  a. Select the type of SSL certificates to use.

     The following table describes the options for the SSL certificates that you can use to secure the Informatica domain:

Option	Description
Use the default Informatica SSL certificates	Use the default SSL certificates contained in the default keystore and truststore. **Note:** If you do not provide an SSL certificate, Informatica uses the same default private key for all Informatica installations. If you use the default Informatica keystore and truststore files, the security of your domain could be compromised. To ensure a high level of security for the domain, select the option to specify the location of the SSL certificate files.
Use custom SSL certificates	Specify the path for the keystore and truststore files that contain the SSL certificates. You must also specify the keystore and truststore passwords. You can provide a self-signed certificate or a certificate issued by a certificate authority (CA). You must provide SSL certificates in PEM format and in Java Keystore (JKS) files. Informatica requires specific names for the SSL certificate files for the Informatica domain. You must use the same SSL certificates for all nodes in the domain. Store the truststore and keystore files in a directory accessible to all the nodes in the domain and specify the same keystore file directory and truststore file directory for all nodes in the same domain.

  b. If you provide the SSL certificate, specify the location and passwords of the keystore and truststore files.

The following table describes the parameters that you must enter for the SSL certificate files:

Property	Description
Keystore file directory	Directory that contains the keystore files. The directory must contain files named infa_keystore.jks.
Keystore password	Password for the keystore infa_keystore.jks.
Truststore file directory	Directory that contains the truststore files. The directory must contain files named infa_truststore.jks and infa_truststore.pem.
Truststore password	Password for the infa_truststore.jks file.

The **Domain Configuration Repository** section appears.

Configure the Domain Repository

After you configure the domain, you can configure domain repository.

- Enter the information for the domain that you want to join.

  The following table describes the properties that you specify for the domain:

Property	Description
Domain name	Name of the domain to join.
Gateway node host	Host name of the machine that hosts the gateway node for the domain.
Gateway node port	Port number of the gateway node.
Domain user name	User name of the administrator for the domain you want to join.
Domain password	Password for the domain administrator.

The **Domain Security - Encryption Key** section appears.

Configure the Encryption Key

After you configure domain repository, you can configure encryption key.

- In the **Domain Security - Encryption Key** section, enter the directory for the encryption key for the Informatica domain.

The following table describes the encryption key parameters that you must specify when you join a domain:

Property	Description
Select the encryption key	Path and file name of the encryption key for the Informatica domain that you want to join. All nodes in the Informatica domain use the same encryption key. You must specify the encryption key file created on the gateway node for the domain that you want to join. If you copied the encryption key file to a temporary directory to make it accessible to the nodes in the domain, specify the path and file name of the encryption key file in the temporary directory.
Encryption key directory	Directory in which to store the encryption key on the node created during this installation. The installer copies the encryption key file for the domain to the encryption key directory on the new node.

The installer sets different permissions to the directory and the files in the directory. For more information about the permissions for the encryption key file and directory, see "Secure Files and Directories" on page 104.

The **Domain and Node Configuration** section appears.

Configure the Domain and Node

After you configure the encryption key, you can configure the domain and node.

1. Enter the information for the domain and the node that you want to join.

   The following table describes the properties that you set for the domain and gateway node.

Property	Description
Node Host name	Host name for the node. The node host name cannot contain the underscore (_) character.Note: Do not use localhost. The host name must explicitly identify the machine.
Node name	Name of the node to join.
Node port number	Port number for the node. The default port number for the node is 6005. If the port number is not available on the machine, the installer displays the next available port number.
Database truststore file	Path and file name of the truststore file for the secure database. Select the same database truststore file used by the master gateway node in the domain. Available when you join a gateway node to a domain that uses a domain configuration repository database that is secured with the SSL protocol.
Truststore password	Password for the database truststore file for the secure database. Available when you join a gateway node to a domain that uses a domain configuration repository database that is secured with the SSL protocol.

2. Select whether to display the default ports for the domain and node components assigned by the installer.

The following table describes the advanced port configuration page:

Prompt	Description
Display advanced port configuration page	Select whether to display the port numbers for the domain and node components assigned by the installer: 1 - No 2 - Yes If you select Yes, the installer displays the default port numbers assigned to the domain components. You can specify the port numbers to use for the domain and node components. You can also specify a range of port numbers to use for the service process that will run on the node. You can use the default port numbers or specify new port numbers. Verify that the port numbers you enter are not used by other applications.

3. If you display the port configuration page, enter new port numbers at the prompt or press Enter to use the default port numbers.

   The following table describes the ports that you can set:

Port	Description
Service Manager port	Port number used by the Service Manager on the node. The Service Manager listens for incoming connection requests on this port. Client applications use this port to communicate with the services in the domain. The Informatica command line programs use this port to communicate to the domain. This is also the port for the SQL data service JDBC/ODBC driver. Default is 6006.
Service Manager Shutdown port	Port number that controls server shutdown for the domain Service Manager. The Service Manager listens for shutdown commands on this port. Default is 6007.
Informatica Administrator port	Port number used by Informatica Administrator. Default is 6008.
Informatica Administrator HTTPS port	No default port. Enter the required port number when you create the service. Setting this port to 0 disables an HTTPS connection to the Administrator tool.
Informatica Administrator shutdown port	Port number that controls server shutdown for Informatica Administrator. Informatica Administrator listens for shutdown commands on this port. Default is 6009.
Minimum port number	Lowest port number in the range of dynamic port numbers that can be assigned to the application service processes that run on this node. Default is 6014.
Maximum port number	Highest port number in the range of dynamic port numbers that can be assigned to the application service processes that run on this node. Default is 6114.

The **Post-Installation Summary** section indicates whether the installation completed successfully. The summary also shows the status of the installed components and their configuration.

CHAPTER 10

Install Informatica Services

This chapter includes the following topics:

Informatica Services Installation Overview

You can install the Informatica services on multiple Linux machines. The installation process creates a service named Informatica that runs as a daemon.

The first time you run the installer, you create a domain. If you are install on multiple machines and you have created a domain, you join the domain.

When you create a domain, the node on the machine where you install becomes a gateway node in the domain. You can choose to set up secure communication between services within the domain. You can also choose create a Model Repository Service and a Data Integration Service during the installation process.

When you join a domain, you can configure the node that you create to be a gateway node. When you create a gateway node, you can select enable a secure HTTPS connection to Informatica Administrator.

Note: When you run the installer in console mode, the words Quit and Back are reserved words. Do not use them as input text.

Create a Domain

Create a domain if you are installing for the first time or if you want to administer nodes in separate domains.

Run the Installer

Perform the following steps to run the installer:

1. Log in to the machine with a system user account.
2. Close all other applications.

3. On a shell command line, run the `install.sh` file from the root directory.

   The installer displays the message to verify that the locale environment variables are set.

4. If the environment variables are not set, press **n** to exit the installer and set them as required.

   If the environment variables are set, press **y** to continue.

5. Press **1** to install Informatica Big Data Suite Products.

   The installer displays different options based on the platform you are installing on.
   The following options appear:

   a. Press **1** to run the Pre-Installation System Check Tool.

      For more information about running the Pre-Installation (i10Pi) System Check Tool, see "Run the Pre-Installation (i10Pi) System Check Tool" on page 35.

   b. Press **2** to run the Informatica Kerberos SPN Format Generator.

      For more information about running the Informatica Kerberos SPN Format Generator, see "Running the SPN Format Generator on Linux" on page 69.

   c. Press **3** to run the installer.

The **Welcome** section appears.

Accept Terms and Conditions

1. Read the terms and conditions for Informatica installation and the product usage toolkit and select **I agree to the terms and conditions**.

   a. Press **1** if you do not want to accept the terms and conditions.

   b. Press **2** to accept the terms and conditions.

2. Version 10.2.2 is for big data products only, such as Big Data Management and Big Data Quality. This version does not support non-big data products, such as PowerCenter or Informatica Data Quality.

   a. Press **1** and type **quit** to quit the installation.

   b. Press **2** to continue the installation.

   If you choose to not accept the terms and condition, the installer prompts you to accept the terms and conditions.

The **Component Selection** sections appears.

Install Informatica Domain Services

After you accept terms and conditions, you can install Informatica domain services.

1. Press **1** to install Informatica domain services.

   This option installs version 10.2.2 domain services and the application service binaries to support Big Data Management and Big Data Streaming.

2. Choose whether you want to run the installer on a network that uses Kerberos authentication.

   a. Press **1** to configure the Informatica domain to run on a network that does not use Kerberos authentication.

   b. Press **2** to configure the Informatica domain to run on a network with Kerberos authentication.

The **Installation Prerequisites** section displays the installation requirements. Verify that all requirements are met before you continue the installation.

Tune the Application Services

After you review the installation prerequisites, you can tune the application services for better performance based on the deployment type in your environment. If you do not tune now, you can tune the services later through infacmd.

1. Select if you want to tune the services now.

Prompt	Description
Do you want to tune the services now?	Select if you want to tune the services. 1 - No 2 - Yes Select no if you do not want to tune the services. Select yes if you want to tune the services.

If you are joining the node to existing domain, ensure the deployment type you select here is same deployment type as the gateway nodes.

2. Select the deployment type associated with the Informatica environment.

Prompt	Description
1. Sandbox	Choose this option if the environment is used for proof of concepts or as a sandbox environment with minimal users. Sandbox environments are typically configured with 16 cores, 32 GB RAM, and about 50 GB disk space.
2. Basic	Choose this option if the environment is used for low volume processing environments with low levels of concurrency. Basic environments are typically single- or multi-node setups configured with 24 cores, 64 GB RAM, and about 100 GB disk space.
3. Standard	Choose this option if the environment is used for high volume processing but with low levels of concurrency. Standard environments are typically multi-node setups configured with 64 GB RAM, more than 100 GB disk space per node, and 48 cores across nodes.
4. Advanced	Advanced Choose this option if the environment is used for high volume processing with high levels of concurrency. Advanced environments are typically multi-node setups configured with 128 GB RAM, more than 100 GB disk space per node, and 96 cores across nodes.

The **License and Installation Directory** section appears.

Specify Installation Directory

After you verify the installation prerequisites, you can specify the installation directory.

1. Enter the path and file name of the Informatica license key and press **Enter**.
2. Enter the absolute path for the installation directory.

   The directory names in the path must not contain spaces or the following special characters: @|* $ # ! % ( ) { } [ ] , ; ' Default is /home/toolinst.

   Note: Informatica recommends using alphanumeric characters in the installation directory path. If you use a special character such as á or €, unexpected results might occur at run time.

If you enabled Kerberos network authentication, the **Service Principal Level** section appears.

If you did not enable Kerberos network authentication, the **Pre-Installation Summary** section appears. Review the installation information and press **Enter** to continue. Skip to "Configure the Domain Options" on page 165.

Configure Security Level

After you specify the installation directory, you can configure security level.

- In the **Service Principal Level** section, select the level at which to set the Kerberos service principals for the domain.

  Note: All nodes in the domain must use the same service principal level. When you join a node to a domain, select the same service principal level used by the gateway node in the domain.

  The following table describes the levels that you can select:

Level	Description
Process Level	Configures the domain to use a unique service principal name (SPN) and keytab file for each node and each application service on a node. The number of SPNs and keytab files required for each node depends on the number of application service processes that run on the node. Use the process level option for domains that require a high level of security, such as productions domains.
Node Level	Configures the domain to share SPNs and keytab files on a node. This option requires one SPN and keytab file for the node and all application services that run on the node. It also requires a separate SPN and keytab file for all HTTP processes on the node. Use the node level option for domains that do not require a high level of security, such as test and development domains.

The **Network Security - Kerberos Authentication** section appears.

Configure Kerberos Authentication

After you configure the security level, you can configure Keberos authentication.

- In the **Network Security - Kerberos Authentication** section, enter the parameters required for Kerberos authentication.

  The following table describes the Kerberos authentication parameters that you must set:

Property	Description
Domain name	Name of the domain. The name must not exceed 128 characters and must be 7-bit ASCII only. The name cannot contain a space or any of the following characters: ` % * + ; " ? , < > \ /
Node name	Name of the Informatica node.

Property	Description
Node host name	Fully qualified host name or the IP address of the machine on which to create the node. The node host name cannot contain the underscore (_) character. **Note:** Do not use *localhost*. The host name must explicitly identify the machine.
Service realm name	Name of the Kerberos realm to which the Informatica domain services belong. The realm name must be in uppercase. The service realm name and the user realm name must be the same.
User realm name	Name of the Kerberos realm to which the Informatica domain users belong. The realm name must be in uppercase. The service realm name and the user realm name must be the same.
Keytab directory	Directory where all keytab files for the Informatica domain are stored. The name of a keytab file in the Informatica domain must follow a format set by Informatica.
Fully qualified path to the kerberos configuration file	Path and file name of the Kerberos configuration file. Informatica requires the following name for the Kerberos configuration file: *krb5.conf*

Important: If you configure the domain to run with Kerberos authentication, the domain and node name and the node host name must match the names you specified when you ran the Informatica Kerberos SPN Format Generator to generate SPN and keytab file names. If you use a different domain, node, or host name, generate the SPN and keytab file names again and ask the Kerberos administrator to add the new SPN to the Kerberos principal database and create the keytab files.

The **Pre-Installation Summary** section appears. Review the installation information.

Configure the Domain Options

After you review the Pre-Installation summary, you can enter the domain information.

1. Press **1** to create a domain.

   When you create a domain, the node that you create becomes a gateway node in the domain. The gateway node contains a Service Manager that manages all domain operations.
2. Select whether you want to enable secure communication for services in the domain.
   a. Press **1** to disable secure communication for the domain.
   b. Press **2** to enable secure communication for the domain.

   By default, if you enable secure communication for the domain, the installer sets up an HTTPS connection for the Informatica Administrator. You can also create a domain configuration repository on a secure database.
3. Specify the connection details for Informatica Administrator.
   a. If you do not enable secure communication for the domain, you can specify whether to set up a secure HTTPS connection for the Informatica Administrator.

The following table describes the options available to enable or disable a secure connection to Informatica Administrator:

Option	Description
Enable HTTPS for Informatica Administrator	Set up a secure connection to Informatica Administrator.
Disable HTTPS	Do not set up a secure connection to Informatica Administrator.

b. If you enable secure communication for the domain or if you enable HTTPS connection for the Informatica Administrator, enter the keystore file and port number for the HTTPS connection to Informatica Administrator.

The following table describes the connection information you must enter if you enable HTTPS:

Option	Description
Port	Port number for the HTTPS connection.
Keystore file	Select whether to use a keystore file generated by the installer or a keystore file you create. You can use a keystore file with a self-signed certificate or a certificate signed by a certification authority. 1 - Use a keystore generated by the installer 2 - Specify a keystore file and password If you select to use a keystore file generated by the installer, the installer creates a self-signed keystore file named Default.keystore in the following location: `<Informatica installation directory>/tomcat/conf/`

c. If you specify the keystore, enter the password and location of the keystore file.

d. If you enabled secure communication for the domain, the **Domain Security - Secure Communication** section appears.

e. If you did not enable secure communication for the domain, the **Domain Configuration Repository** section appears. Skip to "Configure the Domain Repository" on page 168.

4. Select whether to enable SAML authentication to configure Security Assertion Markup Language (SAML)-based single sign-on (SSO) support for web-based Informatica applications in an Informatica domain.

The following table describes the information you must enter to enable SAML authentication:

Prompt	Description
Enable SAML authentication	Select whether to enable SAML authentication: 1 - No If you select No, skip to "Configure Domain Security" on page 167. 2 - Yes If you select Yes, configure the SAML authentication.

5. Enter the Identity Provider URL for the domain.
6. Enter the identity provider assertion signing certificate alias name.

7. Select whether to use the default Informatica SSL certificates or to use your SSL certificates to enable SAML authentication in the domain.

   The following table describes the SSL certificate options for SAML authentication:

Option	Description
Use the default Informatica SSL certificate file.	Select to use the default Informatica truststore file for SAML authentication.
Enter the location of the SSL certificate file.	Select to use a custom truststore file for SAML authentication. Specify the directory containing the custom truststore file on gateway nodes within the domain. Specify the directory only, not the full path to the file.

8. If you provide the security certificates, specify the location and passwords of the keystore and truststore files.

   The following table describes the location and password of the truststore file:

Property	Description
Truststore Directory	Specify the directory containing the custom truststore file on gateway nodes within the domain. Specify the directory only, not the full path to the file.
Truststore Password	The password for the custom truststore file.

The **Configure Domain Security** appears.

Configure Domain Security

After you configure the domain, you can configure domain security.

- In the Domain Security - Secure Communication section, specify whether to use the default Informatica SSL certificates or to use your SSL certificates to secure domain communication.

  a. Select the type of SSL certificates to use.

The following table describes the options for the SSL certificates that you can use to secure the Informatica domain:

Option	Description
Use the default Informatica SSL certificates	Use the default SSL certificates contained in the default keystore and truststore. **Note:** If you do not provide an SSL certificate, Informatica uses the same default private key for all Informatica installations. If you use the default Informatica keystore and truststore files, the security of your domain could be compromised. To ensure a high level of security for the domain, select the option to specify the location of the SSL certificate files.
Use custom SSL certificates	Specify the path for the keystore and truststore files that contain the SSL certificates. You must also specify the keystore and truststore passwords. You can provide a self-signed certificate or a certificate issued by a certificate authority (CA). You must provide SSL certificates in PEM format and in Java Keystore (JKS) files. Informatica requires specific names for the SSL certificate files for the Informatica domain. You must use the same SSL certificates for all nodes in the domain. Store the truststore and keystore files in a directory accessible to all the nodes in the domain and specify the same keystore file directory and truststore file directory for all nodes in the same domain.

b. If you provide the SSL certificate, specify the location and passwords of the keystore and truststore files.

The following table describes the parameters that you must enter for the SSL certificate files:

Property	Description
Keystore file directory	Directory that contains the keystore files. The directory must contain files named infa_keystore.jks.
Keystore password	Password for the keystore infa_keystore.jks.
Truststore file directory	Directory that contains the truststore files. The directory must contain files named infa_truststore.jks and infa_truststore.pem.
Truststore password	Password for the infa_truststore.jks file.

The **Domain Configuration Repository** section appears.

Configure the Domain Repository

After you configure domain security, you can configure domain repository details.

1. Select the database to use for the domain configuration repository details.

The following table lists the databases you can use for the domain configuration repository:

Prompt	Description
Database type	Type of database for the domain configuration repository. Select from the following options: 1 - Oracle 2 - Microsoft SQL Server 3 - IBM DB2 4 - Sybase ASE

The Informatica domain configuration repository stores metadata for domain operations and user authentication. The domain configuration repository must be accessible to all gateway nodes in the domain.

2. Enter the properties for the database user account.

   The following table lists the properties for the database user account:

Property	Description
Database user ID	Name for the domain configuration database user account.
User password	Password for the domain configuration database user account.

3. Select whether to create a secure domain configuration repository.

   You can create a domain configuration repository in a database secured with the SSL protocol. To create a domain configuration repository in a secure database, press 1 and skip to step to create a domain configuration repository.

   To create a domain configuration repository in an unsecure database, press 2.

4. If you do not create a secure domain configuration repository, enter the parameters for the database.

   a. If you select IBM DB2, select whether to configure a tablespace and enter the tablespace name.

      The following table describes the properties that you must configure for the IBM DB2 database:

Property	Description
Configure tablespace	Select whether to specify a tablespace: 1 - No 2 - Yes In a single-partition database, if you select No, the installer creates the tables in the default tablespace. In a multi-partition database, you must select Yes.
Tablespace	Name of the tablespace in which to create the tables. Specify a tablespace that meets the pageSize requirement of 32768 bytes. In a single-partition database, if you select Yes to configure the tablespace, enter the name of the tablespace in which to create the tables. In a multi-partition database, specify the name of the tablespace that resides in the catalog partition of the database.

b. If you select Microsoft SQL Server, enter the schema name for the database.

The following table describes the properties that you must configure for the Microsoft SQL Server database:

Property	Description
Schema name	Name of the schema that will contain domain configuration tables. If this parameter is blank, the installer creates the tables in the default schema.

c. To enter the JDBC connection information using the JDBC URL information, press **1**. To enter the JDBC connection information using a custom JDBC connection string, press **2**.

d. Enter the JDBC connection information.

- To enter the connection information using the JDBC URL information, specify the JDBC URL properties.
  The following table describes the database connection information:

Prompt	Description
Database host name	Host name for the database.
Database port number	Port number for the database.
Database service name	Service or database name : - Oracle: Enter the service name. - Microsoft SQL Server: Enter the database name. - IBM DB2: Enter the service name. - Sybase ASE: Enter the database name.
Configure JDBC Parameters	Select whether to add additional JDBC parameters to the connection string: 1 - Yes 2 - No If you select Yes, enter the parameters or press Enter to accept the default. If you select No, the installer creates the JDBC connection string without parameters.

- To enter the connection information using a custom JDBC connection string, type the connection string.
  Use the following syntax in the JDBC connection string:
  IBM DB2

```
jdbc:Informatica:db2://host_name:port_no;DatabaseName=
```

  Oracle

```
jdbc:Informatica:oracle://host_name:port_no;ServiceName=
```

  Microsoft SQL Server

```
jdbc:Informatica:sqlserver://
host_name:port_no;SelectMethod=cursor;DatabaseName=
```

Microsoft Azure SQL Database

```
jdbc:Informatica:sqlserver://
host_name:port_no;SelectMethod=cursor;DatabaseName=database_name;SnapshotSeria
lizable=true;EncryptionMethod=SSL;HostNameInCertificate=*.database.windows.net
;ValidateServerCertificate=false
```

Sybase

```
jdbc:Informatica:sybase://host_name:port_no;DatabaseName=
```

Verify that the connection string contains all the connection parameters required by your database system.

5. If you create a secure domain configuration repository, enter the parameters for the secure database.

   If you create the domain configuration repository on a secure database, you must provide the truststore information for the database. You must also provide a JDBC connection string that includes the security parameters for the database.

   The following table describes the options available to create a secure domain configuration repository database:

Property	Description
Database truststore file	Path and file name of the truststore file for the secure database.
Database truststore password	Password for the truststore file.
Custom JDBC Connection String	JDBC connection string to connect to the secure database, including the host name and port number and the security parameters for the database.

   In addition to the host name and port number for the database server, you must include the following secure database parameters: You can use the following syntax for the connection strings:

   EncryptionMethod

   Required. Indicates whether data is encrypted when transmitted over the network. This parameter must be set to`SSL`.

   ValidateServerCertificate

   Optional. Indicates whether Informatica validates the certificate that the database server sends.

   If this parameter is set to True, Informatica validates the certificate that the database server sends. If you specify the HostNameInCertificate parameter, Informatica also validates the host name in the certificate.

   If this parameter is set to False, Informatica does not validate the certificate that the database server sends. Informatica ignores any truststore information that you specify

   Default is True.

   HostNameInCertificate

   Optional. Host name of the machine that hosts the secure database. If you specify a host name, Informatica validates the host name included in the connection string against the host name in the SSL certificate.

cryptoProtocolVersion

Required. Specifies the cryptographic protocol to use to connect to a secure database. You can set the parameter to `cryptoProtocolVersion=TLSv1.1` or `cryptoProtocolVersion=TLSv1.2` based on the cryptographic protocol used by the database server:

- Oracle: `jdbc:Informatica:oracle://host_name:port_no;ServiceName=service_name;EncryptionMethod=SSL;HostNameInCertificate=DB_host_name;ValidateServerCertificate=true_or_false`
- IBM DB2: `jdbc:Informatica:db2://host_name:port_no;DatabaseName=database_name;EncryptionMethod=SSL;HostNameInCertificate=DB_host_name;ValidateServerCertificate=true_or_false`
- Microsoft SQL Server: `jdbc:Informatica:sqlserver://host_name:port_no;SelectMethod=cursor;DatabaseName=database_name;EncryptionMethod=SSL;HostNameInCertificate=DB_host_name;ValidateServerCertificate=true_or_false`

Note: The installer does not validate the connection string. Verify that the connection string contains all the connection parameters and security parameters required by your database.

6. If the database contains a domain configuration repository for a previous domain, select to overwrite the data or set up another database.

   The following table describes the options of overwriting the data or setting up another database when you create a domain configuration repository for a previous domain:

Option	Description
1 - OK	Enter the connection information for a new database.
2 - Continue	The installer overwrites the data in the database with new domain configuration.

The **Domain Security - Encryption Key** section appears.

Configure the Encryption Key

After you configure domain repository, you can configure encryption key.

▶ In the **Domain Security - Encryption Key** section, enter the keyword and directory for the encryption key for the Informatica domain.

The following table describes the encryption key parameters that you must specify when you create a domain:

Property	Description
Keyword	Keyword to use to create a custom encryption key to secure sensitive data in the domain. The keyword must meet the following criteria: - From 8 to 20 characters long - Includes at least one uppercase letter - Includes at least one lowercase letter - Includes at least one number - Does not contain spaces The encryption key is created based on the keyword that you provide when you create the Informatica domain.
Encryption key directory	Directory in which to store the encryption key for the domain. By default, the encryption key is created in the following directory: `<Informatica installation directory>/isp/config/keys`.

The installer sets different permissions to the directory and the files in the directory. For more information about the permissions for the encryption key file and directory, see "Secure Files and Directories" on page 104.

The **Domain and Node Configuration** section appears.

Configure the Domain and Node

After you configure the encryption key, you can configure the domain and node.

1. Enter the information for the domain and the node that you want to create.

   The following table describes the properties that you set for the domain and gateway node.

Property	Description
Domain name	Name of the Informatica domain to create. The default domain name is Domain_<MachineName>. The name must not exceed 128 characters and must be 7-bit ASCII only. The name cannot contain a space or any of the following characters: `` ` % * + ; " ? , < > \ / ``
Node name	Name of the node to create.
Node host name	Host name or IP address of the machine on which to create the node. If the machine has a single network name, use the default host name. If the a machine has multiple network names, you can modify the default host name to use an alternate network name. **Note:** The node host name cannot contain the underscore (_) character. Do not use localhost. The host name must explicitly identify the machine.

Property	Description
Node port number	Port number for the node. The default port number for the node is 6005. If the port number is not available on the machine, the installer displays the next available port number.
Domain user name	User name for the domain administrator. You can use this user name to initially log in to Informatica Administrator. Use the following guidelines: - The name is not case sensitive and cannot exceed 128 characters. - The name cannot include a tab, newline character, or the following special characters: % * + / ? ; < >. - The name can include an ASCII space character except for the first and last character. Other space characters are not allowed.

2. Select whether you want to enable password complexity to secure sensitive data in the domain.

   The following table describes the password complexity:

Prompt	Description
Password complexity	Select whether you want to enable password complexity. 1 - Yes 2 - No If you select Yes, the password must meet the following requirements: It must be at least eight characters long and contain at least one alpha character, one numeric character, and one special character.
Domain password	Password for the domain administrator. The password must be more than 2 characters and must not exceed 16 characters. Not available if you configure the Informatica domain to run on a network with Kerberos authentication.
Confirm password	Enter the password again to confirm. Not available if you configure the Informatica domain to run on a network with Kerberos authentication.

3. Select whether to display the default ports for the domain and node components assigned by the installer.

   The following table describes the advanced port configuration page:

Prompt	Description
Display advanced port configuration page	Select whether to display the port numbers for the domain and node components assigned by the installer: 1 - No 2 - Yes If you select Yes, the installer displays the default port numbers assigned to the domain components. You can specify the port numbers to use for the domain and node components. You can also specify a range of port numbers to use for the service process that will run on the node. You can use the default port numbers or specify new port numbers. Verify that the port numbers you enter are not used by other applications.

4. If you display the port configuration page, enter new port numbers at the prompt or press Enter to use the default port numbers.

   The following table describes the ports that you can set:

Port	Description
Service Manager port	Port number used by the Service Manager on the node. The Service Manager listens for incoming connection requests on this port. Client applications use this port to communicate with the services in the domain. The Informatica command line programs use this port to communicate to the domain. This is also the port for the SQL data service JDBC/ODBC driver. Default is 6006.
Service Manager Shutdown port	Port number that controls server shutdown for the domain Service Manager. The Service Manager listens for shutdown commands on this port. Default is 6007.
Informatica Administrator port	Port number used by Informatica Administrator. Default is 6008.
Informatica Administrator HTTPS port	No default port. Enter the required port number when you create the service. Setting this port to 0 disables an HTTPS connection to the Administrator tool.
Informatica Administrator shutdown port	Port number that controls server shutdown for Informatica Administrator. Informatica Administrator listens for shutdown commands on this port. Default is 6009.
Minimum port number	Lowest port number in the range of dynamic port numbers that can be assigned to the application service processes that run on this node. Default is 6014.
Maximum port number	Highest port number in the range of dynamic port numbers that can be assigned to the application service processes that run on this node. Default is 6114.

5. Select if you want to **Configure the Model Repository Service and Data Integration Service**.

   The following table describes the option to configure the application services:

Prompt	Description
Configure the Model Repository Service and Data Integration Service	Select whether you want to configure the Model Repository Service and Data Integration Service . 1 - Yes 2 - No If you select Yes, you can configure the application services. If you select No, you can configure the application services from the Administrator tool.

6. Select if you want to create a **monitoring Model Repository Service to monitor domain statistics**.

The following table describes the option to configure the monitoring Model repository:

Prompt	Description
Create a monitoring Model Repository Service	Select whether you want to create a monitoring Model Repository Service. 1 - Yes 2 - No If you select Yes, you can create a monitoring Model Repository Service. If you select No, you can create a monitoring Model Repository Service from the Administrator tool.

If you choose to configure the Model Repository Service and Data Integration Service, the **Model Repository Database** section appears. If you choose not to configure the Model Repository Service and Data Integration Service, the **Post-Installation Summary** section indicates whether the installation completed successfully. The summary also shows the status of the installed components and their configuration.

Configure the Model Repository Database

After you configure the domain and the node, you can configure the Model repository database properties.

1. Enter the Model Repository Service name.

   Enter the name of the service. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters:

   ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [

   You cannot change the name of the service after you create it.

   If you selected process level SPN, specify the Model Repository Service keytab file. The keytab file for the Model Repository Service process. The keytab file must have the following name: .keytab

2. Select the database to configure Model repository.

   The following table lists the databases you can configure Model repository:

Prompt	Description
Database type	Type of database for the Model repository. Select from the following options: 1 - Oracle 2 - Microsoft SQL Server 3 - IBM DB2

3. Enter the properties for the database user account.

   The following table lists the properties for the database user account:

Property	Description
Database user ID	Name for the Model repository database user account.
User password	Password for the domain configuration database user account.

4. Select whether to create a secure Model repository database.

You can create a model repository service in a database secured with the SSL protocol. To create a model repository service in a secure database, press 1 and skip to step to enter the JDBC information.

To create a model repository service in an unsecured database, press 2.

5. If you chose not to create a secured Model repository, enter the parameters for the database.
   a. If you select IBM DB2, select whether to configure a tablespace and enter the tablespace name.

      The following table describes the properties that you must configure for the IBM DB2 database:

Property	Description
Configure tablespace	Select whether to specify a tablespace: 1 - No 2 - Yes In a single-partition database, if you select No, the installer creates the tables in the default tablespace. In a multi-partition database, you must select Yes.
Tablespace	Name of the tablespace in which to create the tables. Specify a tablespace that meets the pageSize requirement of 32768 bytes. In a single-partition database, if you select Yes to configure the tablespace, enter the name of the tablespace in which to create the tables. In a multi-partition database, specify the name of the tablespace that resides in the catalog partition of the database.

   b. If you select Microsoft SQL Server, enter the schema name for the database.

      The following table describes the properties that you must configure for the Microsoft SQL Server database:

Property	Description
Schema name	Name of the schema that will contain domain configuration tables. If this parameter is blank, the installer creates the tables in the default schema.

   c. To enter the JDBC connection information using the JDBC URL information, press **1**. To enter the JDBC connection information using a custom JDBC connection string, press **2**.

   d. Enter the JDBC connection information.

      - To enter the connection information using the JDBC URL information, specify the JDBC URL properties.
        The following table describes the database connection information:

Prompt	Description
Database host name	Host name for the database.
Database port number	Port number for the database.

Prompt	Description
Database service name	Service or database name : - Oracle: Enter the service name. - Microsoft SQL Server: Enter the database name. - IBM DB2: Enter the service name.
Configure JDBC Parameters	Select whether to add additional JDBC parameters to the connection string: 1 - Yes 2 - No If you select Yes, enter the parameters or press Enter to accept the default. If you select No, the installer creates the JDBC connection string without parameters.

- To enter the connection information using a custom JDBC connection string, type the connection string.
  Use the following syntax in the JDBC connection string:
  IBM DB2

  ```
  jdbc:Informatica:db2://host_name:port_no;DatabaseName=
  ```

  Oracle

  ```
  jdbc:Informatica:oracle://host_name:port_no;ServiceName=
  ```

  Microsoft SQL Server

  ```
  jdbc:Informatica:sqlserver://
  host_name:port_no;SelectMethod=cursor;DatabaseName=
  ```

  Microsoft Azure SQL Database

  ```
  jdbc:Informatica:sqlserver://
  host_name:port_no;SelectMethod=cursor;DatabaseName=database_name;SnapshotSeria
  lizable=true;EncryptionMethod=SSL;HostNameInCertificate=*.database.windows.net
  ;ValidateServerCertificate=false
  ```

  Sybase

  ```
  jdbc:Informatica:sybase://host_name:port_no;DatabaseName=
  ```

  Verify that the connection string contains all the connection parameters required by your database system.

The **Service Parameters** section appears.

Configure the Monitoring Model Repository Database

After you configure Model Repository database, you can configure the monitoring Model repository database properties.

1. Enter the monitoring Model Repository Service name.

   Enter the name of the service. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters:

   ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [

   You cannot change the name of the service after you create it.

   If you selected process level SPN, specify the monitoring Model Repository Service keytab file. The keytab file for the monitoring Model Repository Service process. The keytab file must have the following name: .keytab

2. Select the database type for the monitoring Model repository.

   The following table lists the databases for the monitoring Model repository.

Prompt	Description
Database type	Type of database for the monitoring Model repository. Select from the following options: 1 - Oracle 2 - Microsoft SQL Server 3 - IBM DB2

3. Enter the properties for the database user account.

   The following table lists the properties for the database user account:

Property	Description
Database user ID	Name for the monitoring Model repository database user account.
User password	Password for the monitoring Model repository user account.

4. Select whether to create a secure monitoring Model repository database.

   You can create a monitoring Model repository in a database secured with the SSL protocol. To create a monitoring Model repository in a secure database, press 1 and skip to step to enter the JDBC information.

   To create a monitoring Model repository in an unsecured database, press 2.

5. If you do not create a secure monitoring Model repository, enter the parameters for the database.

   a. If you select IBM DB2, select whether to configure a tablespace and enter the tablespace name.

The following table describes the properties that you must configure for the IBM DB2 database:

Property	Description
Configure tablespace	Select whether to specify a tablespace: 1 - No 2 - Yes In a single-partition database, if you select No, the installer creates the tables in the default tablespace. In a multi-partition database, you must select Yes.
Tablespace	Name of the tablespace in which to create the tables. Specify a tablespace that meets the pageSize requirement of 32768 bytes. In a single-partition database, if you select Yes to configure the tablespace, enter the name of the tablespace in which to create the tables. In a multi-partition database, specify the name of the tablespace that resides in the catalog partition of the database.

b. If you select Microsoft SQL Server, enter the schema name for the database.

The following table describes the properties that you must configure for the Microsoft SQL Server database:

Property	Description
Schema name	Name of the schema that will contain domain configuration tables. If this parameter is blank, the installer creates the tables in the default schema.

c. To enter the JDBC connection information using the JDBC URL information, press **1**. To enter the JDBC connection information using a custom JDBC connection string, press **2**.

d. Enter the JDBC connection information.

- To enter the connection information using the JDBC URL information, specify the JDBC URL properties.
  The following table describes the database connection information:

Prompt	Description
Database host name	Host name for the database.
Database port number	Port number for the database.

Prompt	Description
Database service name	Service or database name : - Oracle: Enter the service name. - Microsoft SQL Server: Enter the database name. - IBM DB2: Enter the service name.
Configure JDBC Parameters	Select whether to add additional JDBC parameters to the connection string: 1 - Yes 2 - No If you select Yes, enter the parameters or press Enter to accept the default. If you select No, the installer creates the JDBC connection string without parameters.

- To enter the connection information using a custom JDBC connection string, type the connection string.
  Use the following syntax in the JDBC connection string:
  IBM DB2

  ```
  jdbc:Informatica:db2://host_name:port_no;DatabaseName=
  ```

  Oracle

  ```
  jdbc:Informatica:oracle://host_name:port_no;ServiceName=
  ```

  Microsoft SQL Server

  ```
  jdbc:Informatica:sqlserver://
  host_name:port_no;SelectMethod=cursor;DatabaseName=
  ```

  Microsoft Azure SQL Database

  ```
  jdbc:Informatica:sqlserver://
  host_name:port_no;SelectMethod=cursor;DatabaseName=database_name;SnapshotSeria
  lizable=true;EncryptionMethod=SSL;HostNameInCertificate=*.database.windows.net
  ;ValidateServerCertificate=false
  ```

  Sybase

  ```
  jdbc:Informatica:sybase://host_name:port_no;DatabaseName=
  ```

  Verify that the connection string contains all the connection parameters required by your database system.

The **Service Parameters** section appears.

Configure the Service Parameters

After you configure the Model Repository database, you can configure the service parameters for the application services.

1. Enter the following service parameter information:

Port	Description
Data Integration Service name	Name of the Data Integration Service to create in the Informatica domain.
HTTP protocol type	Type of connection to the Data Integration Service. Select one of the following options: - HTTP. Requests to the service uses an HTTP connection. - HTTPS. Requests to the service uses a secure HTTP connection. - HTTP&HTTPS. Requests to the service can use either an HTTP or HTTPS connection.
HTTP port	Port number to use for the Data Integration Service. Default is 9085.
HTTPS port	Port number to use for the Data Integration Service. Default is 9085.

2. Select the SSL certificates contained to secure the Data Integration Service.

Option	Description
Use the default Informatica SSL certificate files	Use the default Informatica SSL certificates contained in the default keystore and truststore. **Note:** If you do not provide an SSL certificate, Informatica uses the same default private key for all Informatica installations. If you use the default Informatica keystore and truststore files, the security of your domain could be compromised. To ensure a high level of security for the domain, select the option to specify the location of the SSL certificate files.
Use custom SSL certificates	Use custom SSL certificates. You must specify the location of the keystore and truststore files. You can provide a self-signed certificate or a certificate issued by a certificate authority (CA). You must provide SSL certificates in PEM format and in Java Keystore (JKS) files. Informatica requires specific names for the SSL certificate files for the Informatica domain. You must use the same SSL certificates for all nodes in the domain. Store the truststore and keystore files in a directory accessible to all the nodes in the domain and specify the same keystore file directory and truststore file directory for all nodes in the same domain.

If you choose to use custom SSL certificates, enter the following information.

Property	Description
Keystore file directory	Directory that contains the keystore files. The directory must contain files named infa_keystore.jks and infa_keystore.pem.
Keystore password	Password for the keystore infa_keystore.jks.

Property	Description
Truststore file directory	Directory that contains the truststore files. The directory must contain files named infa_truststore.jks and infa_truststore.pem.
Truststore password	Password for the infa_truststore.jks file.

3. Do you want to enable big data job recovery for the Data Integration Service?
   - Yes
   - No

   If you choose Yes, you can recover mapping jobs that the Data Integration Service pushes to the Spark engine for processing. Default is No.
4. Do you want to create a cluster configuration?

   The cluster configuration enables the Data Integration Service to push mapping logic to the cluster. If you are integrating with the Hadoop environment, you can create a cluster configuration.

   Press 1 if you want to create a cluster configuration.

   Press 2 if you do not want to create a cluster configuration. Default is 1.

   Note: To create a cluster configuration for the Databricks environment, use the Administrator tool after you complete installation.

   After installation, refer to the *Big Data Management Integration Guide* to fully integrate the domain with the Hadoop environment.

The **Post-Installation Summary** section indicates whether the installation completed successfully. The summary also shows the status of the installed components and their configuration.

Join a domain

You can join a domain if you are installing on multiple machines and you have created a domain on another machine.

Run the Installer

Perform the following steps to run the installer:

1. Log in to the machine with a system user account.
2. Close all other applications.
3. On a shell command line, run the `install.sh` file from the root directory.

   The installer displays the message to verify that the locale environment variables are set.
4. If the environment variables are not set, press **n** to exit the installer and set them as required.

   If the environment variables are set, press **y** to continue.
5. Press **1** to install Informatica Big Data Suite Products.

   The installer displays different options based on the platform you are installing on.

The following options appear:

a. Press **1** to run the Pre-Installation System Check Tool.

   For more information about running the Pre-Installation (i10Pi) System Check Tool, see "Run the Pre-Installation (i10Pi) System Check Tool" on page 35.

b. Press **2** to run the Informatica Kerberos SPN Format Generator.

   For more information about running the Informatica Kerberos SPN Format Generator, see "Running the SPN Format Generator on Linux" on page 69.

c. Press **3** to run the installer.

The **Welcome** section appears.

Accept Terms and Conditions

1. Read the terms and conditions for Informatica installation and the product usage toolkit and select **I agree to the terms and conditions**.

   Informatica DiscoveryIQ is a product usage tool that sends routine reports on data usage and system statistics to Informatica. Informatica DiscoveryIQ uploads data to Informatica 15 minutes after you install and configure Informatica domain. Thereafter, the domain sends the data every 30 days. You can choose to disable usage statistics from the Administrator tool.

   a. Press **1** to not accept the terms and conditions

   b. Press **2** to accept the terms and conditions.

2. Version 10.2.2 is for big data products only, such as Big Data Management and Big Data Quality. This version does not support non-big data products, such as PowerCenter or Informatica Data Quality.

   a. Press **1** and type **quit** to quit the installation.

   b. Press **2** to continue the installation.

   If you choose to not accept the terms and condition, the installer prompts you to accept the terms and conditions.

The **Component Selection** sections appears.

Product Installation

After you accept terms and conditions, you can install Informatica domain services.

1. Press **1** to install Informatica domain services.

   This option installs version 10.2.2 domain services and the application service binaries to support Big Data Management and Big Data Streaming.

2. Choose whether you want to run the installer on a network that uses Kerberos authentication.

   a. Press **1** to configure the Informatica domain to run on a network that does not use Kerberos authentication.

   b. Press **2** to configure the Informatica domain to run on a network with Kerberos authentication.

The **Installation Prerequisites** section displays the installation requirements. Verify that all requirements are met before you continue the installation.

Tune the Application Services

After you review the installation prerequisites, you can tune the application services for better performance based on the deployment type in your environment. If you do not tune now, you can tune the services later through infacmd.

1. Select if you want to tune the services now.

Prompt	Description
Do you want to tune the services now?	Select if you want to tune the services. 1 - No 2 - Yes Select no if you do not want to tune the services. Select yes if you want to tune the services.

If you are joining the node to existing domain, ensure the deployment type you select here is same deployment type as the gateway nodes.

2. Select the deployment type associated with the Informatica environment.

Prompt	Description
1. Sandbox	Choose this option if the environment is used for proof of concepts or as a sandbox environment with minimal users. Sandbox environments are typically configured with 16 cores, 32 GB RAM, and about 50 GB disk space.
2. Basic	Choose this option if the environment is used for low volume processing environments with low levels of concurrency. Basic environments are typically single- or multi-node setups configured with 24 cores, 64 GB RAM, and about 100 GB disk space.
3. Standard	Choose this option if the environment is used for high volume processing but with low levels of concurrency. Standard environments are typically multi-node setups configured with 64 GB RAM, more than 100 GB disk space per node, and 48 cores across nodes.
4. Advanced	Advanced Choose this option if the environment is used for high volume processing with high levels of concurrency. Advanced environments are typically multi-node setups configured with 128 GB RAM, more than 100 GB disk space per node, and 96 cores across nodes.

The **License and Installation Directory** section appears.

Specify Installation Directory

After you verify the installation prerequisites, you can specify the installation directory.

1. Enter the path and file name of the Informatica license key and press **Enter**.
2. Enter the absolute path for the installation directory.

   The directory names in the path must not contain spaces or the following special characters: @|* $ # ! % ( ) { } [ ] , ; ' Default is /home/toolinst.

   Note: Informatica recommends using alphanumeric characters in the installation directory path. If you use a special character such as á or €, unexpected results might occur at run time.

If you enabled Kerberos network authentication, the **Service Principal Level** section appears.

If you did not enable Kerberos network authentication, the **Pre-Installation Summary** section appears. Review the installation information and press **Enter** to continue. Skip to "Configure the Domain Options" on page 186.

Configure Security Level

After you specify the installation directory, you can configure security level.

- In the **Service Principal Level** section, select the level at which to set the Kerberos service principals for the domain.

  Note: All nodes in the domain must use the same service principal level. When you join a node to a domain, select the same service principal level used by the gateway node in the domain.

  The following table describes the levels that you can select:

Level	Description
Process Level	Configures the domain to use a unique service principal name (SPN) and keytab file for each node and each application service on a node. The number of SPNs and keytab files required for each node depends on the number of application service processes that run on the node. Use the process level option for domains that require a high level of security, such as productions domains.
Node Level	Configures the domain to share SPNs and keytab files on a node. This option requires one SPN and keytab file for the node and all application services that run on the node. It also requires a separate SPN and keytab file for all HTTP processes on the node. Use the node level option for domains that do not require a high level of security, such as test and development domains.

The **Network Security - Kerberos Authentication** section appears.

Configure the Domain Options

After you review the Pre-Installation summary, you can enter the domain information.

1. Press **2** to join a domain.

   The installer joins a node on the machine where you install.
2. Specify whether the domain you want to join has the secure communication option enabled.

   Press 1 to join an unsecure domain or press 2 to join a secure domain.
3. Select the type of node you want to create.

The following table describes that types of nodes that you can create:

Property	Description
Configure this node as a gateway	Select whether to configure the node as a gateway or worker node. 1 - Yes 2 - No Select 1 to configure a gateway node or 2 to configure a worker node.

If you configure the node as a gateway, you can enable a secure HTTPS connection to the Informatica Administrator.

4. Specify the connection details to Informatica Administrator.
   a. Specify whether to set up a secure HTTPS connection to the Informatica Administrator.

Option	Description
1 - Enable HTTPS for Informatica Administrator	Set up a secure connection to Informatica Administrator.
2 - Disable HTTPS	Do not set up a secure connection to Informatica Administrator.

   b. If you enable HTTPS connection for the Informatica Administrator, enter the keystore file and port number to use to secure the connection.

Option	Description
Port	Port number for the HTTPS connection.
Keystore file	Select whether to use a keystore file generated by the installer or a keystore file you create. You can use a keystore file with a self-signed certificate or a certificate signed by a certification authority. 1 - Use a keystore generated by the installer 2 - Specify a keystore file and password If you select to use a keystore file generated by the installer, the installer creates a self-signed keystore file named Default.keystore in the following location: `<Informatica installation directory>/tomcat/conf/`

   c. If you specify the keystore, enter the password and location of the keystore file.
   d. If you enabled secure communication for the domain, the **Domain Security - Secure Communication** section appears.
   e. If you did not enable secure communication for the domain, the **Domain Configuration Repository** section appears. Skip to "Configure Domain Repository Connection Details" on page 189.

5. Select whether to enable SAML authentication to configure Security Assertion Markup Language (SAML)-based single sign-on (SSO) support for web-based Informatica applications in an Informatica domain.

The following table describes the information you must enter to enable SAML authentication:

Prompt	Description
Does the domain use SAML authentication?	Select if the domain uses SAML authentication: 1 - No If you select No, skip to "Configure Domain Security" on page 188 2 - Yes If you select Yes, configure the SAML authentication.

6. Enter the Identity Provider URL for the domain.
7. Enter the identity provider assertion signing certificate alias name.
8. Select whether to use the default Informatica SSL certificates or to use your SSL certificates to enable SAML authentication in the domain.

   The following table describes the SSL certificate options for SAML authentication:

Option	Description
Use the default Informatica SSL certificate file.	Select to use the default Informatica truststore file for SAML authentication.
Enter the location of the SSL certificate file.	Select to use a custom truststore file for SAML authentication. Specify the directory containing the custom truststore file on gateway nodes within the domain. Specify the directory only, not the full path to the file.

9. If you provide the security certificates, specify the location and passwords of the truststore files.

   The following table describes the location and password of the truststore file:

Property	Description
Truststore Directory	Specify the directory containing the custom truststore file on gateway nodes within the domain. Specify the directory only, not the full path to the file.
Truststore Password	The password for the custom truststore file.

The **Domain Security - Secure Communication** appears.

Configure Domain Security

After you configure the domain, you can configure domain security.

- In the Domain Security - Secure Communication section, specify whether to use the default Informatica SSL certificates or to use your SSL certificates to secure domain communication.
  a. Select the type of SSL certificates to use.

The following table describes the options for the SSL certificates that you can use to secure the Informatica domain:

Option	Description
Use the default Informatica SSL certificates	Use the default SSL certificates contained in the default keystore and truststore. **Note:** If you do not provide an SSL certificate, Informatica uses the same default private key for all Informatica installations. If you use the default Informatica keystore and truststore files, the security of your domain could be compromised. To ensure a high level of security for the domain, select the option to specify the location of the SSL certificate files.
Use custom SSL certificates	Specify the path for the keystore and truststore files that contain the SSL certificates. You must also specify the keystore and truststore passwords. You can provide a self-signed certificate or a certificate issued by a certificate authority (CA). You must provide SSL certificates in PEM format and in Java Keystore (JKS) files. Informatica requires specific names for the SSL certificate files for the Informatica domain. You must use the same SSL certificates for all nodes in the domain. Store the truststore and keystore files in a directory accessible to all the nodes in the domain and specify the same keystore file directory and truststore file directory for all nodes in the same domain.

b. If you provide the SSL certificate, specify the location and passwords of the keystore and truststore files.

The following table describes the parameters that you must enter for the SSL certificate files:

Property	Description
Keystore file directory	Directory that contains the keystore files. The directory must contain files named infa_keystore.jks.
Keystore password	Password for the keystore infa_keystore.jks.
Truststore file directory	Directory that contains the truststore files. The directory must contain files named infa_truststore.jks and infa_truststore.pem.
Truststore password	Password for the infa_truststore.jks file.

The **Domain Configuration Repository** section appears.

Configure Domain Repository Connection Details

After you configure domain security, you can configure domain repository connection details.

- Enter the information for the domain that you want to join.

The following table describes the properties that you specify for the domain:

Property	Description
Domain name	Name of the domain to join.
Gateway node host	Host name of the machine that hosts the gateway node for the domain.
Gateway node port	Port number of the gateway node.
Domain user name	User name of the administrator for the domain you want to join.
Domain password	Password for the domain administrator.
Security domain name	Name of the secure domain.

The **Domain Security - Encryption Key** section appears.

Configure the Encryption Key

After you configure domain repository, you can configure encryption key.

- In the **Domain Security - Encryption Key** section, enter the directory for the encryption key for the Informatica domain.

  The following table describes the encryption key parameters that you must specify when you join a domain:

Property	Description
Select the encryption key	Path and file name of the encryption key for the Informatica domain that you want to join. All nodes in the Informatica domain use the same encryption key. You must specify the encryption key file created on the gateway node for the domain that you want to join. If you copied the encryption key file to a temporary directory to make it accessible to the nodes in the domain, specify the path and file name of the encryption key file in the temporary directory.
Encryption key directory	Directory in which to store the encryption key on the node created during this installation. The installer copies the encryption key file for the domain to the encryption key directory on the new node.

  The installer sets different permissions to the directory and the files in the directory. For more information about the permissions for the encryption key file and directory, see "Secure Files and Directories" on page 104.

The **Domain and Node Configuration** section appears.

Configure the Join Domain and Node

After you configure the encryption key, you can configure the join domain and node.

1. Enter the information for the domain and the node that you want to join.

The following table describes the properties that you set for the current node.

Property	Description
Node host name	Host name or IP address of the machine on which to join the node. If the machine has a single network name, use the default host name. If the a machine has multiple network names, you can modify the default host name to use an alternate network name. **Note:** The node host name cannot contain the underscore (_) character. Do not use localhost. The host name must explicitly identify the machine.
Node name	Name of the node to join.
Node port number	Port number for the node. The default port number for the node is 6005. If the port number is not available on the machine, the installer displays the next available port number.

2. Select whether to display the default ports for the domain and node components assigned by the installer.

   The following table describes the advanced port configuration page:

Prompt	Description
Display advanced port configuration page	Select whether to display the port numbers for the domain and node components assigned by the installer: 1 - No 2 - Yes If you select Yes, the installer displays the default port numbers assigned to the domain components. You can specify the port numbers to use for the domain and node components. You can also specify a range of port numbers to use for the service process that will run on the node. You can use the default port numbers or specify new port numbers. Verify that the port numbers you enter are not used by other applications.

3. If you display the port configuration page, enter new port numbers at the prompt or press Enter to use the default port numbers.

   The following table describes the ports that you can set:

Port	Description
Service Manager port	Port number used by the Service Manager on the node. The Service Manager listens for incoming connection requests on this port. Client applications use this port to communicate with the services in the domain. The Informatica command line programs use this port to communicate to the domain. This is also the port for the SQL data service JDBC/ODBC driver. Default is 6006.
Service Manager Shutdown port	Port number that controls server shutdown for the domain Service Manager. The Service Manager listens for shutdown commands on this port. Default is 6007.
Informatica Administrator port	Port number used by Informatica Administrator. Default is 6008.

Port	Description
Informatica Administrator HTTPS port	No default port. Enter the required port number when you create the service. Setting this port to 0 disables an HTTPS connection to the Administrator tool.
Informatica Administrator shutdown port	Port number that controls server shutdown for Informatica Administrator. Informatica Administrator listens for shutdown commands on this port. Default is 6009.
Minimum port number	Lowest port number in the range of dynamic port numbers that can be assigned to the application service processes that run on this node. Default is 6014.
Maximum port number	Highest port number in the range of dynamic port numbers that can be assigned to the application service processes that run on this node. Default is 6114.

The **Post-Installation Summary** section indicates whether the installation completed successfully. The summary also shows the status of the installed components and their configuration.

Resume the Installer

When the installation process stops midway, you can resume the installation from the point of failure or exit.

When you run the server installer and the installation process fails, you can still resume from the previous service configuration and recover the last entered details for that service installation.

The install process might fail for reasons such as network outage, when you exit the installation before completing the entire installation process, or because of incorrect information entered.

If a service fails or if the installation process fails during a service creation, you can resume the installation process with the server installer. To resume the installation process, ensure that at least one of the services is created and that the domain is up and running from the installation log. For example, if you want to check whether the Model Repository Service is created, check if you have a service creation success text in the server log in the following format:

```
SUCCESS: MRS Service [mrs_name] is created. Command ran successfully.
```

To resume the installation, run the installer again.

Note: You cannot resume the installer if you are running it to configure services after the services have been created. When you run the service configuration wizard, you cannot resume the installer for Big Data, Enterprise Data Lake, or Enterprise Data Catalog. When you join the domain, you also cannot resume the installer.

When you resume the installer while creating a service, the installer retains all the service and database specific information, such as the create service status, service name, service enabled or disabled status. You can confirm and use the previously entered values or specify new values for the service and resume the installation process.

Resuming the Installer

When the installation process stops midway, you can resume the installation from the point of failure or exit.

Before you can resume the installer, complete the following prerequisites:

- In the installation log file present in the installation directory, verify that at least the domain and one service is created. The installer log file name appears in the following syntax: Informatica_<Version>_Services_<timestamp>.log
- Ensure that you do not delete the installInst.obj object file present in the tools folder of the user installation directory.
- For silent installer, ensure that RESUME_INSTALLATION is set to true in the SilentInput.properties file.

1. Open a command prompt and navigate to the location of the installation files.
2. Run the Installer.

   On Linux, run silentInstall.sh to resume the silent installer. To resume the regular installer, run the ./install.sh command.

3. When the regular installer runs, you might get a prompt confirming whether you want to resume previous installer or not.
   - If you do not want to resume installation, enter 1 for No. Default is 1.
   - If you want to resume installation, enter 2 for Yes.

   Before you can resume the installation, the services get validated.

CHAPTER 11

Install Enterprise Data Catalog and Enterprise Data Lake

This chapter includes the following topics:

Overview

You can install the Enterprise Data Catalog and Enterprise Data Lake binaries on a node where the Informatica services are installed. Use this installation option to install the products in a domain where Big Data products run.

After you install the binaries, run the installer to configure the Enterprise Data Catalog and Enterprise Data Lake services. Configure the Enterprise Data Catalog services, and then configure the Enterprise Data Lake services.

Before you run the installer to configure the Enterprise Data Lake services, verify that the domain is integrated with the Hadoop environment and that the Hadoop, HDFS, and Hive connections are associated with the cluster configuration. If the cluster configuration does not exist, use the Administrator tool to create the connections manually after you integrate the domain with the Hadoop environment.

For more information about integrating the domain with the Hadoop environment, see the *Informatica Big Data Management Integration Guide*.

Installation Process

The installation of Enterprise Data Catalog and Enterprise Data Lake consists of multiple phases.

When you plan to install Enterprise Data Catalog and Enterprise Data Lake, you must account for the dependencies for each product. You also must plan the relational databases that each product requires.

Consider the following high-level tasks of the installation process:

Perform pre-installation tasks for the Catalog Service.

Review the Catalog Service requirements. For more information, see "Catalog Service" on page 45.

Perform pre-installation tasks for the Data Preparation Service.

Review the Data Preparation Service requirements. For more information, see "Data Preparation Service" on page 53.

You must prepare the Data Preparation repository database before you create the Data Preparation Service. For more information, see "Data Preparation Repository Database Requirements" on page 53.

If you plan to use rules during data preparation, you can create a Model Repository Service to associate with the Data Preparation Service when you configure the Enterprise Data Lake services. Prepare the Model repository database that contains rule metadata before you create the Data Preparation Service.

Perform pre-installation tasks for the Enterprise Data Lake Service.

Review the Enterprise Data Lake Service requirements. For more information, see "Enterprise Data Lake Service" on page 55.

You can create a Model Repository Service to associate with the Enterprise Data Lake Service when you configure the Enterprise Data Lake services. Prepare the Model repository database before you create the Enterprise Data Lake Service.

Run the installer to install the Enterprise Data Catalog and Enterprise Data Lake binaries.

Run the installer to install the Enterprise Data Catalog and Enterprise Data Lake binaries on the node. For more information, see "Install the Enterprise Data Catalog and Enterprise Data Lake Binaries" on page 195.

Run the installer to configure the Enterprise Data Catalog services.

Run the installer to create and enable the Enterprise Data Catalog services on the node. For more information, see "Configure Enterprise Data Catalog" on page 196.

Run the installer to configure the Enterprise Data Lake services.

Run the installer to create and enable the Enterprise Data Lake services on the node. For more information, see "Configure Enterprise Data Lake" on page 207.

Install the Enterprise Data Catalog and Enterprise Data Lake Binaries

You can install the Enterprise Data Catalog and Enterprise Data Lake binaries on a node on which the Informatica services are installed.

1. On a shell command line, run the `install.sh` file from the root directory.

   The installer displays the message for documentation and copyright information.
2. Press **1** to install the Informatica Big Data products.
3. Press **2** to agree to the terms and conditions.
4. Press **2** to continue with the installation.
5. Press **3** to install Enterprise Data Lake.
6. Press **2** to indicate that the Informatica services are installed on the node.
7. Press **1** to indicate that Enterprise Data Catalog is not installed on the node.

8. Press **2** to tune the application services for better performance based on your deployment type.
9. Enter the directory where you want to install the Enterprise Data Catalog and Enterprise Data Lake binaries.

   To enable Informatica to register Enterprise Data Lake with the domain, you must install the Enterprise Data Catalog and Enterprise Data Lake binaries in the same directory on any gateway node.
10. Choose how to proceed if Enterprise Data Lake is already installed in the specified directory.
    - Press **1** to change the installation directory.
    - Press **2** to overwrite the existing installation.
11. Review the pre-installation summary, and then press **Enter**.
12. Ensure the current node is shut down, and then press **Enter**.
13. After you install the Enterprise Data Catalog and Enterprise Data Lake binaries, run the installer to configure the Enterprise Data Catalog and Enterprise Data Lake services. Configure the Enterprise Data Catalog services, and then configure the Enterprise Data Lake services.

    For more information about configuring the Enterprise Data Catalog services, see "Configure Enterprise Data Catalog" on page 196

    For more information about configuring the Enterprise Data Lake services, see "Configure Enterprise Data Lake" on page 207.

Configure Enterprise Data Catalog

Enterprise Data Catalog requires application services to be created and in running state before you can use it.

You can create the application services using one of the following methods:

Using the installer when you install Enterprise Data Catalog

For more information about using the installer to create the application services when you install Enterprise Data Catalog, see https://network.informatica.com/onlinehelp/edc/Install_Help/index.htm

Using the Informatica Administrator after you install Enterprise Data Catalog

For more information about creating the application services using Informatica Administrator, see the *Create the Application Services* chapter in this guide.

Using the installer after you install Enterprise Data Catalog

For more information about creating the application services using the installer after you install Enterprise Data Catalog, see the steps listed in the *Creating the Enterprise Data Catalog Application Services Using the Installer* topic.

If the application services are not created or if the process failed, you cannot resume the process from the point of failure using the installer. You can restart the process using the installer.

Creating the Enterprise Data Catalog Application Services Using the Installer

Perform the following steps to create the application services using the installer after you install Enterprise Data Catalog:

1. Log in to the machine with a system user account.
2. Close all applications running on the machine.

3. On a shell command line, run the `./install.sh` command to start the installer.
4. Press **y** to proceed with the installation.
5. Press **3** to select the option to install the application services for Enterprise Data Catalog or Enterprise Data Lake.
6. Press **2** to agree to the terms and conditions.
7. Press **2** to accept that you want to proceed with the installation of big data products.
8. Press **1** to configure services for Enterprise Data Catalog.
9. Press **1** to confirm that the latest version of Enterprise Data Catalog services is not installed.
10. Enter the directory where you installed Enterprise Data Catalog and press **Enter**.
11. Enter the following domain details that you had configured when you installed Enterprise Data Catalog:

Proceed to provide the Informatica domain details that you had configured.

Specifying the Informatica Domain Details

Perform the following steps to specify the Informatica domain details that you had configured:

1. Enter the following domain details that you had configured when you installed Enterprise Data Catalog:
   a. Domain name. Provide the name of the Informatica domain that you created and press **Enter**.
   b. Node name. Provide the name of the node that you created on the machine where you installed Enterprise Data Catalog and press **Enter**.
   c. Domain user password. Provide the password you configured for the Informatica domain administrator and press **Enter**.

Selecting the Application Services

Perform the following steps to select the application services that you want to create:

1. Press **1** to confirm that you want to create the Model Repository Service and Data Integration Service.
2. Press **1** if you want to create the Monitoring Model Repository Service to monitor the Informatica domain statistics.
3. Press **2** to specify that you do not want to create a cluster configuration. You must create a cluster configuration in the Informatica domain if you plan to configure Enterprise Data Lake services.
4. Press **1** if you want to create the profiling warehouse connection.
5. Press **1** to configure the Content Management Service.
6. You need to create the Informatica Cluster Service if you deploy Enterprise Data Catalog on an embedded cluster. Press **1** if you want to configure the Informatica Cluster Service.
   See the following points to decide how you want to create the Informatica Cluster Service:
   - Create the Informatica Cluster Service. The installer creates the Informatica Cluster Service.
   - Select the option to specify that you do not want to create the Informatica Cluster Service. The installer prompts you to specify if you want to associate an Informatica Cluster Service with the Catalog Service. If you select this option, the installer does not create a new Informatica Cluster Service. The installer prompts you for an Informatica Cluster Service that you want to associate with the Catalog Service.
   - Select the options to specify that you do not want to create the Informatica Cluster Service and associate an existing Informatica Cluster Service with the Catalog Service. The installer does not create the Informatica Cluster Service and proceeds to create the Catalog Service.
7. Press **1** if you want to configure the Catalog Service.

Creating the Model Repository Service

Perform the following steps to create the Model Repository Service:

1. Name of the Model Repository Service.
2. Name of the node on which the Model Repository Service must run.
3. The license that you want to associate with the Model Repository Service.
4. Select the database that you want to configure for the model repository from the following options:
   - Oracle
   - SQL Server
   - DB2

   Default is Oracle.
5. Type the username to access the database in the **Database user ID** parameter and press **Enter**. Default is **admin**.
6. Type the password for the username in the **User password** parameter and press Enter.
7. Press **1** if the database is secured with SSL.
   If you selected the option to specify that the database is SSL-enabled, provide the following parameters:

Secure Database Parameter	Description
EncryptionMethod	Indicates whether data is encrypted when transmitted over the network. This parameter must be set to SSL.
ValidateServerCertificate	Indicates whether Informatica validates the certificate that the database server sends. If this parameter is set to True, Informatica validates the certificate that the database server sends. If you specify the HostNameInCertificate parameter, Informatica also validates the host name in the certificate. If this parameter is set to False, Informatica does not validate the certificate that the database server sends. Informatica ignores any truststore information that you specify.
HostNameInCertificate	Host name of the machine that hosts the secure database. If you specify a host name, Informatica validates the host name included in the connection string against the host name in the SSL certificate.
cryptoProtocolVersion	Specifies the cryptographic protocol to use to connect to a secure database. You can set the parameter to cryptoProtocolVersion=TLSv1.1 or cryptoProtocolVersion=TLSv1.2 based on the cryptographic protocol used by the database server.
TrustStore	Path and file name of the truststore file that contains the SSL certificate for the database. If you do not include the path for the truststore file, Informatica looks for the file in the following default directory: `<Informatica installation directory>/tomcat/bin`
TrustStorePassword	Password for the truststore file for the secure database.

8. Press **1** to specify the JDBC URL to connect to the database.
9. Specify the database address in the following format for the **Database address** parameter: `<Fully qualified domain name of the host>:<port>`

10. Specify the database service name in the following format for the **Database service name** parameter:
    `<Fully qualified domain name of the service>`
11. Press **1** to specify that you want to configure the JDBC parameters.
12. Specify the required values for the parameters or press **Enter** to apply the default values. Press **Enter** to retain the default values.
    Use the following syntax for the connection string of the selected database type:

Database Type	Connection String Syntax
IBM DB2	`jdbc:informatica:db2://<host_name>:<port_number>;DatabaseName=<database_name>;BatchPerformanceWorkaround=true;DynamicSections=3000`

Microsoft SQL Server	- Microsoft SQL Server that uses the default instance
	`jdbc:informatica:sqlserver://<host_name>:<port_number>;DatabaseName=<database_name>;SnapshotSerializable=true`
	- Microsoft SQL Server that uses a named instance
	`jdbc:informatica:sqlserver://<host_name>\<named_instance_name>;DatabaseName=<database_name>;SnapshotSerializable=true`

Oracle	`jdbc:informatica:oracle://<host_name>:<port_number>;SID=<database_name>;MaxPooledStatements=20;CatalogOptions=0;BatchPerformanceWorkaround=true`

    The installer validates the node name and license, and then creates the Model Repository Service. The installer proceeds to create the Data Integration Service.

Creating the Data Integration Service

Perform the following steps to create the Data Integration Service:

1. Name of the Data Integration Service.
2. Name of the node on which the Data Integration Service must run.
3. The license that you want to associate with the Data Integration Service.
4. The name of the Model Repository Service that you want to associate with the Data Integration Service.
5. Specify the protocol that you want to use for the service from the following options:
   - http
   - https
   - http&https
     If you select **https** or **http&https** as the protocol for the service, provide the following details:
     1. HTTPS port. Default is 18095.
     2. Specify the SSL certificate that you want to use to secure the Data Integration Service. You can use the default SSL certificates in the default keystore and truststore or use custom SSL certificates. If you choose custom SSL certificates, specify the path that includes the filename of the keystore and truststore files and the passwords to access the keystore and truststore files.
6. Press **1** if you want the Data Integration Service to use Spark engine to run Sqoop mappings or process Java transformations.

The installer validates the node name and the license and creates and enables the Data Integration Service. The installer proceeds to create the profiling warehouse.

Configuring the Profiling Warehouse

Provide the following details to configure the database for the profiling warehouse:

1. Name of the Data Integration Service that you want to associate with the profiling warehouse.
2. Select the database that you want to configure for the profiling warehouse from the following options:
   - Oracle
   - SQL Server
   - DB2

   Default is Oracle.
3. Type the username to access the database in the **Database user ID** parameter and press **Enter**. Default is **admin**.
4. Type the password for the username in the **User password** parameter and press Enter.
5. Press **1** if the database is secured with SSL.
   If you selected the option to specify that the database is SSL-enabled, provide the following parameters:

Secure Database Parameter	Description
EncryptionMethod	Indicates whether data is encrypted when transmitted over the network. This parameter must be set to SSL.
ValidateServerCertificate	Indicates whether Informatica validates the certificate that the database server sends. If this parameter is set to True, Informatica validates the certificate that the database server sends. If you specify the HostNameInCertificate parameter, Informatica also validates the host name in the certificate. If this parameter is set to False, Informatica does not validate the certificate that the database server sends. Informatica ignores any truststore information that you specify.
HostNameInCertificate	Host name of the machine that hosts the secure database. If you specify a host name, Informatica validates the host name included in the connection string against the host name in the SSL certificate.
cryptoProtocolVersion	Specifies the cryptographic protocol to use to connect to a secure database. You can set the parameter to cryptoProtocolVersion=TLSv1.1 or cryptoProtocolVersion=TLSv1.2 based on the cryptographic protocol used by the database server.
TrustStore	Path and file name of the truststore file that contains the SSL certificate for the database. If you do not include the path for the truststore file, Informatica looks for the file in the following default directory: `<Informatica installation directory>/tomcat/bin`
TrustStorePassword	Password for the truststore file for the secure database.

6. Press **1** to specify the JDBC URL to connect to the database.
7. Specify the database address in the following format for the **Database address** parameter: `<Fully qualified domain name of the host>:<port>`
8. Specify the database service name in the following format for the **Database service name** parameter: `<Fully qualified domain name of the service>`

9. Press **1** to specify that you want to configure the JDBC parameters.
10. Specify the required values for the parameters or press **Enter** to apply the default values. Press **Enter** to retain the default values.
    Use the following syntax for the connection string of the selected database type:

Database Type	Connection String Syntax
IBM DB2	`jdbc:informatica:db2://<host_name>:<port_number>;DatabaseName=<database_name>;BatchPerformanceWorkaround=true;DynamicSections=3000`

Microsoft SQL Server	- Microsoft SQL Server that uses the default instance
	`jdbc:informatica:sqlserver://<host_name>:<port_number>;DatabaseName=<database_name>;SnapshotSerializable=true`
	- Microsoft SQL Server that uses a named instance
	`jdbc:informatica:sqlserver://<host_name>\<named_instance_name>;DatabaseName=<database_name>;SnapshotSerializable=true`

Oracle	`jdbc:informatica:oracle://<host_name>:<port_number>;SID=<database_name>;MaxPooledStatements=20;CatalogOptions=0;BatchPerformanceWorkaround=true`

The installer creates the data profiling warehouse and proceeds to create the Content Management Service.

Creating the Content Management Service

Provide the following details to create the Content Management Service:

1. Name of the Model Repository Service that you want to associate with the service.
2. Name of the Data Integration Service that you want to associate with the service.
3. Name of the node on which the Content Management Service must run.
4. The license that you want to associate with the Content Management Service.
5. Name of the Content Management Service.
6. Specify the protocol that you want to use for the service from the following options:
   - http
   - https

   If you select **https** as the protocol for the service, provide the following details:
   1. HTTPS port. Default is 17466.
   2. Specify the SSL certificate that you want to use to secure the Content Management Service. You can use the default SSL certificates in the default keystore or use custom SSL certificates. If you choose custom SSL certificates, specify the path that includes the filename of the keystore file and the password to access the keystore file.
7. Perform the following steps to configure the database for the Content Management Service:
8. Select the database that you want to configure for the Content Management Service from the following options:
   - Oracle

- SQL Server
- DB2

Default is Oracle.

9. Type the username to access the database in the **Database user ID** parameter and press **Enter**. Default is **admin**.
10. Type the password for the username in the **User password** parameter and press Enter.
11. Press **1** if the database is secured with SSL.
    If you selected the option to specify that the database is SSL-enabled, provide the following parameters:

Secure Database Parameter	Description
EncryptionMethod	Indicates whether data is encrypted when transmitted over the network. This parameter must be set to SSL.
ValidateServerCertificate	Indicates whether Informatica validates the certificate that the database server sends. If this parameter is set to True, Informatica validates the certificate that the database server sends. If you specify the HostNameInCertificate parameter, Informatica also validates the host name in the certificate. If this parameter is set to False, Informatica does not validate the certificate that the database server sends. Informatica ignores any truststore information that you specify.
HostNameInCertificate	Host name of the machine that hosts the secure database. If you specify a host name, Informatica validates the host name included in the connection string against the host name in the SSL certificate.
cryptoProtocolVersion	Specifies the cryptographic protocol to use to connect to a secure database. You can set the parameter to cryptoProtocolVersion=TLSv1.1 or cryptoProtocolVersion=TLSv1.2 based on the cryptographic protocol used by the database server.
TrustStore	Path and file name of the truststore file that contains the SSL certificate for the database. If you do not include the path for the truststore file, Informatica looks for the file in the following default directory: `<Informatica installation directory>/tomcat/bin`
TrustStorePassword	Password for the truststore file for the secure database.

12. Press **1** to specify the JDBC URL to connect to the database.
13. Specify the database address in the following format for the **Database address** parameter: `<Fully qualified domain name of the host>:<port>`
14. Specify the database service name in the following format for the **Database service name** parameter: `<Fully qualified domain name of the service>`
15. Press **1** to specify that you want to configure the JDBC parameters.
16. Specify the required values for the parameters or press **Enter** to apply the default values. Press **Enter** to retain the default values.

Use the following syntax for the connection string of the selected database type:

Database Type	Connection String Syntax
IBM DB2	`jdbc:informatica:db2://<host_name>:<port_number>;DatabaseName=<database_name>;BatchPerformanceWorkaround=true;DynamicSections=3000`

Microsoft SQL Server	- Microsoft SQL Server that uses the default instance
	`jdbc:informatica:sqlserver://<host_name>:<port_number>;DatabaseName=<database_name>;SnapshotSerializable=true`
	- Microsoft SQL Server that uses a named instance
	`jdbc:informatica:sqlserver://<host_name>\<named_instance_name>;DatabaseName=<database_name>;SnapshotSerializable=true`

Oracle	`jdbc:informatica:oracle://<host_name>:<port_number>;SID=<database_name>;MaxPooledStatements=20;CatalogOptions=0;BatchPerformanceWorkaround=true`

The installer creates and enables the Content Management Service and proceeds to configure the cluster and application service options.

Configuring the Cluster and Application Service Options

Perform the following steps to configure the cluster and application service options:

1. Press **1** if you want the installer to configure the Apache ZooKeeper, YARN, and HDFS based on the properties that you provide.
2. Select the cluster type from the following options:
   - Hortonworks
   - Cloudera
   - Azure HDInsight
3. Specify if the cluster uses Kerberos authentication.
4. Specify if the cluster is enabled for SSL.

Creating the Informatica Cluster Service

If you are installing Enterprise Data Catalog on an embedded cluster, provide the following details to configure the Informatica Cluster Service:

1. Name of the node on which the Informatica Cluster Service must run.
2. The license that you want to associate with the Informatica Cluster Service.
3. Username for the Apache Ambari Server. Default is root.
4. Name of the Informatica Cluster Service.
5. Hostname for the Informatica Hadoop cluster gateway.
6. Comma-separated list of hadoop nodes where the Apache Ambari agents run.
7. Port number of the Apache Ambari Server. Default is 9075.
8. Port number of the Informatica Hadoop cluster gateway. Default is 8080.
9. Specify if you want to change the default password for Ambari.

10. Provide the following properties if you want to enable SSL for the cluster:
    a. HTTPS port for the Informatica Cluster Service. Default is 7500.
    b. Press **1** if you want to use the default keystore generated by the installer.
    c. If you select to use a keystore file generated by the installer, the installer creates a self-signed keystore file named Default.keystore in the following location: <Informatica installation directory>/tomcat/conf/. Alternatively, you can use a keystore file with a self-signed certificate or a certificate signed by a certification authority. If you do not plan to use the keystore file generated by the installer, verify that you provide SSL certificates in PEM format and in Java Keystore (JKS) files. Informatica requires specific names for the SSL certificate files for the Informatica domain. You must use the same SSL certificates for all nodes in the domain. Store the truststore and keystore files in a directory accessible to all the nodes in the domain and specify the same keystore file directory and truststore file directory for all nodes in the same domain.
    d. Specify the truststore file location accessible to all the nodes in the domain and press **Enter**. Default is `/opt/ssl`.
11. Press **2** if you want to specify a directory to store HDFS, YARN, and ZooKeeper log files and data directories.
12. If you selected **2** for the previous step, specify the path to the directory where you want to store the log files and press **Enter**.
13. The installer creates Informatica Cluster Service and proceeds to create the Catalog Service.

Creating the Catalog Service

Provide the following details to configure the Catalog Service:

1. Name of the Catalog Service.
2. Name of the Model Repository Service that you wan to associate with the Catalog Service.
3. Name of the node on which the Catalog Service must run.
4. For Enterprise Data catalog deployed on an embedded cluster, if you had selected the option to associate an Informatica Cluster Service with the Catalog Service, provide the name of the Informatica Cluster Service.
5. The license that you want to associate with the Catalog Service.
6. The cluster Hadoop distribution URL.
7. Username to access the cluster Hadoop distribution URL. Default is admin.
8. Password to access the cluster Hadoop distribution URL.
9. Specify the following properties if you deployed Enterprise Data Catalog on an existing cluster:

Property	Description
Name of the cluster	If you selected Cloudera as the cluster type, you can provide a name for the cluster.
HDFS Service Name for High Availability	Applies to highly available existing cluster. Specify the HDFS service name.
Yarn resource manager scheduler URI	Scheduler URI value for the Yarn resource manager.

Note: If you select ClouderaManager or Hortonworks as the Hadoop distribution for an existing cluster, Enterprise Data Catalog automatically identifies the following properties for the Hadoop-distribution type:

- ZooKeeper Cluster URI
- HDFS Namenode URI
- Yarn resource manager URI
- Yarn resource manager HTTPS or HTTP URI
- History Server HTTP URI
- HDFS Service Name for High Availability
- Yarn resource manager scheduler URI

10. If you deployed Enterprise Data Catalog on an Azure HDInsight cluster, specify the following properties for the Catalog Service:

Property	Description
Cluster type	External Cluster
Hadoop distribution	HDInsight
Cluster URL	Fully qualified host name to access the cluster.
Cluster URL username	User name to access the cluster.
Cluster URL password	Password for the Cluster URL username.

After you create the Catalog Service, configure the following custom properties in Informatica Administrator for the Catalog Service:

Custom Property	Description
LdmCustomOptions.deployment.azure.account.key	The key to authenticate the Catalog Service to connect to Azure storage account . The value of the Azure storage account key might be encrypted or non encrypted. You can retrieve the value from `fs.azure.account.key.<storage account name>` property in `core-site.xml` file present in the Azure HDInsight cluster.
LdmCustomOptions.deployment.azure.key.decryption.script.path	If the key specified in the `LdmCustomOptions.deployment.azure.account.key` property is in encrypted format, you can use the decrypt shell script to decrypt the key using the key certificate. You must verify that you copy the decrypt shell script and key certificate file to the (same path as cluster machine) domain machine before enabling Catalog Service. Make sure that you maintain the path in the Azure HDInsight cluster machine for the copied files in the domain machine. The value for the property is the location of the decrypt shell script. For example, `/usr/lib/python2.7/dist-packages/hdinsight_common/decrypt.sh`. The key certificate file, key_decryption_cert.prv, is present in the `/usr/lib/hdinsight-common/certs/key_decryption_cert.prv` directory of Azure HDInsight cluster.
LdmCustomOptions.deployment.hdfs.default.fs	Address of the WASB storage account to which the Catalog Service must connect. The address includes the WASB storage container name with the storage account name. The value for the property is the complete WASB address with the container and storage account names. You can retrieve the value for the property from the `fs.defaultFS` property in the `core-site.xml` file present in the Azure HDInsight cluster.

11. The service cluster name.
12. Provide the following properties if you want to enable secure access to the Catalog Service:
    a. Provide the HTTPS port that you want to configure for the Catalog Service. Default is 9124.
    b. Press **1** if you want to use the default keystore generated by the installer, else press **2** to use a custom keystore file.
    If you do not plan to use the keystore file generated by the installer, verify that you provide SSL certificates in PEM format and in Java Keystore (JKS) files. Informatica requires specific names for the SSL certificate files for the Informatica domain. You must use the same SSL certificates for all nodes in the domain. Store the truststore and keystore files in a directory accessible to all the nodes in the domain and specify the same keystore file directory and truststore file directory for all nodes in the same domain.

    If you selected the option to use a custom keystore file, provide the following details:

    1. Path to the keystore file.
    2. The keystore alias.

3. Password of the keystore file.
4. Password of the Solr keystore.

13. Select the metadata load size that you want to ingest into the catalog from the following options:
    - demo
    - low
    - medium
    - high

Configure Enterprise Data Lake

After you install the Enterprise Data Lake binaries on a node, run the installer to create and enable the Enterprise Data Lake services on the node.

You can create the Data Preparation Service and Enterprise Data Lake Service during the configuration process. To create the services, the domain must be integrated with the Hadoop environment before you run the installer. For more information about integrating the domain with the Hadoop environment, see the *Informatica 10.2.2 Big Data Management Integration Guide*.

If you plan to use rules, you must associate the Data Preparation Service with the Model Repository Service that manages the Model repository that contains the rule objects and metadata. You must also associate a Data Integration Service with the Data Preparation Service that runs rules during data preparation. You can create a Model Repository Service and Data Integration Service to associate with the Data Preparation Service, or you can associate existing services with the Data Preparation Service.

You must associate a Model Repository Service and a Data Integration Service with the Enterprise Data Lake Service. You can create a Model Repository Service and Data Integration Service to associate with the Enterprise Data Lake Service, or you can associate existing services with the Enterprise Data Lake Service.

If you create a Model Repository Service, you must specify the details for the Model repository database used by the Model Repository Service.

Configure the Enterprise Data Lake Services

When you configure the Enterprise Data Lake services on a domain node on which the application binaries are already installed, you indicate that domain services and Enterprise Data Catalog are already installed on the node.

1. On a shell command line, run the `install.sh` file from the root directory.
2. Press **3** to configure the Enterprise Data Catalog or Enterprise Data Lake services.
3. Press **2** to agree to the terms and conditions.
4. Press **2** to continue with the installation.
5. Press **2** to configure the Enterprise Data Lake services.
6. Press **2** to indicate that the Enterprise Data Catalog services exist on the node.
7. Enter the directory containing the Enterprise Data Lake binaries.

The Domain Details section appears.

Configure the Domain Details

Provide the domain authentication details.

▶ Enter the domain administrator user name and password.

The Associated Services section appears.

Configure the Associated Services

Configure the application services required by Enterprise Data Lake.

1. Enter the name of the Catalog Service to associate with Enterprise Data Lake.
2. Enter the name of the Model Repository Service associated with the Catalog Service.
3. Enter the name of the Data Integration Service associated with the Catalog Service.
4. Choose whether to create a Model Repository Service and a Data Integration Service to associate with the Enterprise Data Lake Service, or to associate existing application services with Enterprise Data Lake.
   - To create to create a Model Repository Service and a Data Integration Service to associate with the Enterprise Data Lake Service on the node, press **1**.
   - To associate an existing Model Repository Service and an existing Data Integration Service with the Enterprise Data Lake Service, press **2**.

     After you specify the service names, skip to "Configure the Data Preparation Repository Database Details" on page 144.

The Application Services for Enterprise Data Lake section appears.

Configure the Model Repository Service and Model Repository Database Details

If you choose to create a Model Repository Service to associate with Enterprise Data Lake, specify the application service details and the connection details for the Model repository database.

1. Enter the name of the Model Repository Service to associate with Enterprise Data Lake.

   The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
2. Enter the name of the domain node where the service runs.
3. Enter the name of the Informatica license to associate with the service.
4. Specify the connection details for the Model repository database.

   The following table describes the parameters you set:

Property	Description
Database Type	Database for the Model repository managed by the Model Repository Service.
Database User ID	User name of the database user account to use to log in to the Model repository database.
User Password	Password for the Model repository database user account.

Property	Description
Tablespace	Configure for a IBM DB2 database. Name of the tablespace in which to create the tables. The tablespace must be defined on a single node, and the page size must be 32K. This option is required for a multi-partition database. If this option is not selected for a single partition database, the installer creates the tables in the default tablespace.
Schema Name	Configure for a Microsoft SQL Server database. Name of the schema that will contain domain configuration tables. If not selected, the installer creates the tables in the default schema.

5. Specify the truststore details required to access a secure Model repository database.

   The following table describes the properties you set:

Property	Description
Database Truststore File	Path and file name of the truststore file for the secure database.
Database Truststore Password	Password for the truststore.

6. Choose whether to configure the database connection using a JDBC URL or a custom JDBC connection string.
   - Press **1** to configure the database connection using a JDBC URL.

The following table describes the properties you set:

Property	Description
Database Address	Host name and port number for the database in the format `<host name>:<port>`.
Database Service Name	Service name for Oracle and IBM DB2 databases, or database name for Microsoft SQL Server.
JDBC Parameters	The JDBC connection string used to connect to the Model repository database. You can use the default parameters or add or modify the parameters based on your database requirements. Verify that the parameter string is valid. The installer does not validate the parameter string before it adds the string to the JDBC URL. If not selected, the installer creates the JDBC URL without additional parameters. Use the following JDBC connect string syntax for each supported database: - IBM DB2. `jdbc:informatica:db2://<host name>:<port>;DatabaseName=<database name>;BatchPerformanceWorkaround=true;DynamicSections=3000` - Microsoft SQL Server that uses the default instance. `jdbc:informatica:sqlserver://<host name>:<port>;DatabaseName=<database name>;SnapshotSerializable=true` - Microsoft SQL Server that uses a named instance. `jdbc:informatica:sqlserver://<host name>\<named instance name>;DatabaseName=<database name>;SnapshotSerializable=true` - Azure SQL Server. `jdbc:informatica:sqlserver://<host name>:<port>;DatabaseName=<database name>;SnapshotSerializable=true;SnapshotSerializable=true;EncryptionMethod=SSL;HostNameInCertificate=*.<host name in certificate>;ValidateServerCertificate=true` - Oracle. `jdbc:informatica:oracle://<host name>:<port>;SID=<database name>;MaxPooledStatements=20;CatalogOptions=0;BatchPerformanceWorkaround=true`

- Press **2** to configure the database connection using a custom JDBC connection string.
  The following table describes the properties you set:

Property	Description
EncryptionMethod	Required. Indicates whether data is encrypted when transmitted over the network. This parameter must be set to SSL.
ValidateServerCertificate	Optional. Indicates whether Informatica validates the certificate that is sent by the database server. If this parameter is set to True, Informatica validates the certificate that is sent by the database server. If you specify the HostNameInCertificate parameter, Informatica also validates the host name in the certificate. If this parameter is set to false, Informatica does not validate the certificate that is sent by the database server. Informatica ignores any truststore information that you specify.
HostNameInCertificate	Optional. Host name of the machine that hosts the secure database. If you specify a host name, Informatica validates the host name included in the connection string against the host name in the SSL certificate.

The Service Parameters section appears.

Application Service Details

If you create a Model Repository Service and a Data Integration Service to associate with Enterprise Data Lake during installation, specify the properties required to create the Data Integration Service.

1. Specify the name of the Data Integration Service.

   The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
2. Enter the name of the node where the service runs.
3. Enter the name of the Informatica license to associate with the service.
4. Enter the name of the Model Repository Service to associate with the Data Integration Service.
5. Specify the HTTP protocol type for the Data Integration Service, and then enter the port for each protocol you select.
   - To select HTTP only, press **1**.
   - To select HTTPS only, press **2**.
   - To select both HTTP and HTTPS, press **3**.
6. If you select HTTPS or both HTTP and HTTPS, select the SSL certificate to use.
   - To use the default Informatica SSL certificate contained in the default keystore and the default truststore, press **1**.
   - To use a custom SSL certificate contained in custom keystore and truststore files, press **2**, and then enter the path and file name for the keystore and truststore files. You must also enter the keystore and truststore passwords.

The Data Preparation Repository Database section appears.

Configure the Data Preparation Repository Database Details

Specify the Data Preparation repository database connection details. You can choose to use an Oracle database or a MySQL database for the Data Preparation repository database.

If you do not have the database details, you can enter placeholder values, and then create the Data Preparation Service. If you continue without specifying the database connection details, you cannot enable the Data Preparation Service.

Oracle

1. To use an Oracle database for the Data Preparation repository, press **1**.
2. Enter the connection properties for the database.

   The following table describes the connection properties:

Property	Description
Database Host Name	Host name of the machine that hosts the database.
Database Port Number	Port number for the database.

Property	Description
JDBC Parameters	Parameters required to connect to the database.
Custom JDBC Connection String	JDBC connection string to connect to the database. Format the string as follows: `jdbc:informatica:oracle://<database host name>:<port>;ServiceName=<service name>`

3. To connect to a secure database, press **2**, and then enter the secure connection properties.

   The following table describes the secure connection properties:

Property	Description
Truststore File	Path and file name for the database truststore file.
Truststore Password	Password for the database truststore file.
Secure JDBC Parameters	List of secure database parameters to connect to the database. Format the parameters as follows: `EncryptionMethod=SSL;HostNameInCertificate=<secure database host name>;ValidateServerCertificate=true`

4. Press **2** to continue.

The Data Preparation Service Details section appears.

MySQL

1. To use a MySQL database or a MariaDB database for the Data Preparation repository, press **2**.
2. Enter the connection properties for the database.
   The following table describes the connection properties:

Property	Description
Database Host Name	Host name of the machine that hosts the Data Preparation repository database.
Database User Name	Database user account to use to connect to the Data Preparation repository.
Database User Password	Password for the Data Preparation repository database user account.
Database Port Number	Port number for the database.
Database Name	Schema or database name of the Data Preparation repository database.

3. To connect to a secure database, press **2**, and then enter the secure connection properties.

The following table describes the secure connection properties:

Property	Description

Custom JDBC Connection String	Connection string to connect to the database.
	To connect to a non-secure database, format the string as follows:
	`jdbc:mysql://<database host name>:<port>`
	The connection string is optional if you connect to a non-secure database.
	To connect to an SSL-enabled database, format the string as follows:
	`verifyServerCertificate=true&useSSL=true&requireSSL=true`

Secure JDBC Parameters	String containing the path and file name for the database truststore file, and the truststore password. Format the string as follows: `trustCertificateKeyStoreUrl=file://<truststore path/truststore file name>&trustCertificatekeyStorePassword=<truststore password>`

4. Press **Enter** to continue.

The Data Preparation Service Details section appears.

Data Preparation Service Details

Create the Data Preparation Service. If you run the installer to create the Enterprise Data Lake Service and the Data Preparation Service, you must create both services on the same node.

1. Specify the name of the Data Preparation Service.

   The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
2. If you plan to use rules, you must associate a Data Integration Service and a Model Repository Service with the Data Preparation Service.
   - To skip associating a Model Repository Service and a Data Integration Service with the Data Preparation Service, press **1**.
   - To associate a Model Repository Service and a Data Integration Service with the Data Preparation Service, press **2**, and then enter the service names.
3. Enter the name of the node where the Data Preparation Service runs.
   - To create the service during installation, enter the name of the current node.
   - If you do not want to create the service during installation, do not enter a value. You can use the Administrator tool to create the service after installation.
4. Enter the name of the Informatica license to associate with the service.
5. Choose whether to enable secure communication for the service.
   - To enable secure communication for the service, press **1**.
   - To disable secure communication, press **2**.
6. If you enable secure communication for the service, select the SSL certificate to use.
   - To use the default Informatica SSL certificate contained in the default keystore and the default truststore, press **1**.
   - To use a custom SSL certificate contained in a custom keystore and truststore, press **2**, and then enter the path and file name for the keystore and truststore files. You must also enter the keystore and truststore passwords.

7. If you enable secure communication for the service, enter the port number for the HTTPS connection. If you enable non-secure communication for the service, enter the port number for the HTTP connection.
8. Select the Hadoop authentication mode.
   - To select the non-secure authentication mode, press **1**.
   - To select Kerberos authentication, press **2**.
9. If you select Kerberos, enter the authentication parameters.

   The following table describes the authentication parameters that you set if you select Kerberos:

Property	Description
HDFS Principal Name	Service Principal Name (SPN) for the data preparation Hadoop cluster. Specify the service principal name in the following format: user/_HOST@REALM.
Hadoop Impersonation User Name	User name to use in Hadoop impersonation as shown in the Impersonation User Name property for the Hadoop connection in the Administrator tool. If the Hadoop cluster uses Kerberos authentication, the Hadoop impersonation user must have read, write, and execute permissions on the HDFS storage location folder.
Kerberos Keytab File	Path and file name of the SPN keytab file for the user account to impersonate when connecting to the Hadoop cluster. The keytab file must be in a directory on the machine where the Data Preparation Service runs.
Fully Qualified Path to the Kerberos Configuration File	Path to the krb5.conf Kerberos configuration file.

10. Specify the HDFS storage location, HDFS connection, local storage location, and Solr port number details.

    The following table describes the properties you set:

Property	Description
HDFS Storage Location	HDFS location for data preparation file storage. If the Hadoop cluster uses Kerberos authentication, the Hadoop impersonation user must have read, write, and execute permissions on the HDFS storage location folder.
HDFS Connection	HDFS connection for data preparation file storage.
Local Storage Location	Directory for data preparation file storage on the node on which the Data Preparation Service runs. If the connection to the local storage fails, the Data Preparation Service recovers data preparation files from the HDFS storage location.
Solr port	Solr port number for the Apache Solr server used to provide data preparation recommendations.

11. Choose whether to enable the Data Preparation Service.
    - To enable the service at a later time using the Administrator tool, press **1**.
    - To enable the service after you complete the installation process, press **2**.

The Enterprise Data Lake Service Details section appears.

Enterprise Data Lake Service Details

Create the Enterprise Data Lake Service. If you run the installer to create the Enterprise Data Lake Service and the Data Preparation Service, you must create both services on the same node.

1. Specify the details for the Enterprise Data Lake Service.

   The following table describes the properties that you set:

Property	Description
Enterprise Data Lake Service Name	Name of the Enterprise Data Lake Service. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
Data Preparation Service Name	Name of the Data Preparation Service to associate with the Enterprise Data Lake Service.
Model Repository Service Name	Name of the Model Repository Service to associate with the Enterprise Data Lake Service.
Data Integration Service Name	Name of the Data Integration Service associated with the Enterprise Data Lake Service.
Node Name	To create the Enterprise Data Lake Service during installation, enter the name of the current node. If you do not want to create the service during installation, do not enter a value. You can use the Administrator tool to create the service after installation. If you create the Enterprise Data Lake Service and the Data Preparation Service during installation, you must create both services on the same node.
License	Enter the name of the Informatica license to associate with the service.

2. Choose whether to enable secure communication for the service.
   - To enable secure communication for the service, press **1**.
   - To disable secure communication, press **2**.
3. If you enable secure communication for the service, select the SSL certificate to use.
   - To use the default Informatica SSL certificate contained in the default keystore and the default truststore, press **1**.
   - To use a custom SSL certificate contained in a custom keystore and truststore, press **2**, and then enter the path and file name for the keystore and truststore files. You must also enter the keystore and truststore passwords.
4. If you enable secure communication for the service, enter the port number for the HTTPS connection. If you enable non-secure communication for the service, enter the port number for the HTTP connection.
5. Specify the data lake connection options.

The following table describes the properties that you set for the data lake connections:

Property	Description
HDFS Connection	HDFS connection for the data lake.
HDFS Working Directory	HDFS directory where the Enterprise Data Lake Service copies temporary data and files necessary for the service to run.
Hadoop Connection	Hadoop connection for the data lake.
Hive Connection	Hive connection for the data lake.
Hive Table Storage Format	Data storage format for the Hive tables. Select from the following options: - DefaultFormat - ORC - Parquet
Local System Directory	Local directory that contains the files downloaded from Enterprise Data Lake application, such as .csv and .tde files. The default directory is `/home/toolprod`.

6. Choose whether to enable logging of user activity events.
   - To disable logging of user activity events, press **1**.
   - To enable logging of user activity events, press **2**.
7. Select the Hadoop authentication mode.
   - To select the non-secure authentication mode, press **1**.
   - To select Kerberos authentication, press **2**.
8. If you select Kerberos, enter the authentication parameters.

   The following table describes the authentication properties that you must set if you select Kerberos:

Property	Description
Kerberos Principal	If the Hadoop cluster uses Kerberos authentication, specify the Service Principal Name (SPN) of the user account to impersonate when connecting to the data lake Hadoop cluster.
Kerberos KeyTab File	If the Hadoop cluster uses Kerberos authentication, specify the path and file name of the SPN keytab file for the user account to impersonate when connecting to the Hadoop cluster. The keytab file must be in a directory on the machine where the Enterprise Data Lake Service runs.

9. Choose whether to enable the Enterprise Data Lake Service immediately after you create the service.
   - To enable the service at a later time using the Administrator tool, press **1**.
   - To enable the service immediately after you create the service, press **2**.

CHAPTER 12

Install Enterprise Data Catalog

This chapter includes the following topics:

Install Enterprise Data Catalog Overview

Enterprise Data Catalog must be installed in the Informatica services installation directory. When you install Enterprise Data Catalog, you can also install Informatica services, or you can install it on a machine where Informatica services are already running.

This chapter contains instructions for the following Enterprise Data Catalog installations:

- If Informatica services are not installed, you can create a domain and install Enterprise Data Catalog. If a domain exists, you can install Informatica services on another machine to join the domain and install Enterprise Data Catalog.
- You can install Enterprise Data Catalog on an existing node in the Informatica domain.

Install Informatica Services with Enterprise Data Catalog

The first time you run the installer, you must create a domain. After you create a domain, you can run the installer on another machine to join the domain and install Enterprise Data Catalog.

Installing by Creating a Domain

Create a domain if you are installing for the first time or if you want to administer nodes in separate domains.

1. Log in to the machine with a system user account.
2. Close all other applications.

3. Run the ./install.sh command to start the installer.

   The installer displays the message to read Informatica documentation before you proceed with the installation.
4. Press **Y** to continue the installation.
5. Press **1** to install Informatica Big Data suite products.
6. Press **1** to run the Pre-installation System Check tool. The tool verifies if your machine meets the minimum system requirements to install or upgrade Informatica.

   Note: You can skip this step if you are sure that your machine meets the minimum system requirements to install or upgrade Informatica.
7. Press **3** to install Informatica.
8. Press **2** to agree to the terms and conditions of the installation or upgrade.
9. Press **2** to agree that you understand version 10.2.2 is specific to Big Data suite of products and continue with the installation.
10. Press **2** to install Informatica application services with Enterprise Data Catalog.

    The installer prompts you to confirm that the current version of the Informatica application services is not installed on the node.
11. Press **1** if you do not have the current version of the Informatica application services installed, else, press **2**.
12. Choose the Hadoop cluster type for Enterprise Data Catalog. Press **2** to deploy Enterprise Data Catalog on an embedded Hadoop distribution. Press **1** to deploy Enterprise Data Catalog on an existing Hadoop distribution.
    - If you chose the embedded Hadoop distribution, provide the following information after configuring the Informatica domain, the Model Repository Service, and the Data Integration Service:

Option	Description
SSH username	Username for the password-less Secure Shell (SSH) connection
Informatica Cluster service name	Name of the Informatica Cluster Service for the embedded cluster.
Informatica Cluster service port	Port number for the Informatica Cluster Service.
Ambari server host	Host information for the Ambari server. Ambari is a web-based tool for provisioning, managing, and monitoring Apache Hadoop clusters, which includes support for Hadoop HDFS, Hadoop MapReduce, Hive, HBase and ZooKeeper.
Comma-separated Ambari agent hosts	Applies to high availability. If you use multiple Ambari agent hosts, specify the comma-separated values of multiple Ambari agent host names.
Ambari web port	Port number where the Ambari server needs to run.
Catalog service name	Name of the catalog service.
Catalog service port	Port number of the catalog service.
Keytab Location	Applies to a Kerberos-enabled cluster. Location of the merged user and host keytab file.

Option	Description
Kerberos configuration file	Applies to a Kerberos-enabled cluster. Location of the Kerberos configuration file.

- Specify the following details if you select an existing cluster:

Property	Description
Hadoop Distribution	Select one of the following options: - ClouderaManager - HDInsight - Hortonworks
Cluster URL	Fully qualified host name to access the cluster.
Cluster URL Username	Username to access the cluster.
Cluster URL password	Password for the Cluster URL username.

- If you chose the existing Hadoop distribution as `ClouderaManager` or `Hortonworks`, provide the following information:

Option	Description
Catalog service name	Name of the catalog service.
Catalog service port	Port number of the catalog service.
Yarn resource manager URI	The service within Hadoop that submits the MapReduce tasks to specific nodes in the cluster. Use the following format: `<hostname>:<port>` Where • `hostname`is the name or IP address of the Yarn resource manager. • `port` is the port on which Yarn resource manager listens for Remote Procedure Calls (RPC).
Yarn resource manager http URI	http URI value for the Yarn resource manager.
Yarn resource manager scheduler URI	Scheduler URI value for the Yarn resource manager.
Zookeeper Cluster URI	The URI for the Zookeeper service, which is a high-performance coordination service for distributed applications.

Option	Description
HDFS namenode URI	The URI to access HDFS. Use the following format to specify the NameNode URI in the Cloudera distribution: `hdfs://<namenode>:<port>` Where • <namenode> is the host name or IP address of the NameNode. • <port> is the port that the NameNode listens for Remote Procedure Calls (RPC).
Service cluster name	Name of the service cluster. Ensure that you have a directory `/Informatica/LDM/<ServiceClusterName>` in HDFS before the installation is complete. **Note:** If you do not specify a service cluster name, Enterprise Data Catalog considers `DomainName_CatalogServiceName` as the default value. You must then have the `/Informatica/LDM/<DomainName>_<CatalogServiceName>` directory in HDFS. Otherwise, Catalog Service might fail.
History Server HTTP URI	HTTP URI to access the history server.
Is Cluster Secure ?	Set this property to one of the following values if you have an existing cluster that is secure: • 1: specifies that the existing cluster is not secure. • 2: specifies that the existing cluster is secure. Default is 1.
Is Cluster SSL Enabled?	Applicable only if you had selected the Hadoop distribution as `Hortonworks` and `ClouderaManager`. Set this property to one of the following values if you have an existing cluster that is enabled for SSL: • 1: specifies that the existing cluster is not enabled for SSL. • 2: specifies that the existing cluster is enabled for SSL. Default is 1.
Enable Kerberos Authentication	Applicable only if you had selected the Hadoop distribution as `Hortonworks` and `ClouderaManager`. Set this property to one of the following values if you have an existing cluster that is enabled for Keberos: • 1: specifies that the existing cluster is not enabled for Kerberos. • 2: specifies that the existing cluster is enabled for Kerberos.

Depending on the settings that you specify, Enterprise Data Catalog creates an Informatica Cluster Service for embedded Hadoop distribution.

13. Press **Enter** to continue.
14. Press **2** if you want the installer to tune the Informatica application services based on the size of data that you want to deploy.

    The installer displays the following options for various data sizes:

    - Sandbox
    - Basic

- Standard
- High Concurrency and High Volume

15. Type the path and file name of the Informatica license key and press **Enter**.
16. Type the absolute path for the installation directory.

    The directory names in the path must not contain spaces or the following special characters: @|* $ # ! % ( ) { } [ ] , ; ' Default is /home/toolinst.

    Note: Informatica recommends using alphanumeric characters in the installation directory path. If you use a special character such as á or €, unexpected results might occur at run time.
17. Press **2** to run the pre-validation utility. The utility helps you validate the prerequisites to install Enterprise Data Catalog in an embedded cluster. The utility also validates the Informatica domain, cluster hosts, and the Hadoop cluster services configuration.

    The installer prompts you to confirm if you want to enable Kerberos authentication for the cluster.
18. Press **2** if you want to enable Kerberos authentication for the cluster and provide the following details:
    a. **Keytab Location**. Location of the merged user and host keytab file.
    b. **Kerberos Configuration File**. Location of the Kerberos configuration file.
19. Type the gateway user name and press **Enter**. Default is **root**.
20. Type the Informatica Hadoop cluster gateway hostname in the following format: `<hostname>.<FQDN>` and press **Enter**.
21. Type the list of comma-separated Informatica Hadoop cluster nodes as shown in the following format: `<hostname>.<FQDN>, <hostname1>.<FQDN>, <hostname2>.<FQDN>` and press **Enter**.
22. Type the Informatica Hadoop cluster gateway port and press **Enter**. Default is **8080**.

    To avoid a port conflict, make sure that you do not configure Oracle with port 8080 on the same machine where Informatica Cluster Service runs.
23. Type the path to the working directory, and press **Enter**. The path indicates the location where you want to mount the Informatica Cluster Service.

    The installer starts the pre-validation utility.
24. Press **Enter** to continue after running the pre-validation utility.
25. Review the installation information, and press **Enter** to continue.

    The installer copies the Enterprise Data Catalog files to the installation directory. You see a prompt to create or join a domain.
26. Press **1** to create a domain.

    When you create a domain, the node that you create becomes a gateway node in the domain. The gateway node contains a Service Manager that manages all domain operations.
27. To enable secure communication for services in the domain, press **2**. To disable secure communication for the domain, press **1**.

    By default, if you enable secure communication for the domain, the installer sets up an HTTPS connection for the Informatica Administrator. You can also create a domain configuration repository on a secure database.
28. Type the connection details for Informatica Administrator.
    a. If you do not enable secure communication for the domain, you can specify whether to set up a secure HTTPS connection for the Informatica Administrator.

The following table describes the options available to enable or disable a secure connection to Informatica Administrator:

Option	Description
1 - Enable HTTPS for Informatica Administrator	Set up a secure connection to Informatica Administrator.
2 - Disable HTTPS	Do not set up a secure connection to Informatica Administrator.

b. If you enable secure communication for the domain or if you enable HTTPS connection for the Informatica Administrator, enter the keystore file and port number for the HTTPS connection to Informatica Administrator.

The following table describes the connection information you must enter if you enable HTTPS:

Option	Description
Port	Port number for the HTTPS connection.
Keystore file	Select whether to use a keystore file generated by the installer or a keystore file you create. You can use a keystore file with a self-signed certificate or a certificate signed by a certification authority. 1 - Use a keystore generated by the installer 2 - Specify a keystore file and password If you select to use a keystore file generated by the installer, the installer creates a self-signed keystore file named Default.keystore in the following location: `<Informatica installation directory>/tomcat/conf/`

c. If you specify the keystore, enter the password and location of the keystore file.

29. Press **2** if you want to enable Single sign-on using SAML authentication for Enterprise Data Catalog applications.
30. Type the SAML Identity Provider (IdP) URL and press **Enter**.

    See the section *Configure Single Sign-on with SAML Authentication* for information about configuration you must complete after you install Enterprise Data Catalog.

    If you enabled secure communication for the domain, the **Domain Security - Secure Communication** section appears. If you did not enable secure communication for the domain, the **Domain Configuration Repository** section appears.
31. In the Domain Security - Secure Communication section, specify whether to use the default Informatica SSL certificates or to use your SSL certificates to secure domain communication.
    a. Select the type of SSL certificates to use.

The following table describes the options for the SSL certificates that you can use to secure the Informatica domain:

Option	Description
1 - Use the default Informatica SSL certificate files	Use the default SSL certificates provided by Informatica. **Note:** If you do not provide an SSL certificate, Informatica uses the same default private key for all Informatica installations. If you use the default Informatica keystore and truststore files, the security of your domain could be compromised. To ensure a high level of security for the domain, select the option to specify the location of the SSL certificate files.
2 - Specify the location of the SSL certificate files	Use SSL certificates that you provide. You must specify the location of the keystore and truststore files. You can provide a self-signed certificate or a certificate issued by a certificate authority (CA). You must provide SSL certificates in PEM format and in Java Keystore (JKS) files. Informatica requires specific names for the SSL certificate files for the Informatica domain. You must use the same SSL certificates for all nodes in the domain. Store the truststore and keystore files in a directory accessible to all the nodes in the domain and specify the same keystore file directory and truststore file directory for all nodes in the same domain.

b. If you provide the SSL certificate, specify the location and passwords of the keystore and truststore files.

The following table describes the parameters that you must enter for the SSL certificate files:

Property	Description
Keystore file directory	Directory that contains the keystore files. The directory must contain files named infa_keystore.jks and infa_keystore.pem.
Keystore password	Password for the keystore infa_keystore.jks.
Truststore file directory	Directory that contains the truststore files. The directory must contain files named infa_truststore.jks and infa_truststore.pem.
Truststore password	Password for the infa_truststore.jks file.

The Domain Configuration Repository section appears.

32. Select the database to use for the domain configuration repository.

The following table lists the databases you can use for the domain configuration repository:

Prompt	Description
Database type	Type of database for the domain configuration repository. Select from the following options: 1 - Oracle 2 - Microsoft SQL Server 3 - IBM DB2 4 - Sybase ASE

The Informatica domain configuration repository stores metadata for domain operations and user authentication. The domain configuration repository must be accessible to all gateway nodes in the domain.

33. Enter the properties for the database user account.

    The following table lists the properties for the database user account:

Property	Description
Database user ID	Name for the domain configuration database user account.
User password	Password for the domain configuration database user account.

34. Choose whether to create a secure domain configuration repository.

    You can create a domain configuration repository in a database secured with the SSL protocol. To create a domain configuration repository in a secure database, press 1.

    To create a domain configuration repository in an unsecure database, press 2.

35. If you do not want to create a secure domain configuration repository, enter the parameters for the database.

    a. If you select IBM DB2, select whether to configure a tablespace and enter the tablespace name.

       The following table describes the properties that you must configure for the IBM DB2 database:

Property	Description
Configure tablespace	Select whether to specify a tablespace: 1 - No 2 - Yes In a single-partition database, if you select No, the installer creates the tables in the default tablespace. In a multi-partition database, you must select Yes.
Tablespace	Name of the tablespace in which to create the tables. Specify a tablespace that meets the pageSize requirement of 32768 bytes. In a single-partition database, if you select Yes to configure the tablespace, enter the name of the tablespace in which to create the tables. In a multi-partition database, specify the name of the tablespace that resides in the catalog partition of the database.

b. If you select Microsoft SQL Server, enter the schema name for the database.

The following table describes the properties that you must configure for the Microsoft SQL Server database:

Property	Description
Schema name	Name of the schema that will contain domain configuration tables. If this parameter is blank, the installer creates the tables in the default schema.

c. To enter the JDBC connection information using the JDBC URL information, press **1**. To enter the JDBC connection information using a custom JDBC connection string, press **2**.

d. Enter the JDBC connection information.

- To enter the connection information using the JDBC URL information, specify the JDBC URL properties.
  The following table describes the database connection information:

Prompt	Description
Database host name	Host name for the database.
Database port number	Port number for the database.
Database service name	Password for the domain configuration database user account. Service name for Oracle and IBM DB2 databases or database name for Microsoft Microsoft SQL Server and Sybase ASE.
Configure JDBC Parameters	Select whether to add additional JDBC parameters to the connection string: 1 - Yes 2 - No If you select Yes, enter the parameters or press Enter to accept the default. If you select No, the installer creates the JDBC connection string without parameters.

- To enter the connection information using a custom JDBC connection string, type the connection string.
  Use the following syntax for the JDBC connection string for the databases:
  IBM DB2

  ```
  jdbc:Informatica:db2://host_name:port_no;DatabaseName=
  ```

  Oracle

  ```
  jdbc:Informatica:oracle://host_name:port_no;ServiceName=
  ```

  Microsoft SQL Server

  ```
  jdbc:Informatica:sqlserver://
  host_name:port_no;SelectMethod=cursor;DatabaseName=
  ```

  Sybase

  ```
  jdbc:Informatica:sybase://host_name:port_no;DatabaseName=
  ```

  Verify that the connection string contains all the connection parameters required by your database system.

36. If you create a secure domain configuration repository, enter the parameters for the secure database.

    If you create the domain configuration repository on a secure database, you must provide the truststore information for the database. You must also provide a JDBC connection string that includes the security parameters for the database.

    The following table describes the options available to create a secure domain configuration repository database:

Property	Description
Database truststore file	Path and file name of the truststore file for the secure database.
Database truststore password	Password for the truststore file.
Custom JDBC Connection String	Complete JDBC connection for the secure database, including the host name and port number and the secure database parameters.

    In addition to the host name and port number for the database server, you must include the following secure database parameters:
    EncryptionMethod

    Required. Indicates whether data is encrypted when transmitted over the network. This parameter must be set to `SSL`.

    ValidateServerCertificate

    Optional. Indicates whether Informatica validates the certificate that the database server sends.

    If this parameter is set to True, Informatica validates the certificate that the database server sends. If you specify the HostNameInCertificate parameter, Informatica also validates the host name in the certificate.

    If this parameter is set to False, Informatica does not validate the certificate that the database server sends. Informatica ignores any truststore information that you specify.

    Default is True.

HostNameInCertificate

Optional. Host name of the machine that hosts the secure database. If you specify a host name, Informatica validates the host name included in the connection string against the host name in the SSL certificate.

cryptoProtocolVersion

Required. Specifies the cryptographic protocol to use to connect to a secure database. You can set the parameter to `cryptoProtocolVersion=TLSv1.1` or `cryptoProtocolVersion=TLSv1.2` based on the cryptographic protocol used by the database server.

You can use the following syntax for the connection strings:

- Oracle: `jdbc:Informatica:oracle://host_name:port_no;ServiceName=service_name;EncryptionMethod=SSL;HostNameInCertificate=DB_host_name;ValidateServerCertificate=true_or_false`
- IBM DB2: `jdbc:Informatica:db2://host_name:port_no;DatabaseName=database_name;EncryptionMethod=SSL;HostNameInCertificate=DB_host_name;ValidateServerCertificate=true_or_false`
- Microsoft SQL Server: `jdbc:Informatica:sqlserver://host_name:port_no;SelectMethod=cursor;DatabaseName=database_name;EncryptionMethod=SSL;HostNameInCertificate=DB_host_name;ValidateServerCertificate=true_or_false`

Note: The installer does not validate the connection string. Verify that the connection string contains all the connection parameters and security parameters required by your database.

37. If the database contains a domain configuration repository for a previous domain, choose to overwrite the data or set up another database.

    The following table describes the options of overwriting the data or setting up another database when you create a domain configuration repository for a previous domain:

Option	Description
1 - OK	Enter the connection information for a new database.
2 - Continue	The installer overwrites the data in the database with new domain configuration.

38. In the **Domain Security - Encryption Key** section, enter the keyword and encryption key directory for the Informatica domain.

The following table describes the encryption key parameters that you must specify:

Property	Description
Keyword	Keyword to use to create a custom encryption key to secure sensitive data in the domain. The keyword must meet the following criteria: - From 8 to 20 characters long - Includes at least one uppercase letter - Includes at least one lowercase letter - Includes at least one number - Does not contain spaces The encryption key is created based on the keyword that you provide when you create the Informatica domain.
Encryption key directory	Directory in which to store the encryption key for the domain. The default location is the following directory: `<Informatica installation directory>/isp/config/keys`.

The installer sets different permissions to the directory and the files in the directory.

39. Press **Enter**.

    The **Domain and Node Configuration** section appears.

40. Enter the information for the domain and the node that you want to create.

    The following table describes the properties that you set for the domain and gateway node.

Property	Description
Domain name	Name of the domain to create. The default domain name is Domain_<MachineName>. The name must not exceed 128 characters and must be 7-bit ASCII only. It cannot contain a space or any of the following characters: ` % * + ; " ? , < > \ /
Node host name	Host name of the machine on which to create the node. The node host name cannot contain the underscore (_) character. If the machine has a single network name, use the default host name. If the a machine has multiple network names, you can modify the default host name to use an alternate network name. Optionally, you can use the IP address. **Note:** Do not use localhost. The host name must explicitly identify the machine.
Node name	Name of the node to create on this machine. The node name is not the host name for the machine.
Node port number	Port number for the node. The default port number for the node is 6005. If the port number is not available on the machine, the installer displays the next available port number.
Domain user name	User name for the domain administrator. You can use this user name to initially log in to Informatica Administrator. Use the following guidelines: - The name is not case sensitive and cannot exceed 128 characters. - The name cannot include a tab, newline character, or the following special characters: % * + / ? ; < > - The name can include an ASCII space character except for the first and last character. Other space characters are not allowed.

Property	Description
Domain password	Password for the domain administrator. The password must be more than 2 characters and must not exceed 16 characters.
Confirm password	Enter the password again to confirm.

41. Select whether to display the default ports for the domain and node components assigned by the installer.

    The following table describes the advanced port configuration page:

Prompt	Description
Display advanced port configuration page	Select whether to display the port numbers for the domain and node components assigned by the installer: 1 - No 2 - Yes If you select Yes, the installer displays the default port numbers assigned to the domain components. You can specify the port numbers to use for the domain and node components. You can also specify a range of port numbers to use for the service process that will run on the node. You can use the default port numbers or specify new port numbers. Verify that the port numbers you enter are not used by other applications.

42. If you display the port configuration page, enter new port numbers at the prompt or press **Enter** to use the default port numbers.

Port	Description
Service Manager port	Port number used by the Service Manager on the node. The Service Manager listens for incoming connection requests on this port. Client applications use this port to communicate with the services in the domain. The Informatica command line programs use this port to communicate to the domain. This is also the port for the SQL data service JDBC/ODBC driver. Default is 6006.
Service Manager Shutdown port	Port number that controls server shutdown for the domain Service Manager. The Service Manager listens for shutdown commands on this port. Default is 6007.
Informatica Administrator port	Port number used by Informatica Administrator. Default is 6008.
Informatica Administrator HTTPS port	No default port. Enter the required port number when you create the service. Setting this port to 0 disables an HTTPS connection to the Administrator tool.
Informatica Administrator shutdown port	Port number that controls server shutdown for Informatica Administrator. Informatica Administrator listens for shutdown commands on this port. Default is 6009.

Port	Description
Minimum port number	Lowest port number in the range of dynamic port numbers that can be assigned to the application service processes that run on this node. Default is 6014.
Maximum port number	Highest port number in the range of dynamic port numbers that can be assigned to the application service processes that run on this node. Default is 6114.

43. Choose whether you want to create Model Repository Service, Data Integration Service, and Catalog Service as part of the installation. You can create these services after installation in Informatica Administrator. Press **1** to create the services, or press **2** to complete the installation without creating the services.

    If you pressed 1, the **Model Repository Service Database** section appears.
44. If you pressed 1, choose the database type, and enter the database parameters for the Model repository.
45. Choose whether you want to configure a secure database. Press **1** to configure a secure database, or press **2** to skip the step.
46. To configure JDBC connection information, press **1** and enter the JDBC parameters. Press **2** to skip configuring the JDBC connection.
47. Choose the database type for the Model repository, and enter the credentials including the database user ID and user password.
48. Optionally, configure the JDBC connection and its parameters.
49. Enter the following information: Model Repository Service name , Data Integration Service name, and the port number for the Data Integration Service if you do not want to use the default value.

Option	Description
MRS name	Name of the Model Repository Service.
DIS name	Name of the Data Integration Service.
HTTP protocol type	Security protocol that the Data Integration Service uses.
Port	Port number.

    You see messages about creating Model Repository Service and Data Integration Service.
50. Enter the following required information in addition to the Model Repository Service and Data Integration Service to create the profiling warehouse and reference data warehouse databases:

    Reference data warehouse database type

    Database type for the reference data warehouse. The reference data warehouse supports IBM DB2 UDB, Microsoft SQL Server, or Oracle.

    Reference data warehouse database host name

    The name of the machine hosting the reference data warehouse.

    Profiling warehouse database type

    Database type for the profiling warehouse. The profiling warehouse supports IBM DB2 UDB, Microsoft SQL Server, or Oracle.

    Profiling warehouse database host name

    The name of the machine hosting the profiling warehouse.

The Post-installation Summary indicates whether the installation completed successfully. You can view the installation log files to get more information about the tasks performed by the installer and to view configuration properties for the installed components.

Installing by Joining a Domain

You can join a domain if you are installing on multiple machines and you have created a domain on another machine.

1. Log in to the machine with a system user account.
2. Close all other applications.
3. Run the ./install.sh command to start the installer.

   The installer displays the message to read Informatica documentation before you proceed with the installation.
4. Press **Y** to continue the installation.
5. Press **1** from to install Informatica Big Data suite products.
6. Press **1** to run the Pre-installation System Check tool. The tool verifies if your machine meets the minimum system requirements to install or upgrade Informatica.

   Note: You can skip this step if you are sure that your machine meets the minimum system requirements to install or upgrade Informatica.
7. Press **3** to install Informatica.
8. Press **2** to agree to the terms and conditions of the installation or upgrade.
9. Press **2** to agree that you understand version 10.2.2 is specific to Big Data suite of products and continue with the installation.
10. Press **2** to install Informatica Services with Enterprise Data Catalog.

    The installer prompts you to confirm that the current version of the Informatica Services is not installed on the node.
11. Press **1** if you do not have the current version of the Informatica Services installed, else, press **2**.
12. Choose the Hadoop cluster type for Enterprise Data Catalog. Press 2 to deploy Enterprise Data Catalog on an internal Hadoop distribution on HortonWorks using Ambari tool. Press 1 to deploy Enterprise Data Catalog on an existing Hadoop distribution on Cloudera, HortonWorks, or Azure HDInsight.

    Depending on the settings that you specify, Enterprise Data Catalog creates an Informatica Cluster Service for internal Hadoop distribution.
13. If you chose the embedded Hadoop distribution, provide the following information after configuring the Informatica domain, the Model Repository Service, and the Data Integration Service:

    •

Option	Description
SSH username	Username for the password-less Secure Shell (SSH) connection
Informatica Cluster service name	Name of the Informatica Cluster Service for the internal cluster.
Informatica Cluster service port	Port number for the Informatica Cluster Service.

Option	Description
Ambari server host	Host information for the Ambari server. Ambari is a web-based tool for provisioning, managing, and monitoring Apache Hadoop clusters, which includes support for Hadoop HDFS, Hadoop MapReduce, Hive, HBase and ZooKeeper.
Comma-separated Ambari agent hosts	Applies to high availability. If you use multiple Ambari agent hosts, specify the comma-separated values of multiple Ambari agent host names.
Ambari web port	Port number where the Ambari server needs to run.
Catalog service name	Name of the catalog service.
Catalog service port	Port number of the catalog service.
Keytab Location	Applies to a Kerberos-enabled cluster. Location of the merged user and host keytab file.
Kerberos configuration file	Applies to a Kerberos-enabled cluster. Location of the Kerberos configuration file.

- Specify the following details if you select an existing cluster:

Property	Description
Hadoop Distribution	Select one of the following options: - ClouderaManager - HDInsight - Hortonworks
Cluster URL	Fully qualified host name to access the cluster.
Cluster URL Username	Username to access the cluster.
Cluster URL password	Password for the Cluster URL username.

- If you chose the existing Hadoop distribution as `ClouderaManager` or `Hortonworks`, provide the following information:

Option	Description
Catalog service name	Name of the catalog service.
Catalog service port	Port number of the catalog service.

Option	Description
Yarn resource manager URI	The service within Hadoop that submits the MapReduce tasks to specific nodes in the cluster. Use the following format: <hostname>:<port> Where • hostnameis the name or IP address of the Yarn resource manager. • port is the port on which Yarn resource manager listens for Remote Procedure Calls (RPC).
Yarn resource manager http URI	http URI value for the Yarn resource manager.
Yarn resource manager scheduler URI	Scheduler URI value for the Yarn resource manager.
Zookeeper Cluster URI	The URI for the Zookeeper service, which is a high-performance coordination service for distributed applications.
HDFS namenode URI	The URI to access HDFS. Use the following format to specify the NameNode URI in the Cloudera distribution: hdfs://<namenode>:<port> Where • <namenode> is the host name or IP address of the NameNode. • <port> is the port that the NameNode listens for Remote Procedure Calls (RPC).
Service cluster name	Name of the service cluster. Ensure that you have a directory /Informatica/LDM/<ServiceClusterName> in HDFS before the installation is complete. **Note:** If you do not specify a service cluster name, Enterprise Data Catalog considers DomainName_CatalogServiceName as the default value. You must then have the /Informatica/LDM/<DomainName>_<CatalogServiceName> directory in HDFS. Otherwise, Catalog Service might fail.
History Server HTTP URI	HTTP URI to access the history server.
Is Cluster Secure ?	Set this property to one of the following values if you have an existing cluster that is secure: • 1: specifies that the existing cluster is not secure. • 2: specifies that the existing cluster is secure. Default is 1.
Is Cluster SSL Enabled?	Applicable only if you had selected the Hadoop distribution as Hortonworks and ClouderaManager. Set this property to one of the following values if you have an existing cluster that is enabled for SSL: • 1: specifies that the existing cluster is not enabled for SSL. • 2: specifies that the existing cluster is enabled for SSL. Default is 1.

Option	Description
Enable Kerberos Authentication	Set this property to one of the following values if you have an existing cluster that is enabled for Keberos: • 1: specifies that the existing cluster is not enabled for Kerberos. • 2: specifies that the existing cluster is enabled for Kerberos.

14. Press **Enter** to continue.

    You see a prompt message about the license key file.

15. Press **2** if you want the installer to tune the Informatica application services based on the size of data deployed.

    The installer displays the following options for various data sizes:

    - Sandbox
    - Basic
    - Standard
    - High Concurrency and High Volume

16. Type the path and file name of the Informatica license key and press **Enter**.
17. Type the absolute path for the installation directory.

    The directory names in the path must not contain spaces or the following special characters: @|* $ # ! % ( ) { } [ ] , ; ' Default is /home/toolinst.

    Note: Informatica recommends using alphanumeric characters in the installation directory path. If you use a special character such as á or €, unexpected results might occur at run time.

18. Press **2** to run the pre-validation utility. The utility helps you validate the prerequisites to install Enterprise Data Catalog in an embedded cluster. The utility also validates the Informatica domain, cluster hosts, and the Hadoop cluster services configuration.

    The installer prompts you to confirm if you want to enable Kerberos authentication for the cluster.

19. Press **2** if you want to enable Kerberos authentication for the cluster and provide the following details:

    a. **Keytab Location**. Location of the merged user and host keytab file.

    b. **Kerberos Configuration File**. Location of the Kerberos configuration file.

20. Provide the gateway user name and press **Enter**. Default is **root**.
21. Provide the Informatica Hadoop cluster gateway hostname in the following format: `<hostname>.<FQDN>` and press **Enter**.
22. Provide the list of comma-separated Informatica Hadoop cluster nodes as shown in the following format: `<hostname>.<FQDN>, <hostname1>.<FQDN>, <hostname2>.<FQDN>` and press **Enter**.
23. Provide the Informatica Hadoop cluster gateway port and press **Enter**. Default is **8080**.

    Make sure that you do not configure Oracle with port 8080 on the same machine where Informatica Cluster Service runs.

24. Provide the path to the working directory and press **Enter**. The path indicates the location where you want to mount the Informatica Cluster Service.

    The installer starts the pre-validation utility.

25. Press **Enter** to continue after running the pre-validation utility.
26. Review the installation information, and press **Enter** to continue.

The installer copies the Enterprise Data Catalog files to the installation directory. You see a prompt to create or join a domain.

27. Press **2** to join a domain.

    The installer creates a node on the machine where you install. You can specify the type of node to create and the domain to join.

28. Specify whether the domain you want to join has the secure communication option enabled.

    Press 1 to join an unsecure domain, or press 2 to join a secure domain.

29. Select the type of node you want to create.

    The following table describes that types of nodes that you can create:

Property	Description
Configure this node as a gateway	Select whether to configure the node as a gateway or worker node. 1 - Yes 2 - No Select 1 to configure a gateway node or 2 to configure a worker node.

    If you configure the node as a gateway, you can enable a secure HTTPS connection to the Informatica Administrator.

30. Specify the connection details to Informatica Administrator.

    a. Specify whether to set up a secure HTTPS connection to the Informatica Administrator.

       The following table describes the options available to enable or disable a secure connection to Informatica Administrator:

Option	Description
1 - Enable HTTPS for Informatica Administrator	Set up a secure connection to Informatica Administrator.
2 - Disable HTTPS	Do not set up a secure connection to Informatica Administrator.

    b. If you enable HTTPS connection for the Informatica Administrator, enter the keystore file and port number to use to secure the connection.

The following table describes the connection information you must enter if you enable HTTPS:

Option	Description
Port	Port number for the HTTPS connection.
Keystore file	Select whether to use a keystore file generated by the installer or a keystore file you create. You can use a keystore file with a self-signed certificate or a certificate signed by a certification authority. 1 - Use a keystore generated by the installer 2 - Specify a keystore file and password If you select to use a keystore file generated by the installer, the installer creates a self-signed keystore file named Default.keystore in the following location: `<Informatica installation directory>/tomcat/conf/`

c. If you specify the keystore, enter the password and location of the keystore file.

31. Press **2** if you want to enable Single sign-on using SAML authentication for Enterprise Data Catalog applications.
32. Provide the SAML Identity Provider (IdP) URL and press **Enter**.

    See the section *Configure Single Sign-on with SAML Authentication* for information about configuration you must complete after you install Enterprise Data Catalog.

    If you enabled secure communication for the domain, the **Domain Security - Secure Communication** section appears. If you did not enable secure communication for the domain, the **Domain Configuration Repository** section appears.
33. In the Domain Security - Secure Communication section, specify whether to use the default Informatica SSL certificates or to use your SSL certificates to secure domain communication.

    a. Select the type of SSL certificates to use.

    The following table describes the options for the SSL certificates that you can use to secure the Informatica domain:

Option	Description
1 - Use the default Informatica SSL certificate files	Use the default SSL certificates provided by Informatica. **Note:** If you do not provide an SSL certificate, Informatica uses the same default private key for all Informatica installations. If you use the default Informatica keystore and truststore files, the security of your domain could be compromised. To ensure a high level of security for the domain, select the option to specify the location of the SSL certificate files.
2 - Specify the location of the SSL certificate files	Use SSL certificates that you provide. You must specify the location of the keystore and truststore files. You can provide a self-signed certificate or a certificate issued by a certificate authority (CA). You must provide SSL certificates in PEM format and in Java Keystore (JKS) files. Informatica requires specific names for the SSL certificate files for the Informatica domain. You must use the same SSL certificates for all nodes in the domain. Store the truststore and keystore files in a directory accessible to all the nodes in the domain and specify the same keystore file directory and truststore file directory for all nodes in the same domain.

b. If you provide the SSL certificate, specify the location and passwords of the keystore and truststore files.

The following table describes the parameters that you must enter for the SSL certificate files:

Property	Description
Keystore file directory	Directory that contains the keystore files. The directory must contain files named infa_keystore.jks and infa_keystore.pem.
Keystore password	Password for the keystore infa_keystore.jks.
Truststore file directory	Directory that contains the truststore files. The directory must contain files named infa_truststore.jks and infa_truststore.pem.
Truststore password	Password for the infa_truststore.jks file.

The Domain Configuration Repository section appears.

34. At the prompt, enter the information for the domain that you want to join.

The following table describes the properties that you specify for the domain:

Property	Description
Domain name	Name of the domain to join.
Gateway node host	Host name of the machine that hosts the gateway node for the domain.
Gateway node port	Port number of the gateway node.
Domain user name	User name of the administrator for the domain you want to join.
Domain password	Password for the domain administrator.

The **Domain Security - Encryption Key** section appears.

35. Enter the encryption key information for the Informatica domain that you want to join.

If the location of the encryption key in the gateway node is not accessible to the current node, copy the encryption key file to an accessible directory. You might need to assign read permission to the directory that contains the encryption key file on gateway node before you can copy the file.

The following table describes the encryption key parameters that you must specify when you join a domain:

Property	Description
Select the encryption key	Path and file name of the encryption key for the Informatica domain that you want to join. All nodes in the Informatica domain use the same encryption key. You must specify the encryption key file created on the gateway node for the domain that you want to join. If you copied the encryption key file to a temporary directory to make it accessible to the nodes in the domain, specify the path and file name of the encryption key file in the temporary directory.
Encryption key directory	Directory in which to store the encryption key on the node created during this installation. The installer copies the encryption key file for the domain to the encryption key directory on the new node.

36. On the Join Domain Node Configuration section, enter the information for the node you want to create.

    The following table describes the properties that you set for the node:

Property	Description
Node Host name	Host name for the node. The node host name cannot contain the underscore (_) character. **Note:** Do not use localhost. The host name must explicitly identify the machine.
Node name	Name of the Informatica node to create on this machine. The node name is not the host name for the machine.
Node port number	Port number for the node.
Database truststore file	Path and file name of the truststore file for the secure database. Select the same database truststore file used by the master gateway node in the domain. Available when you join a gateway node to a domain that uses a domain configuration repository database that is secured with the SSL protocol.
Truststore password	Password for the database truststore file for the secure database. Available when you join a gateway node to a domain that uses a domain configuration repository database that is secured with the SSL protocol.

37. Select whether to display the default ports for the domain and node components assigned by the installer.

The following table describes the advanced port configuration page:

Prompt	Description
Display advanced port configuration page	Select whether to display the port numbers for the domain and node components assigned by the installer: 1 - No 2 - Yes If you select Yes, the installer displays the default port numbers assigned to the domain components. You can specify the port numbers to use for the domain and node components. You can also specify a range of port numbers to use for the service process that will run on the node. You can use the default port numbers or specify new port numbers. Verify that the port numbers you enter are not used by other applications.

38. If you display the port configuration page, enter new port numbers at the prompt or press **Enter** to use the default port numbers.

Port	Description
Service Manager port	Port number used by the Service Manager on the node. The Service Manager listens for incoming connection requests on this port. Client applications use this port to communicate with the services in the domain. The Informatica command line programs use this port to communicate to the domain. This is also the port for the SQL data service JDBC/ODBC driver. Default is 6006.
Service Manager Shutdown port	Port number that controls server shutdown for the domain Service Manager. The Service Manager listens for shutdown commands on this port. Default is 6007.
Informatica Administrator port	Port number used by Informatica Administrator. Default is 6008.
Informatica Administrator HTTPS port	No default port. Enter the required port number when you create the service. Setting this port to 0 disables an HTTPS connection to the Administrator tool.
Informatica Administrator shutdown port	Port number that controls server shutdown for Informatica Administrator. Informatica Administrator listens for shutdown commands on this port. Default is 6009.
Minimum port number	Lowest port number in the range of dynamic port numbers that can be assigned to the application service processes that run on this node. Default is 6014.
Maximum port number	Highest port number in the range of dynamic port numbers that can be assigned to the application service processes that run on this node. Default is 6114.

39. Choose whether you want to configure a secure database. Press **1** to configure a secure database, or press **2** to skip the step.
40. To configure JDBC connection information, press **1** and enter the JDBC parameters. Press **2** to skip configuring the JDBC connection.
41. Choose the database type for the Model repository, and enter the credentials including the database user ID and user password.
42. Optionally, configure the JDBC connection and its parameters.

43. Enter the following information: Model Repository Service name , Data Integration Service name, and the port number for the Data Integration Service if you do not want to use the default value.

Option	Description
MRS name	Name of the Model Repository Service.
DIS name	Name of the Data Integration Service.
HTTP protocol type	Security protocol that the Data Integration Service uses.
Port	Port number.

You see messages about creating Model Repository Service and Data Integration Service.

The Post-installation Summary indicates whether the installation completed successfully. You can view the installation log files to get more information about the tasks performed by the installer and to view configuration properties for the installed components.

Installing Enterprise Data Catalog on a Domain Node

You can use the Informatica installer to install Enterprise Data Catalog after installing Informatica. To install Enterprise Data Catalog after installing Informatica, perform the following steps:

1. Log in to the machine with a system user account.
2. Shut down the Informatica domain.
3. Close all applications.
4. On a shell command line, run the install.sh from the root directory.

   The installer displays the message to verify that the locale environment variables are set.
5. Press 1 to select the option to install or upgrade Informatica.

   The installer checks if the current version of Informatica is installed.
6. Press 2 to install Informatica Services with Enterprise Data Catalog.

   The installer prompts you to confirm that the current version of the Informatica Services is installed.
7. Press 2 to install Enterprise Data Catalog. Pressing this option assumes that the present version of Informatica is installed.
8. Provide the *<INFA_HOME>* location when prompted by the installer to complete the installation. INFA_HOME refers to the directory where Enterprise Data Catalog must be installed.

The Post-installation Summary indicates whether the installation completed successfully. You can view the installation log files to get more information about the tasks performed by the installer and to view configuration properties for the installed components.

Note: The instructions provided in this section assume that you have created the Informatica application services when you installed Informatica. If you had not created the services, see the *Installing By Joining a Domain* section for more information about creating application services.

After you complete the installation, perform the following steps:

1. Delete the following directories:
   - INFA_HOME/service/work_dir
   - INFA_HOME/tomcat/bin/workspace/.metadata
2. Start the Informatica domain.
3. Enable the Model Repository Service and upgrade the Model Repository Service content using one of the following methods:
   - Informatica Administrator: select the Model Repository Service and click **Actions** > **Repository Contents** > **Upgrade**.
   - Informatica Command Line Interface: run the INFA_HOME/isp/bin/infacmd.sh mrs upgradeContents -dn DOMAINNAME -un domainUsername -pw domainPassword -sn MRSServiceName command
4. Create and enable the Catalog Service. Make sure that you use the upgraded Model Repository Service.

Resume the Installer

When the installation process stops midway, you can resume the installation from the point of failure or exit.

When you run the server installer and the installation process fails, you can still resume from the previous service configuration and recover the last entered details for that service installation.

The install process might fail for reasons such as network outage, when you exit the installation before completing the entire installation process, or because of incorrect information entered.

If a service fails or if the installation process fails during a service creation, you can resume the installation process with the server installer. To resume the installation process, ensure that at least one of the services is created and that the domain is up and running from the installation log. For example, if you want to check whether the Model Repository Service is created, check if you have a service creation success text in the server log in the following format:

```
SUCCESS: MRS Service [mrs_name] is created. Command ran successfully.
```

To resume the installation, run the installer again.

Note: You cannot resume the installer if you are running it to configure services after the services have been created. When you run the service configuration wizard, you cannot resume the installer for Big Data, Enterprise Data Lake, or Enterprise Data Catalog. When you join the domain, you also cannot resume the installer.

When you resume the installer while creating a service, the installer retains all the service and database specific information, such as the create service status, service name, service enabled or disabled status. You can confirm and use the previously entered values or specify new values for the service and resume the installation process.

Resuming the Installer

When the installation process stops midway, you can resume the installation from the point of failure or exit.

Before you can resume the installer, complete the following prerequisites:

- In the installation log file present in the installation directory, verify that at least the domain and one service is created. The installer log file name appears in the following syntax: Informatica_<Version>_Services_<timestamp>.log
- Ensure that you do not delete the installInst.obj object file present in the tools folder of the user installation directory.
- For silent installer, ensure that RESUME_INSTALLATION is set to true in the SilentInput.properties file.

1. Open a command prompt and navigate to the location of the installation files.
2. Run the Installer.

   On Linux, run silentInstall.sh to resume the silent installer. To resume the regular installer, run the ./install.sh command.

3. When the regular installer runs, you might get a prompt confirming whether you want to resume previous installer or not.
   - If you do not want to resume installation, enter 1 for No. Default is 1.
   - If you want to resume installation, enter 2 for Yes.

   Before you can resume the installation, the services get validated.

CHAPTER 13

Install Enterprise Data Lake

This chapter includes the following topics:

Installation Overview

You can run the installer to install Enterprise Data Lake on a node on which Enterprise Data Catalog is installed.

You can create the Enterprise Data Lake Service and the Data Preparation Service during the installation process. If you create the Enterprise Data Lake Service and the Data Preparation Service during the installation process, you must create both services on the same node.

You can also create a Model Repository Service and a Data Integration Service to associate with the Enterprise Data Lake Service during installation.

Install Enterprise Data Lake on a Node with Enterprise Data Catalog

You can run the installer to install Enterprise Data Lake on a domain node where Enterprise Data Catalog is installed.

You can create the Enterprise Data Lake Service and the Data Preparation Service during the installation process. If you want the installer to create the services, it creates both services on a node. The installer prompts for connection objects associated with the Hadoop environment.

Before you run the installer, verify that the domain is integrated with the Hadoop environment and that the Hadoop, HDFS, and Hive connections are associated with the cluster configuration. If the cluster configuration does not exist, you can use the Administrator tool to create the connections manually after you integrate the domain with the Hadoop environment.

For more information about integrating the domain with the Hadoop environment, see the *Informatica Big Data Management Integration Guide*.

Informatica recommends that you associate dedicated Model Repository Service and Data Integration Service instances with the Enterprise Data Lake Service. You can create a Model Repository Service and Data Integration Service to associate with the Enterprise Data Lake Service during installation, or you can associate existing services with the Enterprise Data Lake Service.

If you create a Model Repository Service, you must specify the details for the Model repository database used by the Model Repository Service.

Install the Enterprise Data Lake Binaries

When you install Enterprise Data Lake on a domain node on which Enterprise Data Catalog is already installed, you indicate that domain services and Enterprise Data Catalog are already installed on the node.

If the installer stops or is interrupted, the installer prompts you to continue installing Enterprise Data Lake at the point at which it stopped.

1. On a shell command line, run the `install.sh` file from the root directory.
2. Press **1** to install the Informatica Big Data products.
3. Press **3** to run the installer.
4. Press **2** to agree to the terms and conditions.
5. Press **2** to continue with the installation.
6. Press **3** to install Enterprise Data Lake.
7. Press **2** to indicate that the Informatica services are installed on the node.
8. Press **2** to indicate that Enterprise Data Catalog is installed on the node.
9. Press **2** to tune the application services for better performance based on your deployment type.
10. Enter the directory where you want to install Enterprise Data Lake.

    The first time you install Enterprise Data Lake, enter the Enterprise Data Catalog installation directory.
11. Choose how to proceed if Enterprise Data Lake is already installed in the specified directory.
    - Press **1** to change the installation directory.
    - Press **2** to overwrite the existing installation.
12. Review the pre-installation summary, then click **Enter**.
13. Ensure the current node is shut down, then click **Enter**.

The Domain Details section appears.

Configure the Domain Details

Configure the domain details.

1. Press **2** if the current node is the master gateway node for the domain.
2. Enter the domain name.
3. Enter the name of the current node.
4. Enter the domain administrator user name and password.
5. Press **1** to continue with the installation.

The Associated Services section appears.

Configure the Associated Services

Configure the application services required by Enterprise Data Lake.

1. Enter the name of the Catalog Service to associate with Enterprise Data Lake.
2. Enter the name of the Model Repository Service associated with the Catalog Service.
3. Enter the name of the Data Integration Service associated with the Catalog Service.
4. Choose whether to create a Model Repository Service and a Data Integration Service to associate with the Enterprise Data Lake Service, or to associate existing application services with Enterprise Data Lake.
   - To create to create a Model Repository Service and a Data Integration Service to associate with the Enterprise Data Lake Service on the node, press **1**.
   - To associate an existing Model Repository Service and an existing Data Integration Service with the Enterprise Data Lake Service, press **2**.

     After you specify the service names, skip to "Configure the Data Preparation Repository Database Details" on page 144.

The Application Services for Enterprise Data Lake section appears.

Configure the Model Repository Service and Model Repository Database Details

If you choose to create a Model Repository Service to associate with Enterprise Data Lake, specify the application service details and the connection details for the Model repository database.

1. Enter the name of the Model Repository Service to associate with Enterprise Data Lake.

   The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
2. Enter the name of the domain node where the service runs.
3. Enter the name of the Informatica license to associate with the service.
4. Specify the connection details for the Model repository database.

   The following table describes the parameters you set:

Property	Description
Database Type	Database for the Model repository managed by the Model Repository Service.
Database User ID	User name of the database user account to use to log in to the Model repository database.
User Password	Password for the Model repository database user account.
Tablespace	Configure for a IBM DB2 database. Name of the tablespace in which to create the tables. The tablespace must be defined on a single node, and the page size must be 32K. This option is required for a multi-partition database. If this option is not selected for a single partition database, the installer creates the tables in the default tablespace.
Schema Name	Configure for a Microsoft SQL Server database. Name of the schema that will contain domain configuration tables. If not selected, the installer creates the tables in the default schema.

5. Specify the truststore details required to access a secure Model repository database.

The following table describes the properties you set:

Property	Description
Database Truststore File	Path and file name of the truststore file for the secure database.
Database Truststore Password	Password for the truststore.

6. Choose whether to configure the database connection using a JDBC URL or a custom JDBC connection string.
   - Press **1** to configure the database connection using a JDBC URL.
     The following table describes the properties you set:

Property	Description
Database Address	Host name and port number for the database in the format `<host name>:<port>`.
Database Service Name	Service name for Oracle and IBM DB2 databases, or database name for Microsoft SQL Server.
JDBC Parameters	The JDBC connection string used to connect to the Model repository database. You can use the default parameters or add or modify the parameters based on your database requirements. Verify that the parameter string is valid. The installer does not validate the parameter string before it adds the string to the JDBC URL. If not selected, the installer creates the JDBC URL without additional parameters. Use the following JDBC connect string syntax for each supported database: - IBM DB2. `jdbc:informatica:db2://<host name>:<port>;DatabaseName=<database name>;BatchPerformanceWorkaround=true;DynamicSections=3000` - Microsoft SQL Server that uses the default instance. `jdbc:informatica:sqlserver://<host name>:<port>;DatabaseName=<database name>;SnapshotSerializable=true` - Microsoft SQL Server that uses a named instance. `jdbc:informatica:sqlserver://<host name>\<named instance name>;DatabaseName=<database name>;SnapshotSerializable=true` - Azure SQL Server. `jdbc:informatica:sqlserver://<host name>:<port>;DatabaseName=<database name>;SnapshotSerializable=true;SnapshotSerializable=true;EncryptionMethod=SSL;HostNameInCertificate=*.<host name in certificate>;ValidateServerCertificate=true` - Oracle. `jdbc:informatica:oracle://<host name>:<port>;SID=<database name>;MaxPooledStatements=20;CatalogOptions=0;BatchPerformanceWorkaround=true`

   - Press **2** to configure the database connection using a custom JDBC connection string.

The following table describes the properties you set:

Property	Description
EncryptionMethod	Required. Indicates whether data is encrypted when transmitted over the network. This parameter must be set to SSL.
ValidateServerCertificate	Optional. Indicates whether Informatica validates the certificate that is sent by the database server. If this parameter is set to True, Informatica validates the certificate that is sent by the database server. If you specify the HostNameInCertificate parameter, Informatica also validates the host name in the certificate. If this parameter is set to false, Informatica does not validate the certificate that is sent by the database server. Informatica ignores any truststore information that you specify.
HostNameInCertificate	Optional. Host name of the machine that hosts the secure database. If you specify a host name, Informatica validates the host name included in the connection string against the host name in the SSL certificate.

The Service Parameters section appears.

Configure the Application Service Properties

If you create a Model Repository Service and a Data Integration Service to associate with Enterprise Data Lake during installation, specify the properties required to create the Data Integration Service.

1. Specify the name of the Data Integration Service.

   The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
2. Enter the name of the node where the service runs.
3. Enter the name of the Informatica license to associate with the service.
4. Enter the name of the Model Repository Service to associate with the Data Integration Service.
5. Specify the HTTP protocol type for the Data Integration Service, and then enter the port for each protocol you select.
   - To select HTTP only, press **1**.
   - To select HTTPS only, press **2**.
   - To select both HTTP and HTTPS, press **3**.
6. If you select HTTPS or both HTTP and HTTPS, select the SSL certificate to use.
   - To use the default Informatica SSL certificate contained in the default keystore and the default truststore, press **1**.
   - To use a custom SSL certificate contained in custom keystore and truststore files, press **2**, and then enter the path and file name for the keystore and truststore files. You must also enter the keystore and truststore passwords.

The Data Preparation Repository Database section appears.

Configure the Data Preparation Repository Database Details

Specify the Data Preparation repository database connection details. You can choose to use an Oracle database or a MySQL database for the Data Preparation repository database.

If you do not have the database details, you can enter placeholder values, and then create the Data Preparation Service. If you continue without specifying the database connection details, you cannot enable the Data Preparation Service.

Oracle

1. To use an Oracle database for the Data Preparation repository, press **1**.
2. Enter the connection properties for the database.

   The following table describes the connection properties:

Property	Description
Database Host Name	Host name of the machine that hosts the database.
Database Port Number	Port number for the database.
JDBC Parameters	Parameters required to connect to the database.
Custom JDBC Connection String	JDBC connection string to connect to the database. Format the string as follows: `jdbc:informatica:oracle://<database host name>:<port>;ServiceName=<service name>`

3. To connect to a secure database, press **2**, and then enter the secure connection properties.

   The following table describes the secure connection properties:

Property	Description
Truststore File	Path and file name for the database truststore file.
Truststore Password	Password for the database truststore file.
Secure JDBC Parameters	List of secure database parameters to connect to the database. Format the parameters as follows: `EncryptionMethod=SSL;HostNameInCertificate=<secure database host name>;ValidateServerCertificate=true`

4. Press **2** to continue.

The Data Preparation Service Details section appears.

MySQL

1. To use a MySQL database or a MariaDB database for the Data Preparation repository, press **2**.
2. Enter the connection properties for the database.

The following table describes the connection properties:

Property	Description
Database Host Name	Host name of the machine that hosts the Data Preparation repository database.
Database User Name	Database user account to use to connect to the Data Preparation repository.
Database User Password	Password for the Data Preparation repository database user account.
Database Port Number	Port number for the database.
Database Name	Schema or database name of the Data Preparation repository database.

3. To connect to a secure database, press **2**, and then enter the secure connection properties.

   The following table describes the secure connection properties:

Property	Description

Custom JDBC Connection String	Connection string to connect to the database.
	To connect to a non-secure database, format the string as follows:
	`jdbc:mysql://<database host name>:<port>`
	The connection string is optional if you connect to a non-secure database.
	To connect to an SSL-enabled database, format the string as follows:
	`verifyServerCertificate=true&useSSL=true&requireSSL=true`

Secure JDBC Parameters	String containing the path and file name for the database truststore file, and the truststore password. Format the string as follows: `trustCertificateKeyStoreUrl=file://<truststore path/truststore file name>&trustCertificatekeyStorePassword=<truststore password>`

4. Press **Enter** to continue.

The Data Preparation Service Details section appears.

Create the Data Preparation Service

Create the Data Preparation Service. If you create the Enterprise Data Lake Service and the Data Preparation Service during installation, you must create both services on the same node.

1. Specify the name of the Data Preparation Service.

   The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [

2. If you plan to use rules, you must associate a Data Integration Service and a Model Repository Service with the Data Preparation Service.

   - To skip associating a Model Repository Service and a Data Integration Service with the Enterprise Data Lake Service, press **1**.
   - To associate a Model Repository Service and a Data Integration Service with the Data Preparation Service, press **2**, and then enter the service names.

3. Enter the name of the node where the Data Preparation Service runs.

- To create the service during installation, enter the name of the current node.
- If you do not want to create the service during installation, do not enter a value. You can use the Administrator tool to create the service after installation.

4. Enter the name of the Informatica license to associate with the service.
5. Choose whether to enable secure communication for the Data Preparation Service.
   - To enable secure communication for the service, press **1**.
   - To disable secure communication, press **2**.
6. If you enable secure communication for the service, select the SSL certificate to use.
   - To use the default Informatica SSL certificate contained in the default keystore and the default truststore, press **1**.
   - To use a custom SSL certificate contained in a custom keystore and truststore, press **2**, and then enter the path and file name for the keystore and truststore files. You must also enter the keystore and truststore passwords.
7. If you enable secure communication for the service, enter the port number for the HTTPS connection. If you enable non-secure communication for the service, enter the port number for the HTTP connection.
8. Select the Hadoop authentication mode.
   - To select the non-secure authentication mode, press **1**.
   - To select Kerberos authentication, press **2**.
9. If you select Kerberos, enter the authentication parameters.

   The following table describes the authentication parameters that you set if you select Kerberos:

Property	Description
HDFS Principal Name	Service Principal Name (SPN) for the data preparation Hadoop cluster. Specify the service principal name in the following format: user/_HOST@REALM.
Hadoop Impersonation User Name	User name to use in Hadoop impersonation as shown in the Impersonation User Name property for the Hadoop connection in the Administrator tool. If the Hadoop cluster uses Kerberos authentication, the Hadoop impersonation user must have read, write, and execute permissions on the HDFS storage location folder.
Kerberos Keytab File	Path and file name of the SPN keytab file for the user account to impersonate when connecting to the Hadoop cluster. The keytab file must be in a directory on the machine where the Data Preparation Service runs.
Fully Qualified Path to the Kerberos Configuration File	Path to the krb5.conf Kerberos configuration file.

10. Specify the HDFS storage location, HDFS connection, local storage location, and Solr port number details.

The following table describes the properties you set:

Property	Description
HDFS Storage Location	HDFS location for data preparation file storage. If the Hadoop cluster uses Kerberos authentication, the Hadoop impersonation user must have read, write, and execute permissions on the HDFS storage location folder.
HDFS Connection	HDFS connection for data preparation file storage.
Local Storage Location	Directory for data preparation file storage on the node on which the Data Preparation Service runs. If the connection to the local storage fails, the Data Preparation Service recovers data preparation files from the HDFS storage location.
Solr port	Solr port number for the Apache Solr server used to provide data preparation recommendations.

11. Choose whether to enable the Data Preparation Service.
    - To enable the service at a later time using the Administrator tool, press **1**.
    - To enable the service after you complete the installation process, press **2**.

The Enterprise Data Lake Service Details section appears.

Create the Enterprise Data Lake Service

Create the Enterprise Data Lake Service. If you create the Enterprise Data Lake Service and the Data Preparation Service during installation, you must create both services on the same node.

1. Specify the details for the Enterprise Data Lake Service.

   The following table describes the properties that you set:

Property	Description
Enterprise Data Lake Service Name	Name of the Enterprise Data Lake Service. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
Data Preparation Service Name	Name of the Data Preparation Service to associate with the Enterprise Data Lake Service.
Model Repository Service Name	Name of the Model Repository Service to associate with the Enterprise Data Lake Service.
Data Integration Service Name	Name of the Data Integration Service associated with the Enterprise Data Lake Service.

Property	Description
Node Name	To create the Enterprise Data Lake Service during installation, enter the name of the current node. If you do not want to create the service during installation, do not enter a value. You can use the Administrator tool to create the service after installation. If you create the Enterprise Data Lake Service and the Data Preparation Service during installation, you must create both services on the same node.
License	Enter the name of the Informatica license to associate with the service.

2. Choose whether to enable secure communication for the service.
   - To enable secure communication, press **1**.
   - To disable secure communication, press **2**.
3. If you enable secure communication for the service, select the SSL certificate to use.
   - To use the default Informatica SSL certificate contained in the default keystore and the default truststore, press **1**.
   - To use a custom SSL certificate contained in a custom keystore and truststore, press **2**, and then enter the path and file name for the keystore and truststore files. You must also enter the keystore and truststore passwords.
4. If you enable secure communication for the service, enter the port number for the HTTPS connection. If you enable non-secure communication for the service, enter the port number for the HTTP connection.
5. Specify the data lake connection options.

   The following table describes the properties that you set for the data lake connections:

Property	Description
HDFS Connection	HDFS connection for the data lake.
HDFS Working Directory	HDFS directory where the Enterprise Data Lake Service copies temporary data and files necessary for the service to run.
Hadoop Connection	Hadoop connection for the data lake.
Hive Connection	Hive connection for the data lake.
Hive Table Storage Format	Data storage format for the Hive tables.
Local System Directory	Local directory that contains the files downloaded from Enterprise Data Lake application, such as .csv and .tde files. The default directory is `/home/toolprod`.

6. Choose whether to enable logging of user activity events.
   - To disable logging of user activity events, press **1**.
   - To enable logging of user activity events, press **2**.
7. Select the Hadoop authentication mode.
   - To select the non-secure authentication mode, press **1**.
   - To select Kerberos authentication, press **2**.

8. If you select Kerberos, enter the authentication parameters.

   The following table describes the authentication properties that you must set if you select Kerberos:

Property	Description
Kerberos Principal	If the Hadoop cluster uses Kerberos authentication, specify the Service Principal Name (SPN) of the user account to impersonate when connecting to the data lake Hadoop cluster.
Kerberos KeyTab File	If the Hadoop cluster uses Kerberos authentication, specify the path and file name of the SPN keytab file for the user account to impersonate when connecting to the Hadoop cluster. The keytab file must be in a directory on the machine where the Enterprise Data Lake Service runs.

9. Choose whether to enable the Enterprise Data Lake Service immediately after you create the service.
   - To enable the service at a later time using the Administrator tool, press **1**.
   - To enable the service immediately after you create the service, press **2**.

Resume the Installer

When the installation process stops midway, you can resume the installation from the point of failure or exit.

When you run the server installer and the installation process fails, you can still resume from the previous service configuration and recover the last entered details for that service installation.

The install process might fail for reasons such as network outage, when you exit the installation before completing the entire installation process, or because of incorrect information entered.

If a service fails or if the installation process fails during a service creation, you can resume the installation process with the server installer. To resume the installation process, ensure that at least one of the services is created and that the domain is up and running from the installation log. For example, if you want to check whether the Model Repository Service is created, check if you have a service creation success text in the server log in the following format:

```
SUCCESS: MRS Service [mrs_name] is created. Command ran successfully.
```

To resume the installation, run the installer again.

Note: You cannot resume the installer if you are running it to configure services after the services have been created. When you run the service configuration wizard, you cannot resume the installer for Big Data, Enterprise Data Lake, or Enterprise Data Catalog. When you join the domain, you also cannot resume the installer.

When you resume the installer while creating a service, the installer retains all the service and database specific information, such as the create service status, service name, service enabled or disabled status. You can confirm and use the previously entered values or specify new values for the service and resume the installation process.

Resuming the Installer

When the installation process stops midway, you can resume the installation from the point of failure or exit.

Before you can resume the installer, complete the following prerequisites:

- In the installation log file present in the installation directory, verify that at least the domain and one service is created. The installer log file name appears in the following syntax: Informatica_<Version>_Services_<timestamp>.log
- Ensure that you do not delete the installInst.obj object file present in the tools folder of the user installation directory.
- For silent installer, ensure that RESUME_INSTALLATION is set to true in the SilentInput.properties file.

1. Open a command prompt and navigate to the location of the installation files.
2. Run the Installer.

   On Linux, run silentInstall.sh to resume the silent installer. To resume the regular installer, run the ./install.sh command.

3. When the regular installer runs, you might get a prompt confirming whether you want to resume previous installer or not.
   - If you do not want to resume installation, enter 1 for No. Default is 1.
   - If you want to resume installation, enter 2 for Yes.

   Before you can resume the installation, the services get validated.

CHAPTER 14

Run the Silent Installer

This chapter includes the following topics:

Installing the Informatica Services in Silent Mode

To install the Informatica services without user interaction, install in silent mode. Use a properties file to specify the installation options. The installer reads the file to determine the installation options. You can use silent mode installation to install the Informatica services on multiple machines on the network or to standardize the installation across machines.

Copy the Informatica installation files to the hard disk on the machine where you plan to install the Informatica. If you install on a remote machine, verify that you can access and create files on the remote machine.

To install in silent mode, complete the following tasks:

1. Configure the installation properties file and specify the installation options in the properties file.
2. Run the installer with the installation properties file.
3. Secure the passwords in the installation properties file.

Configure the Properties File

Configure the properties file that contains the configuration properties required to install the Informatica services in silent mode.

Informatica provides two versions of the properties file. Use either file to specify the options for your installation.

Silent input properties file

The silent input properties file contains the configuration properties required to install the Informatica services in silent mode. Use the file if you want to consider the appropriate value to set for each property in the file.

Default silent input properties file

The default silent input properties file contains default values for many configuration properties. The properties are listed in the bottom portion of the file. Use the file if you plan to install the Informatica services using the default property values.

The file contains properties set to the default value for the following options:

- Application service names.
- Secure Sockets Layer authentication.
- Kerberos authentication.
- Port number assignment for domain and node components.

To configure the properties file that contains the configuration properties required to install the Informatica services in silent mode, complete the following steps:

1. Go to the root of the directory that contains the installation files.
2. Create a backup copy of the `SilentInput.properties` file.
3. Open either the `SilentInput.properties` file or the `SilentInput_Default.properties` file.
4. Configure the properties in the file.
5. Save the file with the name `SilentInput.properties`.

Run the Installer

After you configure the properties file, open a command prompt to start the silent installation.

1. Open a command prompt.
2. Go to the root of the directory that contains the installation files.
3. Verify that the directory contains the file SilentInput.properties that you edited and resaved.
4. Run the silent installation. On Linux, run silentInstall.sh.

The silent installer runs in the background. The process can take a while. The silent installation is complete when the Informatica_<Version>_Services_InstallLog<timestamp>.log file is created in the installation directory.

The silent installation fails if you incorrectly configure the properties file or if the installation directory is not accessible. View the installation log files and correct the errors. Then run the silent installation again.

Resume the Installer

When the installation process stops midway, you can resume the installation from the point of failure or exit.

When you run the server installer and the installation process fails, you can still resume from the previous service configuration and recover the last entered details for that service installation.

The install process might fail for reasons such as network outage, when you exit the installation before completing the entire installation process, or because of incorrect information entered.

If a service fails or if the installation process fails during a service creation, you can resume the installation process with the server installer. To resume the installation process, ensure that at least one of the services is created and that the domain is up and running from the installation log. For example, if you want to check whether the Model Repository Service is created, check if you have a service creation success text in the server log in the following format:

```
SUCCESS: MRS Service [mrs_name] is created. Command ran successfully.
```

To resume the installation, run the installer again.

Note: You cannot resume the installer if you are running it to configure services after the services have been created. When you run the service configuration wizard, you cannot resume the installer for Big Data, Enterprise Data Lake, or Enterprise Data Catalog. When you join the domain, you also cannot resume the installer.

When you resume the installer while creating a service, the installer retains all the service and database specific information, such as the create service status, service name, service enabled or disabled status. You can confirm and use the previously entered values or specify new values for the service and resume the installation process.

Resuming the Installer

When the installation process stops midway, you can resume the installation from the point of failure or exit.

Before you can resume the installer, complete the following prerequisites:

- In the installation log file present in the installation directory, verify that at least the domain and one service is created. The installer log file name appears in the following syntax: Informatica_<Version>_Services_<timestamp>.log
- Ensure that you do not delete the installInst.obj object file present in the tools folder of the user installation directory.
- For silent installer, ensure that RESUME_INSTALLATION is set to true in the SilentInput.properties file.

1. Open a command prompt and navigate to the location of the installation files.
2. Run the Installer.

   On Linux, run silentInstall.sh to resume the silent installer. To resume the regular installer, run the ./install.sh command.
3. When the regular installer runs, you might get a prompt confirming whether you want to resume previous installer or not.
   - If you do not want to resume installation, enter 1 for No. Default is 1.
   - If you want to resume installation, enter 2 for Yes.

   Before you can resume the installation, the services get validated.

Secure the Passwords in the Properties File

After you run the silent installer, ensure that passwords in the properties file are kept secure.

When you configure the properties file for a silent installation, you enter passwords in plain text. After you run the silent installer, use one of the following methods to secure the passwords:

- Remove the passwords from the properties file.
- Delete the properties file.
- Store the properties file in a secure location.

CHAPTER 15

Troubleshooting

This chapter includes the following topics:

Installation Troubleshooting Overview

The topics in this section provides you information on troubleshooting probable issues that you might encounter during Informatica installation process. The examples included in the topics describe general troubleshooting strategies and are not a comprehensive list of possible causes of installation issues.

Troubleshooting with Installation Log Files

You can use the following log files to troubleshoot an Informatica installation:

Installation log files

The installer produces log files during and after the installation. You can use these logs to get more information about the tasks completed by the installer and errors that occurred during installation. The installation log files include the following logs:

- Debug logs
- File installation logs

Service Manager log files

Log files generated when the Service Manager starts on a node.

Debug Log Files

The installer writes actions and errors to the debug log file. The name of the log file depends on the Informatica component you install.

The following table describes the properties of the debug log files:

Property	Description
Log File Name	- Informatica_<Version>_Services.log - Informatica_<Version>_Client.log - Informatica_<Version>_Services_Upgrade.log - Informatica_<Version>_Client_Upgrade.log
Location	Installation directory.
Usage	Get more information about the actions performed by the installer and get more information about installation errors. The installer writes information to this file during the installation. If the installer generates an error, you can use this log to troubleshoot the error.
Contents	Detailed summary of each action performed by the installer, the information you entered in the installer, each command line command used by the installer, and the error code returned by the command.

The debug log contains output from the infacmd and infasetup commands used to create the domain, node, and application services. It also contains information about starting the application services.

File Installation Log File

The file installation log file contains information about the installed files.

The following table describes the properties of the installation log file:

Property	Description
Log File Name	- Informatica_<Version>_Services_InstallLog.log - Informatica_<Version>_Client_InstallLog.log
Location	Installation directory.
Usage	Get information about the files installed and registry entries created.
Contents	Directories created, names of the files installed and commands run, and status for each installed file.

Service Manager Log Files

The installer starts the Informatica service. The Informatica service starts the Service Manager for the node. The Service Manager generates log files that indicate the startup status of a node. Use these files to troubleshoot issues when the Informatica service fails to start and you cannot log in to Informatica Administrator. The Service Manager log files are created on each node.

The following table describes the files generated by the Service Manager:

Property	Description
catalina.out	Log events from the Java Virtual Machine (JVM) that runs the Service Manager. For example, a port is available during installation, but is in use when the Service Manager starts. Use this log to get more information about which port was unavailable during startup of the Service Manager. The catalina.out file is in the following directory: `<Informatica installation directory>/logs/<node name>/catalina.out`
node.log	Log events generated during the startup of the Service Manager on a node. You can use this log to get more information about why the Service Manager for a node failed to start. For example, if the Service Manager cannot connect to the domain configuration database after 30 seconds, the Service Manager fails to start. The node.log file is in the /tomcat/logs directory.

Note: The Service Manager also uses node.log to record events when the Log Manager is unavailable. For example, if the machine where the Service Manager runs does not have enough available disk space to write log event files, the Log Manager is unavailable.

Troubleshooting Domains and Nodes

The installer can generate errors when creating and configuring domains and nodes during the Informatica installation.

You can encounter errors with the following installer tasks:

- Adding the domain configuration database
- Creating or joining a domain
- Starting Informatica
- Pinging the domain
- Adding a license

Creating the Domain Configuration Repository

If you create a domain, the installer creates a domain configuration repository to store domain metadata. The installer uses the options you enter during installation to add configuration metadata to the domain configuration repository. The installer uses JDBC to communicate with the database. You do not need to configure ODBC or native connectivity on the machine where you install the Informatica services.

The installer creates and drops a table in the domain configuration repository database to verify the connection information. The user account for the database must have create privileges on the database. Each domain must have a separate domain configuration repository.

Creating or Joining a Domain

The installer completes different tasks depending on whether you create a domain or join a domain:

- **Creating a domain.** The installer runs the infasetup DefineDomain command to create the domain and the gateway node for the domain on the current machine based on the information you enter in the Configure Domain window.
- **Joining a domain.** The installer runs the infasetup DefineWorkerNode command to create a node on the current machine, and runs the infacmd AddDomainNode command to add the node to the domain. The installer uses the information you enter in the Configure Domain window to run the commands.

The infasetup and infacmd commands fail if the gateway node is unavailable. If the gateway node is unavailable, you cannot log in to Informatica Administrator.

For example, the DefineDomain command fails if you click Test Connection and the connection test passes but the database becomes unavailable before you click Next. The DefineDomain command can also fail if the host name or IP address does not belong to the current machine. Verify that the database for the domain configuration is available and that the host name is correct and try again.

If the AddDomainNode command fails, verify that the Informatica service is running on the gateway node and try again.

Starting Informatica

The installer runs infaservice to start the Informatica service. To troubleshoot issues when Informatica fails to start, use the information in the installation debug log and the node.log and catalina.out Service Manager log files to identify the cause of the error.

If you create a domain, log in to Informatica Administrator after the Informatica service starts to verify that the domain is available. If you join a domain, log in to Informatica Administrator after the Informatica service starts to verify that the node was successfully created and started.

Informatica can fail to start for the following reasons:

- **The Service Manager is out of system memory.** The Java Runtime Environment (JRE) that starts Informatica and runs the Service Manager may not have enough system memory to start. Set the INFA_JAVA_OPTS environment variable to configure the amount of system memory used by Informatica. On UNIX, you can set the memory configuration when you start Informatica.
- **The domain configuration database is not available.** Informatica fails to start on a node if the Service Manager on a gateway node cannot connect to the domain configuration database within 30 seconds. Verify that the domain configuration repository is available.
- **Some of the folders in the Informatica installation directory do not have the appropriate execute permissions.** Grant execute permission on the Informatica installation directory.

Pinging the Domain

The installer runs the *infacmd* Ping command to verify that the domain is available before it continues the installation. The domain must be available so that license objects can be added to the domain. If the Ping command fails, start Informatica on the gateway node.

Adding a License

The installer runs the *infacmd* AddLicense command to read the Informatica license key file and create a license object in the domain. To run the application services in Informatica Administrator, a valid license object must exist in the domain.

If you use an incremental license and join a domain, the serial number of the incremental license must match the serial number for an existing license object in the domain. If the serial numbers do not match, the AddLicense command fails.

You can get more information about the contents of the license key file used for installation, including serial number, version, expiration date, operating systems, and connectivity options in the installation debug log. You can get more information about existing licenses for the domain in Informatica Administrator.

Part IV: After You Install the Services

This part contains the following chapters:

CHAPTER 16

Complete the Domain Configuration

This chapter includes the following topics:

Checklist to Complete the Domain Configuration

This chapter contains information about domain configuration tasks that you need to complete after installation. Use this checklist to track domain configuration tasks.

❑ Verify locale settings and code page compatibility:

- Verify that the domain configuration database is compatible with the code pages of the application services that you create in the domain.
- Verify that the locale settings on machines that access the Administrator tool and the Informatica client tools is compatible with the code pages of repositories in the domain.
- Configure the locale environment variables.

❑ Configure the following environment variables:

- Informatica environment variables to store memory, domain, and location settings.
- Library path environment variables on the machines that run the Data Integration Service.
- Kerberos environment variables if you configure the Informatica domain to run on a network with Kerberos authentication.

Complete the Domain Configuration Overview

After you install Informatica services and before you create the application services, complete the configuration for the domain services.

Domain configuration includes tasks such as verifying code pages, configuring the environment variables for the domain, and configuring the firewall.

Verify Locale Settings and Code Page Compatibility

The code pages for application services must be compatible with code pages in the domain.

Verify and configure the locale settings and code pages:

Verify that the domain configuration database is compatible with the code pages of the application services that you create in the domain.

The Service Manager synchronizes the list of users in the domain with the list of users and group in each application service. If a user name in the domain has characters that the code page of the application service does not recognize, characters do not convert correctly and inconsistencies occur.

Verify that the locale settings on machines that access the Administrator tool and the Informatica client tools are compatible with code pages of repositories in the domain.

If the locale setting is not compatible with the repository code page, you cannot create an application service.

Configure Locale Environment Variables

Verify that the locale setting is compatible with the code page for the repository. If the locale setting is not compatible with the repository code page, you cannot create an application service.

Use LANG, LC_CTYPE, or LC_ALL to set the UNIX code page.

Different UNIX operating systems require different values for the same locale. The value for the locale variable is case sensitive.

Use the following command to verify that the value for the locale environment variable is compatible with the language settings for the machine and the type of code page you want to use for the repository:

```
locale -a
```

The command returns the languages installed on the UNIX operating system and the existing locale settings.

Set the following locale environment variables:

Locale on Linux

All UNIX operating systems except Linux have a unique value for each locale. Linux allows different values to represent the same locale. For example, “utf8,” “UTF-8,” “UTF8,” and “utf-8” represent the same locale on a Linux machine. Informatica requires that you use a specific value for each locale on a Linux machine. Make sure that you set the LANG environment variable appropriately for all Linux machines.

Locale for Oracle database clients

For Oracle database clients, set NLS_LANG to the locale that you want the database client and server to use with the login. A locale setting consists of the language, territory, and character set. The value of NLS_LANG depends on the configuration.

For example, if the value is american_america.UTF8, set the variable in a C shell with the following command:

```
setenv NLS_LANG american_america.UTF8
```

To read multibyte characters from the database, set the variable with the following command:

```
setenv NLS_LANG=american_america.AL32UTF8
```

You must set the correct variable on the Data Integration Service machine so that the Data Integration Service can read the Oracle data correctly.

Configure Environment Variables

Informatica uses environment variables to store configuration information when it runs the application services and connects to the clients. Configure the environment variables to meet the Informatica requirements.

Incorrectly configured environment variables can cause the Informatica domain or nodes to fail to start or can cause connection problems between the Informatica clients and the domain.

To configure environment variables, log in with the system user account you used to install Informatica.

Configure Informatica Environment Variables

You can configure Informatica environment variables to store memory, domain, and location settings.

Set the following environment variables:

INFA_JAVA_OPTS

By default, Informatica uses a maximum of 512 MB of system memory.

The following table lists the minimum requirement for the maximum heap size settings, based on the number of users and services in the domain:

Number of Domain Users	Maximum Heap Size (1-5 Services)	Maximum Heap Size (6-10 Services)
1,000 or less	512 MB (default)	1024 MB
5,000	2048 MB	3072 MB
10,000	3072 MB	5120 MB
20,000	5120 MB	6144 MB
30,000	5120 MB	6144 MB

Note: The maximum heap size settings in the table are based on the number of application services in the domain.

If the domain has more than 1,000 users, update the maximum heap size based on the number of users in the domain.

You can use the INFA_JAVA_OPTS environment variable to configure the amount of system memory used by Informatica. For example, to configure 1 GB of system memory for the Informatica daemon in a C shell, use the following command:

```
setenv INFA_JAVA_OPTS "-Xmx1024m"
```

Restart the node for the changes to take effect.

INFA_DOMAINS_FILE

The installer creates a domains.infa file in the Informatica installation directory. The domains.infa file contains the connectivity information for the gateway nodes in a domain, including the domain names, domain host names, and domain host port numbers.

Set the value of the INFA_DOMAINS_FILE variable to the path and file name of the domains.infa file.

Configure the INFA_DOMAINS_FILE variable on the machine where you install the Informatica services.

INFA_HOME

Use INFA_HOME to designate the Informatica installation directory. If you modify the Informatica directory structure, you need to set the environment variable to the location of the Informatica installation directory or the directory where the installed Informatica files are located.

For example, you use a softlink for any of the Informatica directories. To configure INFA_HOME so that any Informatica application or service can locate the other Informatica components it needs to run, set INFA_HOME to the location of the Informatica installation directory.

INFA_TRUSTSTORE

If you enable secure communication for the domain, set the INFA_TRUSTSTORE variable with the directory that contains the truststore files for the SSL certificates. The directory must contain truststore files named infa_truststore.jks and infa_truststore.pem.

You must set the INFA_TRUSTSTORE variable if you use the default SSL certificate provided by Informatica or a certificate that you provide.

INFA_TRUSTSTORE_PASSWORD

If you enable secure communication for the domain and you specify the SSL certificate to use, set the INFA_TRUSTSTORE_PASSWORD variable with the password for the infa_truststore.jks that contains the SSL certificate. The password must be encrypted. Use the command line program pmpasswd to encrypt the password.

Configure Library Path Environment Variables

Configure library path environment variables on the machines that run the Data Integration Service processes. The variable name and requirements depend on the platform and database.

Configure the LD_LIBRARY_PATH environment variable.

The following table describes the values that you set for the LD_LIBRARY_PATH for the different databases:

Database	Value
Oracle	<Database path>/lib
IBM DB2	<Database path>/lib
Sybase ASE	"${SYBASE_OCS}/lib:${SYBASE_ASE}/lib:${LD_LIBRARY_PATH}"

Database	Value
Informix	<Database path>/lib
Teradata	<Database path>/lib
ODBC	<CLOSEDODBCHOME>/lib

Configure Kerberos Environment Variables

If you configure the Informatica domain to run on a network with Kerberos authentication, you must set the Kerberos configuration and credential cache environment variables.

Set the following environment variables:

KRB5_CONFIG

Use the KRB5_CONFIG environment variable to store the path and file name of the Kerberos configuration file. The name of the Kerberos configuration file is *krb5.conf*. You must set the KRB5_CONFIG environment variable on each node in the Informatica domain.

KRB5CCNAME

Set the KRB5CCNAME environment variable with the path and file name of the Kerberos user credential cache. Kerberos single sign-on requires Kerberos credential cache for user accounts.

When you cache the user credential, you must use the *forwardable* option. For example, if you use *kinit* to get and cache the user credential, you must use the *-f* option to request forwardable tickets.

CHAPTER 17

Prepare to Create the Application Services

This chapter includes the following topics:

Checklist for Preparing to Create Application Services

This chapter contains tasks that you need to complete before you create or configure the Analyst Service, the Data Integration Service, and the Content Management Service. When you configure the services you configure properties based on the connections and directories that you create. Use this checklist to track the configuration tasks.

❑ Create the following directories for the Analyst Service:
- Flat file caches
- Temporary business glossary files
- Glossary assets

❑ Create the following connections for the Data Integration Service:
- Data object cache database
- Workflow database
- Profiling warehouse

❑ Create the following connection for the Content Management Service:
- Reference data warehouse

Create Directories for the Analyst Service

Before you create the Analyst Service, you must create directories for the Analyst tool to store temporary files.

Create the following directories on the node that runs the Analyst Service:

Flat file cache directory

Create a directory for the flat file cache where the Analyst tool stores uploaded flat files. The Data Integration Service must also be able to access this directory. If the Analyst Service and the Data Integration Service run on different nodes, configure the flat file directory to use a shared directory. If the Data Integration Service runs on primary and back-up nodes or on a grid, each Data Integration Service process must be able to access the files in the shared directory.

For example, you can create a directory named "flatfilecache" in the following mapped drive that all Analyst Service and Data Integration Service processes can access:

```
F:\shared\<InformaticaInstallationDir>\server
```

When you import a reference table or flat file source, the Analyst tool uses the files from this directory to create a reference table or flat file data object.

Temporary export file directory

Create a directory to store the temporary business glossary files that the business glossary export process creates. Create the directory on the node that runs the Analyst Service.

For example, you can create a directory named "exportfiledirectory" in the following location:
`<Informatica installation directory>/server`

Asset attachments directory

Create a directory to store the files that content managers add as attachments to Glossary assets. Create the directory on the node that runs the Analyst Service.

For example, you can create a directory named "attachmentdirectory" in the following location:
`<Informatica installation directory>/server`

Log In to Informatica Administrator

You must have a user account to log in to the Informatica Administrator web application.

If the Informatica domain runs on a network with Kerberos authentication, you must configure the browser to allow access to the Informatica web applications. In Microsoft Internet Explorer and Google Chrome, add the URL of the Informatica web application to the list of trusted sites. If you are using Chrome version 41 or later, you must also set the `AuthServerWhitelist` and `AuthNegotiateDelegateWhitelist` policies.

1. Start a Microsoft Internet Explorer or Google Chrome browser.
2. In the **Address** field, enter the URL for the Administrator tool:
   - If the Administrator tool is not configured to use a secure connection, enter the following URL:
     ```
     http://<fully qualified hostname>:<http port>/administrator/
     ```
   - If the Administrator tool is configured to use a secure connection, enter the following URL:
     ```
     https://<fully qualified hostname>:<http port>/administrator/
     ```

Host name and port in the URL represent the host name and port number of the master gateway node. If you configured secure communication for the domain, you must use HTTPS in the URL to ensure that you can access the Administrator tool.

If you use Kerberos authentication, the network uses single sign on. You do not need to log in to the Administrator tool with a user name and password.

3. If you do not use Kerberos authentication, enter the user name, password, and security domain for your user account, and then click **Login**.

   The **Security Domain** field appears when the Informatica domain contains an LDAP security domain. If you do not know the security domain that your user account belongs to, contact the Informatica domain administrator.

   Note: If this is the first time you log in with the user name and password provided by the domain administrator, change your password to maintain security.

Troubleshooting the Login to Informatica Administrator

If the Informatica domain uses Kerberos authentication, you might encounter the following issues when logging in to the Administrator tool:

I cannot log in to the Administrator tool from the same machine where I created the domain gateway node.

After installation, if you cannot log in to the Administrator tool from the same machine where you created the domain gateway node, clear the browser cache. When you initially log in to the Administrator tool after installation, you can only log in with the Administrator user account created during installation. If a different user credential is stored in the browser cache, the login can fail.

A blank page appears after I log in to the Administrator tool.

If a blank page appears after you log in to the Administrator tool, verify that you enabled delegation for all user accounts with service principals used in the Informatica domain. To enable delegation, in the Microsoft Active Directory Service, set the **Trust this user for delegation to any service (Kerberos only)** option for each user account that you set an SPN.

Create Connections

In the Administrator tool, create connections to the databases that the application services use. You need to specify the connection details while you configure the application service.

When you create the database connection, specify the database connection properties and test the connection.

The following table describes the database connections that you must create before you create the associated application services:

Database Connection	Description
Data object cache database	To access the data object cache, create the data object cache connection for the Data Integration Service.
Workflow database	To store run-time metadata for workflows, create the workflow database connection for the Data Integration Service.

Database Connection	Description
Profiling warehouse database	To create and run profiles and scorecards, create the profiling warehouse database connection for the Data Integration Service. Use this instance of the Data Integration Service when you configure the run-time properties of the Analyst Service. **Note:** To use the Microsoft SQL Server database as the profiling warehouse, choose ODBC as the provider type, and clear the **use DSN** option in the **Microsoft SQL Server connection properties** dialog box when you configure the Microsoft SQL Server connection.
Reference data warehouse	To store reference table data, create the reference data warehouse connection for the Content Management Service.

IBM DB2 Connection Properties

Use a DB2 for LUW connection to access tables in a DB2 for LUW database.

The following table describes the DB2 for LUW connection properties:

Property	Description
User name	Database user name.
Password	Password for the user name.
Connection String for metadata access	Connection string to import physical data objects. Use the following connection string: `jdbc:informatica:db2://<host>:50000;databaseName=<dbname>`
Connection String for data access	Connection string to preview data and run mappings. Enter `dbname` from the alias configured in the DB2 client.
Code Page	Database code page.
Environment SQL	Optional. Enter SQL commands to set the database environment when you connect to the database. The Data Integration Service executes the connection environment SQL each time it connects to the database.
Transaction SQL	Optional. Enter SQL commands to set the database environment when you connect to the database. The Data Integration Service executes the transaction environment SQL at the beginning of each transaction.
Retry Period	This property is reserved for future use.
Tablespace	Tablespace name of the DB2 for LUW database.

Property	Description
SQL Identifier Character	The type of character used to identify special characters and reserved SQL keywords, such as WHERE. The Data Integration Service places the selected character around special characters and reserved SQL keywords. The Data Integration Service also uses this character for the Support Mixed-case Identifiers property.
Support Mixed-case Identifiers	When enabled, the Data Integration Service places identifier characters around table, view, schema, synonym, and column names when generating and executing SQL against these objects in the connection. Use if the objects have mixed-case or lowercase names. By default, this option is not selected.

Microsoft SQL Server Connection Properties

Use a Microsoft SQL Server connection to access tables in a Microsoft SQL Server database.

The following table describes the Microsoft SQL Server connection properties:

Property	Description
User name	Database user name.
Password	Password for the user name.
Use Trusted Connection	Optional. When enabled, the Data Integration Service uses Windows authentication to access the Microsoft SQL Server database. The user name that starts the Data Integration Service must be a valid Windows user with access to the Microsoft SQL Server database.
Connection String for metadata access	Connection string to import physical data objects. Use the following connection string: `jdbc:informatica:sqlserver://<host>:<port>;databaseName=<dbname>`
Connection String for data access	Connection string to preview data and run mappings. Enter `<ServerName>@<DBName>`
Domain Name	Optional. Name of the domain where Microsoft SQL Server is running.
Packet Size	Required. Optimize the ODBC connection to Microsoft SQL Server. Increase the packet size to increase performance. Default is 0.
Code Page	Database code page.
Owner Name	Name of the schema owner. Specify for connections to the profiling warehouse database or data object cache database.
Schema Name	Name of the schema in the database. Specify for connections to the profiling warehouse or data object cache database. You must specify the schema name for the profiling warehouse if the schema name is different from the database user name. You must specify the schema name for the data object cache database if the schema name is different from the database user name and you manage the cache with an external tool.

Property	Description
Environment SQL	Optional. Enter SQL commands to set the database environment when you connect to the database. The Data Integration Service executes the connection environment SQL each time it connects to the database.
Transaction SQL	Optional. Enter SQL commands to set the database environment when you connect to the database. The Data Integration Service executes the transaction environment SQL at the beginning of each transaction.
Retry Period	This property is reserved for future use.
SQL Identifier Character	The type of character used to identify special characters and reserved SQL keywords, such as WHERE. The Data Integration Service places the selected character around special characters and reserved SQL keywords. The Data Integration Service also uses this character for the Support Mixed-case Identifiers property.
Support Mixed-case Identifiers	When enabled, the Data Integration Service places identifier characters around table, view, schema, synonym, and column names when generating and executing SQL against these objects in the connection. Use if the objects have mixed-case or lowercase names. By default, this option is not selected.

Note: When you use a Microsoft SQL Server connection to access tables in a Microsoft SQL Server database, the Developer tool does not display the synonyms for the tables.

Oracle Connection Properties

Use an Oracle connection to access tables in an Oracle database.

The following table describes the Oracle connection properties:

Property	Description
User name	Database user name.
Password	Password for the user name.
Connection String for metadata access	Connection string to import physical data objects. Use the following connection string: `jdbc:informatica:oracle://<host>:1521;SID=<sid>`
Connection String for data access	Connection string to preview data and run mappings. Enter `dbname.world` from the TNSNAMES entry.
Code Page	Database code page.
Environment SQL	Optional. Enter SQL commands to set the database environment when you connect to the database. The Data Integration Service executes the connection environment SQL each time it connects to the database.
Transaction SQL	Optional. Enter SQL commands to set the database environment when you connect to the database. The Data Integration Service executes the transaction environment SQL at the beginning of each transaction.

Property	Description
Retry Period	This property is reserved for future use.
Parallel Mode	Optional. Enables parallel processing when loading data into a table in bulk mode. Default is disabled.
SQL Identifier Character	The type of character used to identify special characters and reserved SQL keywords, such as WHERE. The Data Integration Service places the selected character around special characters and reserved SQL keywords. The Data Integration Service also uses this character for the Support Mixed-case Identifiers property.
Support Mixed-case Identifiers	When enabled, the Data Integration Service places identifier characters around table, view, schema, synonym, and column names when generating and executing SQL against these objects in the connection. Use if the objects have mixed-case or lowercase names. By default, this option is not selected.

Creating a Connection

In the Administrator tool, you can create relational database, social media, and file systems connections.

1. In the Administrator tool, click the **Manage** tab.
2. Click the **Connections** view.
3. In the Navigator, select the domain.
4. In the Navigator, click **Actions** > **New** > **Connection**.

   The **New Connection** dialog box appears.
5. In the **New Connection** dialog box, select the connection type, and then click **OK**.

   The **New Connection** wizard appears.
6. Enter the connection properties.

   The connection properties that you enter depend on the connection type. Click **Next** to go to the next page of the **New Connection** wizard.
7. When you finish entering connection properties, you can click **Test Connection** to test the connection.
8. Click **Finish**.

CHAPTER 18

Create and Configure Application Services

This chapter includes the following topics:

Checklist to Create and Configure Application Services

This chapter contains instructions to create and configure application services. Even if you created services during installation, you might still need to configure some services. Use this checklist to track completion of application service configuration.

- ❑ Review your notes for planning the application services.
- ❑ Identify the services that you created during installation, and complete additional configuration for the service.
- ❑ Create and configure other services that you want in the domain.

Create and Configure the Application Services Overview

Use the Administrator tool to create the application services.

Some application services depend on other application services. When you create these dependent application services, you must provide the name of other running application services. Review the application service dependencies to determine the order that you must create the services. For example, you must create a Model Repository Service and a Data Integration Service before you create an Analyst Service.

Before you create the application services, verify that you have completed the prerequisite tasks required by the installation and configuration process.

Create and Configure the Model Repository Service

The Model Repository Service is an application service that manages the Model repository. The Model repository stores metadata created by Informatica clients and application services in a relational database to enable collaboration among the clients and services.

When you access a Model repository object from the Developer tool, the Analyst tool, the Administrator tool, or the Data Integration Service, the client or service sends a request to the Model Repository Service. The Model Repository Service process fetches, inserts, and updates the metadata in the Model repository database tables.

Create the Model Repository Service

Use the service creation wizard in the Administrator tool to create the service.

1. In the Administrator tool, click the **Manage** tab.
2. Click **Actions** > **New** > **Model Repository Service**.

   The **New Model Repository Service** dialog box appears.
3. On the **New Model Repository Service - Step 1 of 2** page, enter the following properties:

Property	Description
Name	Name of the service. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
Description	Description of the service. The description cannot exceed 765 characters.
Location	Domain and folder where the service is created. Click **Browse** to choose a different folder. You can move the service after you create it.
License	License object that allows use of the service.

Property	Description
Node	Node on which the service runs.
Backup Nodes	If your license includes high availability, nodes on which the service can run if the primary node is unavailable.

4. Click **Next**.

   The **New Model Repository Service - Step 2 of 2** page appears.

5. Enter the following properties for the Model repository database:

Property	Description
Database Type	The type of the repository database.
Username	The database user name for the repository.
Password	Repository database password for the database user.
Database Schema	Available for Microsoft SQL Server. Name of the schema that will contain Model repository tables.
Database Tablespace	Available for IBM DB2. Name of the tablespace in which to create the tables. For a multi-partition IBM DB2 database, the tablespace must span a single node and a single partition.

6. Enter the JDBC connection string that the service uses to connect to the Model repository database.

   Use the following syntax for the connection string for the selected database type:

Database Type	Connection String Syntax
IBM DB2	`jdbc:informatica:db2://<host_name>:<port_number>;DatabaseName=<database_name>;BatchPerformanceWorkaround=true;DynamicSections=3000`
Microsoft SQL Server	- Microsoft SQL Server that uses the default instance `jdbc:informatica:sqlserver://<host_name>:<port_number>;DatabaseName=<database_name>;SnapshotSerializable=true` - Microsoft SQL Server that uses a named instance `jdbc:informatica:sqlserver://<host_name>\<named_instance_name>;DatabaseName=<database_name>;SnapshotSerializable=true` - Azure SQL Server. `jdbc:informatica:sqlserver://<host_name>:<port_number>;DatabaseName=<database_name>;SnapshotSerializable=true;SnapshotSerializable=true;EncryptionMethod=SSL;HostNameInCertificate=*.<hostnameincertificate>;ValidateServerCertificate=true`
Oracle	`jdbc:informatica:oracle://<host_name>:<port_number>;SID=<database_name>;MaxPooledStatements=20;CatalogOptions=0;BatchPerformanceWorkaround=true`

7. If the Model repository database is secured with the SSL protocol, you must enter the secure database parameters in the **Secure JDBC Parameters** field.

   Enter the parameters as `name=value` pairs separated by semicolon characters (;). For example:

   ```
   param1=value1;param2=value2
   ```

   Enter the following secure database parameters:

Secure Database Parameter	Description
EncryptionMethod	Required. Indicates whether data is encrypted when transmitted over the network. This parameter must be set to `SSL`.
ValidateServerCertificate	Optional. Indicates whether Informatica validates the certificate that the database server sends. If this parameter is set to True, Informatica validates the certificate that the database server sends. If you specify the HostNameInCertificate parameter, Informatica also validates the host name in the certificate. If this parameter is set to False, Informatica does not validate the certificate that the database server sends. Informatica ignores any truststore information that you specify.
HostNameInCertificate	Optional. Host name of the machine that hosts the secure database. If you specify a host name, Informatica validates the host name included in the connection string against the host name in the SSL certificate.
cryptoProtocolVersion	Required. Specifies the cryptographic protocol to use to connect to a secure database. You can set the parameter to `cryptoProtocolVersion=TLSv1.1` or `cryptoProtocolVersion=TLSv1.2` based on the cryptographic protocol used by the database server.
TrustStore	Required. Path and file name of the truststore file that contains the SSL certificate for the database. If you do not include the path for the truststore file, Informatica looks for the file in the following default directory: `<Informatica installation directory>/tomcat/bin`
TrustStorePassword	Required. Password for the truststore file for the secure database.

   Note: Informatica appends the secure JDBC parameters to the JDBC connection string. If you include the secure JDBC parameters directly in the connection string, do not enter any parameter in the **Secure JDBC Parameters** field.

8. Click **Test Connection** to verify that you can connect to the database.
9. Select **No content exists under specified connection string. Create new content.**
10. Click **Finish**.

    The domain creates the Model Repository Service, creates content for the Model repository in the specified database, and enables the service.

    Note: When you update the Model Repository Service properties, you must restart the Model Repository Service for the modifications to take effect.

After you create the service through the wizard, you can edit the properties or configure other properties.

Create the Model Repository User

If the domain does not use Kerberos authentication, the domain uses a user account to authenticate other application services that make requests to the Model Repository Service. You must create a user account and assign the user the Administrator role for the Model Repository Service.

When you create an application service that depends on the Model Repository Service, you provide the name of the Model Repository Service and of this Model repository user.

1. In the Administrator tool, click the **Security** tab.
2. On the Security Actions menu, click **Create User** to create a native user account.

   Note: If you set up LDAP authentication in the domain, you can use an LDAP user account for the Model repository user.
3. Enter the following properties for the user:

Property	Description
Login Name	Login name for the user account. The login name for a user account must be unique within the security domain to which it belongs. The name is not case sensitive and cannot exceed 128 characters. It cannot include a tab, newline character, or the following special characters: , + " \ < > ; / * % ? & The name can include an ASCII space character except for the first and last character. All other space characters are not allowed.
Password	Password for the user account. The password can be from 1 through 80 characters long.
Confirm Password	Enter the password again to confirm. You must retype the password. Do not copy and paste the password.
Full Name	Full name for the user account. The full name cannot include the following special characters: < > “
Description	Description of the user account. The description cannot exceed 765 characters or include the following special characters: < > “

4. Click **OK**.

   The user properties appear.
5. Click the **Privileges** tab.
6. Click **Edit**.

   The **Edit Roles and Privileges** dialog box appears.
7. On the **Roles** tab, expand the Model Repository Service.
8. Under **System Defined Roles**, select Administrator and click **OK**.

Create and Configure the Data Integration Service

The Data Integration Service is an application service that performs data integration jobs for the Analyst tool, the Developer tool, and external clients.

When you preview or run data profiles, SQL data services, and mappings in the Analyst tool or the Developer tool, the client tool sends requests to the Data Integration Service to perform the data integration jobs. When you run SQL data services, mappings, and workflows from the command line program or an external client, the command sends the request to the Data Integration Service.

Create the Data Integration Service

Use the service creation wizard in the Administrator tool to create the service.

Before you create the Data Integration Service, verify that you have created and enabled the Model Repository Service. If the domain does not use Kerberos authentication, verify that you have created a Model repository user that the Data Integration Service can use to access the Model Repository Service.

1. In the Administrator tool, click the **Manage** tab.
2. Click the **Services and Nodes** view.
3. In the Domain Navigator, select the domain.
4. Click **Actions** > **New** > **Data Integration Service**.

   The **New Data Integration Service** wizard appears.
5. On the **New Data Integration Service - Step 1 of 14** page, enter the following properties:

Property	Description
Name	Name of the service. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
Description	Description of the service. The description cannot exceed 765 characters.
Location	Domain and folder where the service is created. Click **Browse** to choose a different folder. You can move the service after you create it.
License	License object that allows use of the service.
Assign	Select **Node** to configure the service to run on a node. If your license includes grid, you can create a grid and assign the service to run on the grid after you create the service.
Node	Node on which the service runs.
Backup Nodes	If your license includes high availability, nodes on which the service can run if the primary node is unavailable.
Model Repository Service	Model Repository Service to associate with the service.

Property	Description
Username	User name that the service uses to access the Model Repository Service. Enter the Model repository user that you created.
Password	Password for the Model repository user.
Security Domain	LDAP security domain for the Model repository user. The field appears when the Informatica domain contains an LDAP security domain. Not available for a domain with Kerberos authentication.

6. Click **Next**.

   The **New Data Integration Service - Step 2 of 14** page appears.
7. Enter the HTTP port number to use for the Data Integration Service.
8. Accept the default values for the remaining security properties. You can configure the security properties after you create the Data Integration Service.
9. Select **Enable Service**.

   The Model Repository Service must be running to enable the Data Integration Service.
10. Verify that the **Move to plugin configuration page** is not selected.
11. Click **Next**.

    The **New Data Integration Service - Step 3 of 14** page appears.
12. Set the **Launch Job Options** property to one of the following values:
    - In the service process. Configure when you run SQL data service and web service jobs. SQL data service and web service jobs typically achieve better performance when the Data Integration Service runs jobs in the service process.
    - In separate local processes. Configure when you run mapping, profile, and workflow jobs. When the Data Integration Service runs jobs in separate local processes, stability increases because an unexpected interruption to one job does not affect all other jobs.

    If you configure the Data Integration Service to run on a grid after you create the service, you can configure the service to run jobs in separate remote processes.
13. Accept the default values for the remaining execution options and click **Next**.

    The **New Data Integration Service - Step 4 of 14** page appears.
14. If you created the data object cache database for the Data Integration Service, click **Select** to select the cache connection. Select the data object cache connection that you created for the service to access the database.
15. Accept the default values for the remaining properties on this page and click **Next**.

    The **New Data Integration Service - Step 5 of 14** page appears.
16. For optimal performance, enable the Data Integration Service modules that you plan to use.

The following table lists the Data Integration Service modules that you can enable:

Module	Description
Web Service Module	Runs web service operation mappings.
Mapping Service Module	Runs mappings and previews.
Profiling Service Module	Runs profiles and scorecards.
SQL Service Module	Runs SQL queries from a third-party client tool to an SQL data service.
Workflow Orchestration Service Module	Runs workflows.

17. Click **Next**.

    The **New Data Integration Service - Step 6 of 14** page appears.

    You can configure the HTTP proxy server properties to redirect HTTP requests to the Data Integration Service. You can configure the HTTP configuration properties to filter the web services client machines that can send requests to the Data Integration Service. You can configure these properties after you create the service.

18. Accept the default values for the HTTP proxy server and HTTP configuration properties and click **Next**.

    The **New Data Integration Service - Step 7 of 14** page appears.

    The Data Integration Service uses the result set cache properties to use cached results for SQL data service queries and web service requests. You can configure the properties after you create the service.

19. Accept the default values for the result set cache properties and click **Next**.

    The **New Data Integration Service - Step 8 of 14** page appears.

20. If you created the profiling warehouse database for the Data Integration Service, select the Profiling Service module.

21. If you created the workflow database for the Data Integration Service, select the Workflow Orchestration Service module.

22. Verify that the remaining modules are not selected.

    You can configure properties for the remaining modules after you create the service.

23. Click **Next**.

    The **New Data Integration Service - Step 11 of 14** page appears.

24. If you created the profiling warehouse database for the Data Integration Service, click **Select** to select the database connection. Select the profiling warehouse connection that you created for the service to access the database.

25. Select whether or not content exists in the profiling warehouse database.

    If you created a new profiling warehouse database, select **No content exists under specified connection string**.

26. Click **Next**.

    The **New Data Integration Service - Step 12 of 14** page appears.

27. Accept the default values for the advanced profiling properties and click **Next**.

    The **New Data Integration Service - Step 14 of 14** page appears.

28. If you created the workflow database for the Data Integration Service, click **Select** to select the database connection. Select the workflow database connection that you created for the service to access the database.
29. Click **Finish**.

    The domain creates and enables the Data Integration Service.

After you create the service through the wizard, you can edit the properties or configure other properties.

Verify the Host File Configuration on Linux

If you configured the Data Integration Service on linux to launch jobs as separate processes, verify that the host file on the node that runs the service contains a localhost entry. Otherwise, jobs fail when the **Launch Jobs as Separate Processes** property for the Data Integration Service is enabled.

Verify the Maximum Heap Size for Data Integration Service

If you work with rule specifications in the Analyst tool or in the Developer tool, verify the Maximum Heap Size property on the Data Integration Service. The property determines the amount of memory that the Data Integration Service can use to test rule specifications and to run mappings that contain rule specifications.

Find the Maximum Heap Size property in the Advanced Properties on the Data Integration Service. Verify that the Maximum Heap Size value is at least 2048 MB.

Create and Configure the Data Preparation Service

The Data Preparation Service manages data preparation within Enterprise Data Lake. When an analyst prepares data in a project, the Data Preparation Service stores worksheet metadata in the Data Preparation repository.

The service connects to the Hadoop cluster to read sample data from Hive tables. The service connects to the HDFS system in the Hadoop cluster to store the sample data being prepared in the worksheet.

Create the Data Preparation Service before you create the Enterprise Data Lake Service. You must associate the Enterprise Data Lake Service with a Data Preparation Service.

Creating the Data Preparation Service

Use the service creation wizard in the Administrator tool to create the service.

1. In the Administrator tool, click the **Manage** tab.
2. Click the **Services and Nodes** view.
3. In the Domain Navigator, select the domain.
4. Click **Actions** > **New** > **Data Preparation Service.**

5. Enter the following properties:

Property	Description
Name	Name of the Data Preparation service. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
Description	Description of the Data Preparation service. The description cannot exceed 765 characters.
Location	Location of the Data Preparation Service in the Informatica domain. You can create the service within a folder in the domain.
License	License object with the data lake option that allows the use of the Data Preparation Service.
Node Assignment	Type of node in the Informatica domain on which the Data Preparation Service runs. Select **Single Node** if a single service process runs on the node or **Primary and Backup Nodes** if a service process is enabled on each node for high availability. However, only a single process runs at any given time, and the other processes maintain standby status. The **Primary and Backup Nodes** option will be available for selection based on the license configuration. Select the **Grid** option to ensure horizontal scalability by using grid for the Data Preparation Service with multiple Data Preparation Service nodes. Improved scalability supports high performance, interactive data preparation during increased data volumes and number of users. Each user is assigned a node in the grid using round-robin method to distribute the load across the nodes. Default is Single Node.
Node	Name of the node on which the Data Preparation Service runs.
Backup Nodes	If your license includes high availability, nodes on which the service can run if the primary node is unavailable. Select each backup node on which the service runs.
Grid	Select the grid that you want to use for the Data Preparation Service.

6. Click **Next**.
7. If you plan to use rules, you must associate the Data Preparation Service with the Model Repository Service that manages the Model repository that contains the rule objects and metadata. You must also associate a Data Integration Service with the Data Preparation Service that runs rules during data preparation.

Enter the following properties for the Model Repository Service and the Data Integration Service required to enable rules:

Property	Description
Model Repository Service Name	Name of the Model Repository Service. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [You cannot change the name of the service after you create it.
Model Repository Service User Name	User name to access the Model Repository Service.
Model Repository Service Password	Password to access the Model Repository Service.
Data Integration Service Name	Name of the Data Integration Service.

8. Click **Next**.
9. Enter the following communication properties:

Property	Description
HTTP Port	Port number for the HTTP connection to the Data Preparation Service.
Enable Secure Communication	Use a secure connection to connect to the Data Preparation Service. If you enable secure communication, you must set all required HTTPS properties, including the keystore and truststore properties.
HTTPS Port	Port number for the HTTPS connection to the Data Preparation Service.
Keystore File	Path and the file name of keystore file that contains key and certificates required for HTTPS communication.
Keystore Password	Password for the keystore file.
Truststore File	Path and the file name of truststore file that contains authentication certificates for the HTTPS connection.
Truststore Password	Password for the truststore file.

10. Click **Next**.
11. Enter the following Data Preparation repository database connection properties:

Property	Description
Database Type	Type of database to use for the Data Preparation repository.
Host Name	Host name of the machine that hosts a MySQL database.

Property	Description
Port Number	Port number for a MySQL database.
Connection String	Connection string used to access an Oracle database. Use the following connection string format: `jdbc:informatica:oracle://<database host name>:<port>;ServiceName=<database name>`
Secure JDBC Parameters	Secure JDBC parameters required to access a secure Oracle database. If the database is secure, information such as TrustStore and TrustStorePassword can be included in this field. The information is saved in an encrypted format. Parameters usually configured include the following: `EncryptionMethod=<encryption method>;HostNameInCertificate=<host name>;TrustStore=<truststore file name and path>;TrustStorePassword=<truststore password>;KeyStore==<keystore file name and path>;KeyStorePassword=<keystore password>;ValidateServerCertificate=<true}false>`
Database User Name	Database user account to use to connect to the database.
Database User Password	Password for the database user account.
Schema Name	Schema or database name for a MySQL database.

12. Click **Next**.
13. Enter the following rules execution property:

Property	Description
Rules Server Port	Port used by the rules server managed by the Data Preparation Service. Set the value to an available port on the node where the Data Preparation Service runs.

14. Click **Next**.
15. Enter the following Solr property:

Property	Description
Solr Port	Port number for the Apache Solr server used to provide data preparation recommendations.

16. Click **Next**.

17. Enter the following data preparation properties:

Property	Description
Local Storage Location	Directory for data preparation file storage on the node where the Data Preparation Service runs.
HDFS Connection	HDFS connection for data preparation file storage.
HDFS Storage Location	HDFS location for data preparation file storage. If the connection to the local storage fails, the Data Preparation Service recovers data preparation files from the HDFS location.

18. Click **Next**.
19. Enter the following Hive security properties:

Property	Description
Hadoop Authentication Mode	Security mode enabled for the Hadoop cluster for data preparation storage. If the Hadoop cluster uses Kerberos authentication, you must set the required Hadoop security properties for the cluster.
HDFS Service Principal Name	Service Principal Name (SPN) for the data preparation Hadoop cluster. Specify the service principal name in the following format: user/_HOST@REALM
Hadoop Impersonation User Name	User name to use in Hadoop impersonation as set in the Hadoop connection properties. Use the Administrator tool to view Hadoop connection properties.
SPN Keytab File for User Impersonation	Path and file name of the SPN keytab file for the user account to impersonate when connecting to the Hadoop cluster. The keytab file must be in a directory on the machine where the Data Preparation Service runs.

20. Click **Next**.
21. Enter the following logging configuration property:

Property	Description
Log Severity	Severity of messages to include in the logs. Select from the following values: - FATAL. Writes FATAL messages to the log. FATAL messages include nonrecoverable system failures that cause the service to shut down or become unavailable. - ERROR. Writes FATAL and ERROR code messages to the log. ERROR messages include connection failures, failures to save or retrieve metadata, service errors. - WARNING. Writes FATAL, WARNING, and ERROR messages to the log. WARNING errors include recoverable system failures or warnings. - INFO. Writes FATAL, INFO, WARNING, and ERROR messages to the log. INFO messages include system and service change messages. - TRACE. Write FATAL, TRACE, INFO, WARNING, and ERROR code messages to the log. TRACE messages log user request failures. - DEBUG. Write FATAL, DEBUG, TRACE, INFO, WARNING, and ERROR messages to the log. DEBUG messages are user request logs. Default value is INFO.

22. Click **Finish**.

23. Select the Data Integration Service in the Domain Navigator, and then select **Actions** > **Create Repository** to create the repository contents.
24. Select **Actions** > **Enable Service** to enable the Data Preparation Service.

Create and Configure the Enterprise Data Lake Service

The Enterprise Data Lake Service runs the Enterprise Data Lake application in the Informatica domain. Enterprise Data Lake requires the Enterprise Data Lake Service to complete operations.

When an analyst uploads data, the Enterprise Data Lake Service connects to the HDFS system in the Hadoop cluster to temporarily stage the data. When an analyst previews data, the Enterprise Data Lake Service connects to the Hadoop cluster to read the data.

Creating the Enterprise Data Lake Service

Use the service creation wizard in the Administrator tool to create the service.

1. In the Administrator tool, click the **Manage** tab.
2. Click the **Services and Nodes** view.
3. In the Domain Navigator, select the domain.
4. Click **Actions** > **New** > **Enterprise Data Lake Service.**
5. Enter the following properties:

Property	Description
Name	Name of the Enterprise Data Lake Service. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
Description	Description of theEnterprise Data Lake Service. The description cannot exceed 765 characters.
Location	Location of the Enterprise Data Lake Service in the Informatica domain. You can create the service within a folder in the domain.
License	License object that allows the use of the Enterprise Data Lake Service.
Node Assignment	Type of node in the Informatica domain on which the Enterprise Data Lake Service runs. Select **Single Node** if a single service process runs on the node or **Primary and Backup Nodes** if a service process is enabled on each node for high availability. However, only a single process runs at any given time, and the other processes maintain standby status. The **Primary and Backup Nodes** option is available based on the license configuration. Default is Single Node.

Property	Description
Node	Name of the node on which theEnterprise Data Lake Service runs.
Backup Nodes	If your license includes high availability, nodes on which the service can run if the primary node is unavailable.

6. Click **Next**.
7. Enter the following properties for the Model Repository Service:

Property	Description
Model Repository Service	Name of the Model Repository Service associated with the Enterprise Data Lake Service.
Model Repository Service User Name	User account to use to log in to the Model Repository Service.
Model Repository Service User Password	Password for the Model Repository Service user account.

8. Click **Next**.
9. Enter the following properties for the Data Preparation Service, Data Integration Service, and Catalog Service:

Property	Description
Data Preparation Service	Name of the Data Preparation Service associated with the Enterprise Data Lake Service.
Data Integration Service	Name of the Data Integration Service associated with the Enterprise Data Lake Service.
Catalog Service	Name of the Catalog Service associated with the Enterprise Data Lake Service.
Catalog Service User Name	User account to use to log in to the Catalog Service.
Catalog Service User Password	Password for the Catalog Service user account.

10. Click **Next**.

11. Enter the following data lake security properties:

Property	Description
Hadoop Authentication Mode	Security mode of the Hadoop cluster for the data lake. If the Hadoop cluster uses Kerberos authentication, you must set the required Hadoop security properties for the cluster.
Principal Name for User Impersonation	Service principal name (SPN) of the user account to impersonate when connecting to the data lake Hadoop cluster. The user account for impersonation must be set in the Hadoop connection properties. Use the Administrator tool to view Hadoop connection properties.
SPN Keytab File for User Impersonation	Path and file name of the SPN keytab file for the user account to impersonate when connecting to the Hadoop cluster. The keytab file must be in a directory on the machine where the Enterprise Data Lake Service runs.

12. Click **Next**.
13. Enter the following connection properties:

Property	Description
HDFS Connection	HDFS connection for the data lake.
HDFS Working Directory	HDFS directory where the Enterprise Data Lake Service copies temporary data and files necessary for the service to run.
Hive Connection	Hive connection for the data lake.
Hive Table Storage Format	Data storage format for the Hive tables. Select from the following options: - DefaultFormat - Parquet - ORC
Hadoop Connection	Hadoop connection for the data lake.

14. Click **Next**.
15. Enter the following user event logging properties:

Property	Description
Log User Activity Events	Indicates whether the Enterprise Data Lake Service logs user activity events.
JDBC Port	JDBC port to use to retrieve user activity event data.

16. Click **Next**.

17. Enter the following data asset upload and download properties:

Property	Description
Maximum File Size for Uploads (MB)	The maximum size of the files that can be uploaded.
Maximum Number of Rows to Download	Number of rows to export to a .csv file. You can specify a maximum of 2,000,000,000 rows. Enter a value of -1 to export all rows.

18. Click **Next**.
19. Enter the following asset recommendation property, which configures how Enterprise Data Lake displays recommendations during project creation:

Property	Description
Number of Recommendations to Display	The number of recommended data assets to display on the Projects page. You can specify a maximum of 50 recommendations. A value of 0 means no recommendations will be displayed.

20. Click **Next**.
21. Enter the following sampling properties, which configure how Enterprise Data Lake displays sampling data during data preparation:

Property	Description
Maximum Data Preparation Sample Size	The maximum number of sample rows to fetch for data preparation. You can specify a maximum number of 1,000,000 rows.
Default Data Preparation Sample Size	Number of sample rows to fetch for data preparation. You can specify a maximum number of 1,000,000 rows and a minimum of 1,000 rows.

22. Click **Next**.
23. Enter the following Apache Zeppelin property:

Property	Description
Zeppelin URL	The URL to access the Zeppelin framework. The URL should be in the following format: `http[s]://<zeppelin host name>:<port>`

24. Click **Next**.

25. Enter the following logging configuration property:

Property	Description
Log Severity	Severity of messages to include in the logs. Select from one of the following values: - FATAL. Writes FATAL messages to the log. FATAL messages include nonrecoverable system failures that cause the service to shut down or become unavailable. - ERROR. Writes FATAL and ERROR code messages to the log. ERROR messages include connection failures, failures to save or retrieve metadata, service errors. - WARNING. Writes FATAL, WARNING, and ERROR messages to the log. WARNING errors include recoverable system failures or warnings. - INFO. Writes FATAL, INFO, WARNING, and ERROR messages to the log. INFO messages include system and service change messages. - TRACE. Write FATAL, TRACE, INFO, WARNING, and ERROR code messages to the log. TRACE messages log user request failures. - DEBUG. Write FATAL, DEBUG, TRACE, INFO, WARNING, and ERROR messages to the log. DEBUG messages are user request logs. Default value is INFO.
Log Directory	Location of the directory to save the log files.

26. Click **Next**.
27. Enter the following Hive execution engine and the local system directory properties:

Property	Description
Hive Execution Engine	The Hive execution engine for Enterprise Data Lake Service, which runs mappings in the Hadoop environment.
Local System Directory	Local directory that contains the files downloaded from Enterprise Data Lake, such as .csv or .tde files.

28. Click **Next**.
29. Enter the following advanced properties:

Property	Description
Solr JVM Options	Solr JVM options required to connect to the specified JDBC port used to retrieve user event logs. Set the property to connect from an external client.
Index Directory	Location of a shared NFS directory used by primary and secondary nodes in a multiple node Enterprise Data Lake installation.

30. Click **Next**.

31. Enter the following properties:

Property	Description
HTTP Port	Port number for the HTTP connection to the Enterprise Data Lake Service.
Enable Secure Communication	Use a secure connection to connect to the Enterprise Data Lake Service. If you enable secure communication, you must enter all required HTTPS options.
HTTPS Port	Port number for the HTTPS connection to the Enterprise Data Lake Service.
Keystore File	Path and the file name of keystore file that contains key and certificates required for the HTTPS connection.
Keystore Password	Password for the keystore file.
Truststore File	Path and the file name of the truststore file that contains authentication certificates for the HTTPS connection.
Truststore Password	Password for the truststore file.

32. Select **Enable Service** if you want to enable the service immediately after you create the service.

    If you want to enable the service at a later time, in the Domain Navigator, select the service and then select **Actions** > **Enable Service**.

33. Click **Finish**.

Create and Configure the Analyst Service

The Analyst Service is an application service that runs the Analyst tool in the Informatica domain. The Analyst Service manages the connections between service components and the users that have access to the Analyst tool.

Create the Analyst Service

Use the service creation wizard in the Administrator tool to create the service.

Before you create the Analyst Service, verify that you have created and enabled the following services:

- Model Repository Service

  If the domain does not use Kerberos authentication, verify that you have created a Model repository user that the Analyst Service can use to access the Model Repository Service.

- Data Integration Service

1. In the Administrator tool, click the **Manage** tab.
2. Click **Actions** > **New** > **Analyst Service**.

   The **New Analyst Service** dialog box appears.

3. On the **New Analyst Service - Step 1 of 6** page, enter the following properties:

Property	Description
Name	Name of the service. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
Description	Description of the service. The description cannot exceed 765 characters.
Location	Domain and folder where the service is created. Click **Browse** to choose a different folder. You can move the service after you create it.
License	License object that allows use of the service.
Node	Node on which the service runs.

4. Click **Next**.

   The **New Analyst Service - Step 2 of 6** page appears.

5. Enter the HTTP port number to use for communication from the Analyst tool to the Analyst Service.
6. To enable secure communication from the Analyst tool to the Analyst Service, select **Enable Secure Communication**.

   Enter the following properties to configure secure communication for the Analyst Service:

Property	Description
HTTPS Port	Port number that the Analyst tool runs on when you enable secure communication. Use a different port number than the HTTP port number.
Keystore File	Directory where the keystore file that contains the digital certificates is stored.
Keystore Password	Plain-text password for the keystore file. If this property is not set, the Analyst Service uses the default password `changeit`.
SSL Protocol	Optional. Indicates the protocol to be used. Set this property to `SSL`.

7. Select **Enable Service**.

   The Model Repository Service and the Data Integration Service must be running to enable the Analyst Service.

8. Click **Next**.

   The **New Analyst Service - Step 3 of 6** page appears.

9. Enter the following properties to associate the Model Repository Service with the Analyst Service:

Description	Property
Model Repository Service	Model Repository Service to associate with the service.
User name	User name that the service uses to access the Model Repository Service. Enter the Model repository user that you created.
Password	Password for the Model repository user.
Security Domain	LDAP security domain for the Model repository user. The field appears when the Informatica domain contains an LDAP security domain. Not available for a domain with Kerberos authentication.

10. To enable Analyst tool users to work on Human task data, set the **Data Integration Service** property to the Data Integration Service that you configure to run workflows.

    If Analyst tool users do not need to work on Human task records, do not configure this property.

11. Click **Next**.

    The **New Analyst Service - Step 4 of 6** page appears.

12. Enter the following run-time properties for the Analyst Service:

Property	Description
Data Integration Service	Data Integration Service to associate with the service. The Analyst Service manages the connection to the Data Integration Service that enables users to perform data preview, mapping specification, scorecard, and profile jobs in the Analyst tool. You can associate the Analyst Service with the Data Integration Service that you configured to run workflows. Or, you can associate the Analyst Service with different Data Integration Services for the different operations.
Flat File Cache Directory	Directory of the flat file cache where the Analyst tool stores uploaded flat files. The Data Integration Service must also be able to access this directory. If the Analyst Service and the Data Integration Service run on different nodes, configure the flat file directory to use a shared directory.

13. Click **Next**.

    The **New Analyst Service - Step 5 of 6** page appears.

14. Enter the directory to store the temporary business glossary files that the business glossary export process creates and the directory to store files that content managers attach to the Glossary assets. These directories must be on the node that runs the Analyst Service.

15. Click **Finish**.

    The domain creates and enables the Analyst Service.

After you create the service through the wizard, you can edit the properties or configure other properties.

Create and Configure the Content Management Service

The Content Management Service is an application service that manages reference data. A reference data object contains a set of data values that you can search while performing data quality operations on source data. The Content Management Service also compiles rule specifications into mapplets. A rule specification object describes the data requirements of a business rule in logical terms.

The Content Management Service uses the Data Integration Service to run mappings to transfer data between reference tables and external data sources. The Content Management Service also provides transformations, mapping specifications, and rule specifications with the following types of reference data:

- Address reference data
- Identity populations
- Probabilistic models and classifier models
- Reference tables

Create the Content Management Service

Use the service creation wizard in the Administrator tool to create the service.

Before you create the Content Management Service, verify that you have created and enabled the following services:

- Model Repository Service

  If the domain does not use Kerberos authentication, verify that you have created a Model repository user that the Content Management Service can use to access the Model Repository Service.
- Data Integration Service

1. In the Administrator tool, click the **Manage** tab.
2. Click **Actions** > **New** > **Content Management Service**.

   The **New Content Management Service** dialog box appears.
3. On the **New Content Management Service - Step 1 of 2** page, enter the following properties:

Property	Description
Name	Name of the service. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
Description	Description of the service. The description cannot exceed 765 characters.
Location	Domain and folder where the service is created. Click **Browse** to choose a different folder. You can move the service after you create it.
License	License object that allows use of the service.
Node	Node on which the service runs.

Property	Description
HTTP Port	HTTP port number to use for the Content Management Service.
Data Integration Service	Data Integration Service to associate with the service. The Data Integration Service and the Content Management Service must run on the same node.
Model Repository Service	Model Repository Service to associate with the service.
Username	User name that the service uses to access the Model Repository Service. Enter the Model repository user that you created.
Password	Password for the Model repository user.
Security Domain	LDAP security domain for the Model repository user. The field appears when the Informatica domain contains an LDAP security domain. Not available for a domain with Kerberos authentication.
Reference Data Location	Reference data warehouse connection that you created for the Content Management Service to access the reference data warehouse. Click **Select** to select the connection.

4. Click **Next**.

   The **New Content Management Service - Step 2 of 2** page appears.
5. Accept the default values for the security properties.
6. Select **Enable Service**.

   The Model Repository Service and Data Integration Service must be running to enable the Content Management Service.
7. Click **Finish**.

   The domain creates and enables the Content Management Service.

After you create the service through the wizard, you can edit the properties or configure other properties.

Creating a Catalog Service

Create a Catalog Service to run the Enterprise Data Catalog application and manage the connections between the Enterprise Data Catalog components. You can configure the general, application service, and security properties of the Catalog Service.

If you plan to deploy Enterprise Data Catalog on multiple nodes, ensure that you configure Informatica Cluster Service and Catalog Service on separate nodes.

Note: The Catalog Service has the same privileges as the user account that creates it. Ensure that the user account does not have privileges to read or modify sensitive files on the system.

1. In the Administrator tool, select a domain, and click the **Services and Nodes** tab.
2. On the Actions menu, click **New** > **Catalog Service**.

   The **New Catalog Service Step 1 of 4** dialog box appears.
3. Configure the general properties in the dialog box.

The following table describes the properties:

Property	Description
Name	Name of the service. The name is not case-sensitive and must be unique within the domain. The name cannot exceed 128 characters or begin with @. The name cannot contain character spaces. The characters in the name must be compatible with the code page of the Model repository that you associate with the Catalog Service. The name cannot contain the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
Description	Description of the service. The description cannot exceed 765 characters.
Location	Domain in which the service runs.
License	License to assign to the Catalog Service. Select the license that you installed with Informatica.
Node	Node in the Informatica domain on which the Catalog Service runs. If you change the node, you must recycle the Catalog Service.
Backup Nodes	If your license includes high availability, nodes on which the service can run if the primary node is unavailable.

4. Click **Next**.

   The **New Catalog Service - Step 2 of 4** dialog box appears.

5. Configure the application service properties in the dialog box.

   The following table describes the properties:

Property	Description
Model Repository Service	Model Repository Service to associate with the Catalog Service. The Model Repository Service manages the Model repository that Enterprise Data Catalog uses. If you update the property to specify a different Model Repository Service, recycle the Catalog Service.
User name	The database user name for the Model repository.
Password	An encrypted version of the database password for the Model repository.
Security Domain	Name of the security domain that includes the **User name**.

6. Click **Next**.

   The **New Catalog Service - Step 3 of 4** dialog box appears.

7. Configure the security properties in the dialog box.

The following table describes the properties:

Property	Description
HTTP Port	A unique HTTP port number used for each Data Integration Service process. The defaults is 8085.
Enable Transport Layer Security	Indicates that the Catalog Service must use HTTPS. If you did not configure the Data Integration Service to use HTTPS, the Catalog Service does not start.
HTTPS Port	Port number for the HTTPS connection.
Keystore File	Path and file name of the keystore file. The keystore file contains the keys and certificates required if you use the SSL security protocol with Catalog Administrator. Required if you select Enable Transport layer Security. When Enterprise Data Catalog creates the Catalog Service, Enterprise Data Catalog exports the keystore to a certificate and stores the certificate in the keystore directory. Ensure that you configure the read and write permissions on the directory for Enterprise Data Catalog to successfully store the certificate.
Keystore Password	Password for the keystore file. Required if you select Enable Transport layer Security.
SSL Protocol	Secure Sockets Layer protocol to use.

8. Click **Next**.

   The **New Catalog Service - Step 4 of 4** dialog box appears.

9. Configure the Hadoop cluster properties in the dialog box.

The following table describes the properties:

Property	Description
Cluster Type	Select one of the following options to indicate the deployment type for Enterprise Data Catalog: - External Cluster. Deploy Enterprise Data Catalog in an existing Hadoop cluster on Hortonworks, ClouderaManager, or Azure HDInsight. - Internal Cluster. Deploy Enterprise Data Catalog in the embedded Hadoop cluster on Hortonworks.
Hadoop Distribution	Applicable if you select the **External Cluster** option for **Cluster Type**. Select one of the following options to specify the Hadoop distribution: - ClouderaManager. Use this option if you want to use a ClouderaManager Hadoop distribution. - Hortonworks. Use this option if you want to use a Hortonworks Hadoop distribution. **Note:** If you select ClouderaManager or Hortonworks as the Hadoop distribution, Enterprise Data Catalog automatically identifies the following properties for the Hadoop-distribution type: - ZooKeeper Cluster URI - HDFS Namenode URI - Yarn resource manager URI - Yarn resource manager HTTPS or HTTP URI - History Server HTTP URI - HDFS Service Name for High Availability - Yarn resource manager scheduler URI - HDInsight. Use this option if you want to use an Azure HDInsight Hadoop distribution. - Others. Use this option if you want to manually specify all the properties for a ClouderaManager, Hortonworks, or an Azure HDInsight Hadoop distribution. Make sure that you configure the following custom options for the Catalog Service: - LdmCustomOptions.yarn-site.yarn.application.classpath - LdmCustomOptions.yarn-site.yarn.nodemanager.webapp.address - LdmCustomOptions.yarn-site.yarn.nodemanager.webapp.https.address **Note:** If you select ClouderaManager or Hortonworks, configure the following properties with the other required properties : - Cluster URL. The cluster URL to access the selected Hadoop distribution. - Cluster URL username. The username to access the cluster URL. - Cluster URL password. The password associated with the cluster URL username.
ZooKeeper Cluster URI	Applies to existing cluster. Multiple ZooKeeper addresses in a comma-separated list.
HDFS Namenode URI	Applies to existing cluster. The URI to access HDFS. Use the following format to specify the NameNode URI in the Cloudera distribution:`<Hostname>:<Port>` Where - <host name> is the host name or IP address of the NameNode - <port number> is the port number that the NameNode listens for Remote Procedure Calls (RPC).

Property	Description
Yarn resource manager URI	Applies to existing cluster. The service within Hadoop that submits the MapReduce tasks to specific nodes in the cluster. Use the following format:`<Hostname>:<Port>` Where - <host name> is the name or IP address of the Yarn resource manager. - <port number> is the port number on which Yarn resource manager listens for Remote Procedure Calls (RPC).
Yarn resource manager HTTPS or HTTP URI	Applies to existing cluster. https or http URI value for the Yarn resource manager.
History Server HTTP URI	Applies to existing cluster. Specify a value to generate YARN allocation log files for scanners. Catalog Administrator displays the log URL as part of task monitoring.
HDFS Service Name for High Availability	Applies to highly available existing cluster. Specify the HDFS service name.
Yarn resource manager scheduler URI	Applies to existing cluster. Scheduler URI value for the Yarn resource manager.
Service Cluster Name	Applies to both embedded and existing clusters. Name of the service cluster. Ensure that you have a directory `/Informatica/LDM/<ServiceClusterName>` in HDFS. **Note:** If you do not specify a service cluster name, Enterprise Data Catalog considers `DomainName_CatalogServiceName` as the default value. You must then have the `/Informatica/LDM/<DomainName>_<CatalogServiceName>` directory in HDFS. Otherwise, Catalog Service might fail.
Load Type	Select any of the following options to specify the data size that you plan to load in the catalog: - demo - low - medium - high See the *Tuning Enterprise Data Catalog Performance*How-to-article for more information about data size, load types, and the performance tuning parameter values that Enterprise Data Catalog configures for each load type.
Enable Kerberos Authentication	Select to enable Kerberos authentication for the existing cluster.
HDFS Service Principal Name	Applies to Kerberos authentication. Principal name for the HDFS Service.
YARN Service Principal Name	Applies to Kerberos authentication. Principal name for the YARN Service.
Service Keytab Location	Applies to Kerberos authentication. Path to the keytab file.
Kerberos Domain Name	Applies to Kerberos authentication. Name of the Kerberos domain.

Property	Description
Enable Cluster SSL	Select to enable SSL authentication for secure communication in the existing cluster.
Solr Keystore	Applies to SSL authentication. Path to the Solr keystore file.
Solr Keystore Password	Applies to SSL authentication. Password for the Solr keystore file.
Receive Alerts through Email	Applies to both embedded and existing clusters. Choose to receive email notifications on the Catalog Service status. **Note:** If you select this option, you must enable the Email Service. For more information about enabling Email Service, see the *Administrator Reference for Enterprise Data Catalog* guide.
Enable Catalog Service	Applies to both embedded and existing clusters. Select the option to enable the Catalog Service.
Informatica Cluster Service	Applies to embedded cluster. Name of the Informatica Cluster Service, which is an application service that Enterprise Data Catalog uses in embedded cluster deployment.

10. Click **Finish**.

- Make sure that the krb5.conf file is located in all cluster nodes and domain machines under the `/etc` directory.
- If you did not choose to enable the Catalog Service earlier, you must recycle the service to start it.

Create and Configure the Search Service

The Search Service is an application service that manages search in the Analyst tool and Business Glossary Desktop.

By default, the Search Service returns search results from a Model repository, such as data objects, mapping specifications, profiles, reference tables, rules, scorecards, and business glossary terms. The search results can also include column profile results and domain discovery results from a profiling warehouse.

Create the Search Service

Use the service creation wizard in the Administrator tool to create the service.

Before you create the Search Service, verify that you have created and enabled the following services:

1. Model Repository Service
   If the domain does not use Kerberos authentication, verify that you have created a Model repository user that the Search Service can use to access the Model Repository Service.
2. Data Integration Service
   Verify that the Model repository user has permissions to the Data Integration Service.
3. Analyst Service
   Verify that the Model repository user has permissions to the Analyst Service.

1. In the Administrator tool, click the **Manage** tab.

2. Click **Actions** > **New** > **Search Service**.

   The **New Search Service** dialog box appears.

3. On the **New Search Service - Step 1 of 2** page, enter the following properties:

Property	Description
Name	Name of the service. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
Description	Description of the service. The description cannot exceed 765 characters.
Location	Domain and folder where the service is created. Click **Browse** to choose a different folder. You can move the service after you create it.
License	License object that allows use of the service.
Node	Node on which the service runs.

4. Click **Next**.

   The **New Search Service - Step 2 of 2** page appears.

5. Enter the following search properties for the Search Service:

Description	Property
Port Number	Port number to use for the Search Service.
Index Location	Directory that contains the search index files. Enter a directory on the machine that runs the Search Service. If the directory does not exist, Informatica creates the directory when it creates the Search Service.
Extraction Interval	Interval in seconds at which the Search Service extracts and indexes updated content. Default is 60 seconds.
Model Repository Service	Model Repository Service to associate with the service.
User Name	User name that the service uses to access the Model Repository Service. Enter the Model repository user that you created.
Password	Password for the Model repository user.
Security Domain	LDAP security domain for the Model repository user. The field appears when the Informatica domain contains an LDAP security domain. Not available for a domain with Kerberos authentication.

6. Click **Finish**.

   The domain creates the Search Service. The domain does not enable the Search Service during the creation process. You must enable the Search Service before users can perform searches in the Analyst tool and Business Glossary Desktop.

7. To enable the Search Service, select the service in the Navigator, and then click **Actions** > **Enable Service**.

   The Model Repository Service, Data Integration Service, and Analyst Service must be running to enable the Search Service.

After you create the service through the wizard, you can edit the properties or configure other properties.

Creating an Informatica Cluster Service

You can choose to generate the Informatica Cluster Service when you install Enterprise Data Catalog or create the application service manually using Informatica Administrator.

If you plan to deploy Enterprise Data Catalog on multiple nodes, ensure that you configure Informatica Cluster Service and Catalog Service on separate nodes.

1. In the Administrator tool, select a domain, and click the **Services and Nodes** tab.
2. On the Actions menu, click **New** > **Informatica Cluster Service**.

   The **New Informatica Cluster Service: Step 1 of 4** dialog box appears.
3. Configure the general properties in the dialog box.

   The following table describes the properties:

Property	Description
Name	Name of the service. The name is not case-sensitive and must be unique within the domain. The name cannot exceed 128 characters or begin with @. The name cannot contain character spaces. The characters in the name must be compatible with the code page of the Model repository that you associate with the Catalog Service. The name cannot contain the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
Description	Description of the service. The description cannot exceed 765 characters.
Location	Domain in which the application service runs.
License	License to assign to the Informatica Cluster Service. Select the license that you installed with Enterprise Data Catalog.
Node	Node in the Informatica domain on which the Informatica Cluster Service runs. If you change the node, you must recycle the Informatica Cluster Service.
Backup Nodes	If your license includes high availability, nodes on which the service can run if the primary node is unavailable.

4. Click **Next**.

   The **New Informatica Cluster Service - Step 2 of 4** dialog box appears.
5. Configure the security properties in the dialog box.

The following table describes the properties:

Property	Description
HTTP Port	A unique HTTP port number used for each Data Integration Service process. The defaults is 8085.
Enable Transport Layer Security (TLS)	Select the option to enable TLS for the Informatica Cluster Service.
HTTPS Port	Port number for the HTTPS connection. Required if you select **Enable Transport layer Security**.
Keystore File	Path and file name of the keystore file. The keystore file contains the keys and certificates required if you use the SSL security protocol with Catalog Administrator. Required if you select **Enable Transport layer Security**.
Keystore Password	Password for the keystore file. Required if you select **Enable Transport Layer Security**.
SSL Protocol	Secure Sockets Layer protocol to use.

6. Click **Next**.

   The **New Informatica Cluster Service - Step 3 of 4** dialog box appears.

7. Configure the Hadoop cluster properties in the dialog box.

   The following table describes the properties:

Property	Description
Hadoop Gateway Host	Host where Apache Ambari server runs.
Hadoop Gateway Port	Web port for the Apache Ambari server.
Gateway User	User name for the Apache Ambari server.
Hadoop Nodes	Hosts where the Apache Ambari agents run.
Override default password	Select this option if you want to change the default password for the cluster. Provide the new password in the **Ambari Server Admin Password** text box.

Property	Description
Enable Kerberos Authentication	Select the option to enable Kerberos authentication for the cluster.
KDC Type	Select one of the following Kerberos Key Distribution Center (KDC) types if you had selected the **Enable Kerberos Authentication** option: - Active Directory. Select this option if you want to use Active Directory KDC. - MIT KDC. Select this option if you want to use MIT KDC. Specify the following options after you select the **KDC Type** - KDC Host. Name of the KDC host machine. - Administrator Server Host. The name of the administrator server machine that hosts the KDC server. - Realm. Name of the Kerberos realm on the machine that hosts the KDC server. - Administrator Principal. The Kerberos administrator principal. - Administrator Password. The Kerberos administrator password. - LDAP URL. This property applies to Microsoft Active Directory and represents the URL to the LDAP server directory. - Container DN. This property applies to Microsoft Active Directory and represents the Distinguished Name of the container to which the user belongs. - KDC Certificate Path. Path to the KDC certificate on the Informatica domain machine.

8. Click **Next**.

   The **New Informatica Cluster Service - Step 4 of 4** dialog box appears.

9. Configure the domain security options for Informatica Cluster Service.

   The following table describes the properties:

Property	Description
Domain is SSL Enabled	Specify if the Informatica domain is enabled for SSL.
Domain Truststore File Location	Location to the domain truststore file.
Domain Truststore Password	Password for the domain truststore file.
Enable Service	Select the option to enable the Informatica Cluster Service immediately after you create the service.

10. Click **Finish**.

    Note: After you update the Informatica Cluster Service security options in Informatica Administrator, restart the Informatica Cluster Service.

Before enabling the Informatica Cluster Service in a Kerberos-enabled cluster, verify the following prerequisites:

- You must configure the Key Distribution Center (KDC) hostname and IP address on all cluster nodes and domain machines in the `/etc/hosts`.
- Make sure that the krb5.conf file is located in all cluster nodes and domain machines under the `/etc` directory.
- For an SSL-enabled cluster or a Kerberos-enabled cluster, ensure that the domain truststore file is configured and copied to a common location accessible to all the cluster nodes.

- If the Solr keystore and password are different from the keystore and password of Informatica Cluster Service, you must export the public certificate of Solr to all the cluster nodes and import the certificate to the YARN truststore and domain truststore.

Create and Configure the Metadata Access Service

The Metadata Access Service is an application service that allows the Developer tool to access Hadoop connection information to import and preview metadata.

The Metadata Access Service contains information about the Service Principal Name (SPN) and keytab information if the Hadoop cluster uses Kerberos authentication.

Create the Metadata Access Service

The Metadata Access Service allows the Developer tool to access Hadoop connection information to import and preview metadata from the Hadoop environment. The Metadata Access Service is required for design-time access to the Hadoop environment.

1. In the Administrator tool, click the **Manage** tab.
2. Click the **Services and Nodes** view.
3. In the Domain Navigator, select the domain.
4. Click **Actions** > **New** > **Metadata Access Service**.

   The **New Metadata Access Service** wizard appears.
5. On the **New Metadata Access Service - Step 1 of 3** page, enter the following properties:

Property	Description
Name	Name of the service. The name is not case sensitive and must be unique within the domain. It cannot exceed 128 characters or begin with @. It also cannot contain spaces or the following special characters: ` ~ % ^ * + = { } \ ; : ' " / ? . , < > | ! ( )] [
Description	Description of the service. The description cannot exceed 765 characters.
Location	Domain and folder where the service is created. Click **Browse** to choose a different folder. You can move the service after you create it.
License	License object that allows use of the service.
Node	Node on which the service runs.
Backup Nodes	If your license includes high availability, nodes on which the service can run if the primary node is unavailable.

6. Click **Next**.

   The **New Metadata Access Service - Step 2 of 3** page appears.
7. Select the HTTP Protocol Type and enter the respective port number to use for the Metadata Access Service.

8. Accept the default values for the remaining security properties. You can configure the security properties after you create the Metadata Access Service.
9. Select **Enable Service**.

   The Metadata Access Service does not have any other service dependency.
10. Click **Next**.

    The **New Metadata Access Service - Step 3 of 3** page appears.
11. If applicable, specify the execution options for impersonation user, Kerberos cluster, logging options, and click **Next**.
12. Click **Finish**.

    The domain creates and enables the Metadata Access Service.

CHAPTER 19

Complete the Enterprise Data Lake Configuration

This chapter includes the following topics:

Install Python for Enterprise Data Lake

Enterprise Data Lake uses the Apache Solr indexing capabilities to provide recommendations of related data assets. Apache Solr requires Python modules.

You must install Python with the following modules on every node that hosts the Data Preparation Service:

argparse

sys

getopt

os

urllib

httplib2

ConfigParser

Integrate the Domain with the Hadoop Environment

If you imported the cluster configuration from the Hadoop environment during installation, you must complete the integration between the domain and the Hadoop environment. Integration tasks are required in both the Hadoop environment and the Informatica domain environment.

To integrate the domain with the Hadoop environment, you complete the following high-level tasks:

1. Prepare directories, users, and permissions.

2. Configure *-site.xml files on the Hadoop environment. The properties *-site.xml files must be updated with values required for Informatica processing in the Hadoop environment.
3. Refresh the cluster configuration in the Administrator tool. Refresh the cluster configuration to get the updated properties from the *-site.xml files on the cluster.
4. Update connections in the Administrator tool. Update connections if you want to use property values other than the default values. You will also need to configure environment variables in the Hadoop connection.

Refer to the *Big Data Management Integration Guide* for details about the integration.

Enable Data Preparation of JSON Files on Cloudera CDH

If you integrate Enterprise Data Lake with a Cloudera CDH Hadoop cluster, you must specify the location of a Hive .jar file in the Hive Auxiliary JARs Directory in CDH to enable data preparation of JSON files.

1. Search for hive-hcatalog-core.jar in the Cloudera CDH installation directory.

   You can generally find the .jar file in the following directory:

   `/opt/cloudera/parcels/CDH/lib/hive-hcatalog/share/hcatalog/`
2. Log in to Cloudera Manager.
3. Select **Hive** in the cluster.
4. Click the **Configuration** tab, then select the **Advanced** category.
5. Enter the path to the directory containing the .jar file in the Hive Auxiliary JARs Directory property.

   If the file is in the location noted, the path to the directory is:

   `/opt/cloudera/parcels/CDH/lib/hive-hcatalog/share/hcatalog/`
6. Restart the Hive server.

Part V: Install the Developer Tool

This part contains the following chapter:

CHAPTER 20

Install the Developer Tool

This chapter includes the following topics:

Before You Install Informatica Developer

Before you install the Informatica Developer, verify that the minimum system and third-party software requirements are met. If the machine where you install the Informatica Developer is not configured correctly, the installation can fail.

Verify Installation Requirements

Before you install the Developer tool, verify the installation requirements to run the Developer tool are met.

You can install all the Developer tool on the same machine or on separate machines. You can also install the Developer tool on multiple machines.

Before you install the Developer tool, verify the following installation requirements:

Disk space for the temporary files

The installer writes temporary files to the hard disk. Verify that you have 1 GB disk space on the machine to support the installation. When the installation completes, the installer deletes the temporary files and releases the disk space.

Permissions to install the Developer tool

Verify that the user account that you use to install the Developer tool has write permission on the installation directory and Windows registry.

Minimum system requirements to run the Developer tool

The following table lists the minimum system requirements to run the Developer tool:

Client	Processor	RAM	Disk Space
Informatica Developer	1 CPU	1GB	6 GB

Install Informatica Developer

You can install the Informatica Developer in graphical or silent mode. Complete the pre-installation tasks to prepare for the installation. You can install the Informatica Developer on multiple machines.

Installing in Graphical Mode

You can install the Developer tool in graphical mode on Windows.

1. Close all other applications.
2. Go to the root of the directory for the installation files and run install.bat as administrator.

   To run the file as administrator, right-click the install.bat file and select **Run as administrator**.

   Note: If you do not run the installer as administrator, the Windows system administrator might encounter issues when accessing files in the Informatica installation directory.

   If you encounter problems when you run the install.bat file from the root directory, run the following file:
   `<installer files directory>\client\install.exe`
3. Select **Install Informatica <Version> Clients** and click **Next**.
4. Read the terms and conditions for Informatica installation and the product usage toolkit and select **I agree to the terms and conditions**.

   Informatica DiscoveryIQ is a product usage tool that sends routine reports on data usage and system statistics to Informatica. Informatica DiscoveryIQ uploads data to Informatica 15 minutes after you install and configure Informatica domain. Thereafter, the domain sends the data every 30 days. You can choose to disable usage statistics from the Administrator tool.

   a. Press **1** if you do not want to accept the terms and conditions.

   b. Press **2** to accept the terms and conditions.
5. Version 10.2.2 is for big data products only, such as Big Data Management and Big Data Quality. This version does not support non-big data products, such as PowerCenter or Informatica Data Quality.

   a. Press **1** and type **quit** to quit the installation.

   b. Press **2** to continue the installation.

   If you choose to not accept the terms and condition, the installer prompts you to accept the terms and conditions.
6. The **Installation Pre-requisites** page displays the system requirements. Verify that all installation requirements are met before you continue the installation.
7. On the **Installation Directory** page, enter the absolute path for the installation directory.

   The installation directory must be on the current computer. The maximum length of the path must be less than 260 characters. The directory names in the path must not contain spaces or the following special characters: @|* $ # ! % ( ) { } [ ] , ; '

   Note: Informatica recommends using alphanumeric characters in the installation directory path. If you use a special character such as á or €, unexpected results might occur at run time.
8. Click **Next**.
9. On the **Pre-Installation Summary** page, review the installation information, and click **Install**.

   The installer copies the Developer tool files to the installation directory.

   The **Post-installation Summary** page indicates whether the installation completed successfully.
10. Click **Done** to close the installer.

You can view the installation log files to get more information about the tasks performed by the installer.

Installing in Silent Mode

To install the Informatica clients without user interaction, install in silent mode.

Use a properties file to specify the installation options. The installer reads the file to determine the installation options. You can use silent mode installation to install the Informatica clients on multiple machines on the network or to standardize the installation across machines.

To install in silent mode, complete the following tasks:

1. Configure the installation properties file and specify the installation options in the properties file.
2. Run the installer with the installation properties file.

Configuring the Properties File

Informatica provides a sample properties file that includes the properties required by the installer. Customize the sample properties file to create a properties file and specify the options for your installation. Then run the silent installation.

The sample SilentInput.properties file is stored in the root directory of the DVD or the installer download location.

1. Go to the root of the directory that contains the installation files.
2. Locate the sample `SilentInput.properties` file.
3. Create a backup copy of the SilentInput.properties file.
4. Use a text editor to open and modify the values of the properties in the file.

   The following table describes the installation properties that you can modify:

Property Name	Description
INSTALL_TYPE	Indicates whether to install or upgrade the Informatica clients. If the value is 0, the Informatica clients are installed in the directory you specify. If the value is 1, the Informatica clients are upgraded. Default is 0.
UPGRADE_WITHOUT_PC	Informatica does not support PowerCenter for version 10.2.2. If you want to install or upgrade to this version, the PowerCenter functionality will not be available. Set the value to 1, to continue with the installation. Set the value to 0, to quit the installer.
UPG_BACKUP_DIR	Directory of the previous version of the Informatica client that you want to upgrade.
USER_INSTALL_DIR	Informatica client installation directory.

5. Save the properties file.

Running the Silent Installer

After you configure the properties file, open a command prompt to start the silent installation.

1. Open a command prompt.

2. Go to the root of the directory that contains the installation files.
3. Verify that the directory contains the file SilentInput.properties that you edited and resaved.
4. To run the silent installation, run silentInstall.bat.

   The silent installer runs in the background. The process can take a while. The silent installation is complete when the Informatica_<Version>_Client_InstallLog<timestamp>.log file is created in the installation directory.

   The silent installation fails if you incorrectly configure the properties file or if the installation directory is not accessible. View the installation log files and correct the errors. Then run the silent installation again.

After You install Informatica Developer

After you install Informatica Developer, you can enable secure communication within the domain and start the Developer tool.

Install Languages

To view languages other than the system locale and to work with repositories that use a UTF-8 code page, install additional languages on Windows for use with the Informatica clients.

You also must install languages to use the Windows Input Method Editor (IME).

1. Click **Start** > **Settings** > **Control Panel**.
2. Click **Regional Options**.
3. Under Language settings for the system, select the languages you want to install.
4. Click **Apply**.

If you change the system locale when you install the language, restart the Windows machine.

Configure the Client for a Secure Domain

When you enable secure communication within the domain, you also secure connections between the domain and Informatica client applications, such as the Developer tool. Based on the truststore files used, you might need to specify the location and password for the truststore files in environment variables on each client host.

You might need to set the following environment variables on each client host:

INFA_TRUSTSTORE

Set this variable to the directory that contains the truststore files for the SSL certificates. The directory must contain truststore files named `infa_truststore.jks` and `infa_truststore.pem`.

INFA_TRUSTSTORE_PASSWORD

Set this variable to the password for the `infa_truststore.jks` file. The password must be encrypted. Use the command line program pmpasswd to encrypt the password.

Informatica provides an SSL certificate that you can use to secure the domain. When you install the Informatica clients, the installer sets the environment variables and installs the truststore files in the following directory by default: `<Informatica installation directory>\clients\shared\security`

If you use the default Informatica SSL certificate, and the `infa_truststore.jks` and `infa_truststore.pem` are in the default directory, you do not need to set the INFA_TRUSTSTORE or INFA_TRUSTSTORE_PASSWORD environment variables.

You must set the INFA_TRUSTSTORE and INFA_TRUSTSTORE_PASSWORD environment variables on each client host in the following scenarios:

You use a custom SSL certificate to secure the domain.

If you provide an SSL certificate to use to secure the domain, copy the `infa_truststore.jks` and `infa_truststore.pem` truststore files to each client host. You must specify the location of the files and the truststore password.

You use the default Informatica SSL certificate, but the truststore files are not in the default Informatica directory.

If you use the default Informatica SSL certificate, but the `infa_truststore.jks` and `infa_truststore.pem` truststore files are not in the default Informatica directory, you must specify the location of the files and the truststore password.

Configure the Developer Tool Workspace Directory

Configure Informatica Developer to write the workspace metadata to the machine where the user is logged in.

1. Go to the following directory: `<Informatica installation directory>\clients\DeveloperClient\configuration\`
2. Locate the config.ini file.
3. Create a backup copy of the config.ini file.
4. Use a text editor to open the config.ini file.
5. Add the `osgi.instance.area.default` variable to the end of the config.ini file and set the variable to the directory location where you want to save the workspace metadata. The file path cannot contain non-ANSI characters. Folder names in the workspace directory cannot contain the number sign (#) character. If folder names in the workspace directory contain spaces, enclose the full directory in double quotes.
   - If you run Informatica Developer from the local machine, set the variable to the absolute path of the workspace directory:

     ```
     osgi.instance.area.default=<Drive>/<WorkspaceDirectory>
     ```

     or

     ```
     osgi.instance.area.default=<Drive>\\<WorkspaceDirectory>
     ```
   - If you run Informatica Developer from a remote machine, set the variable to the directory location on the local machine:

     ```
     osgi.instance.area.default=\\\\<LocalMachine>/<WorkspaceDirectory>
     ```

     or

     ```
     osgi.instance.area.default=\\\\<LocalMachine>\\<WorkspaceDirectory>
     ```

     The user must have write permission to the local workspace directory.

Informatica Developer writes the workspace metadata to the workspace directory. If you log into Informatica Developer from a local machine, Informatica Developer writes the workspace metadata to the local machine. If the workspace directory does not exist on the machine from which you logged in, Informatica Developer creates the directory when it writes the files.

You can override the workspace directory when you start Informatica Developer.

Starting the Developer Tool

When you start the Developer tool, you connect to a Model repository. The Model repository stores metadata created in the Developer tool. The Model Repository Service manages the Model repository. Connect to the repository before you create a project.

1. From the Windows Start menu, click **Programs** > **Informatica[Version]** > **Client** > **Developer Client** > **Launch Informatica Developer**.

   The first time you run the Developer tool, the Welcome page displays several icons. The Welcome page does not appear when you run the Developer tool subsequently.
2. Click **Workbench.**

   The first time you start the Developer tool, you must select the repository in which to save the objects you create.
3. Click **File** > **Connect to Repository**.

   The **Connect to Repository** dialog box appears.
4. If you have not configured a domain in the Developer tool, click **Configure Domains** to configure a domain.

   You must configure a domain to access a Model Repository Service.
5. Click **Add** to add a domain.

   The **New Domain** dialog box appears.
6. Enter the domain name, host name, and port number.
7. Click **Finish**.
8. Click **OK**.
9. In the **Connect to Repository** dialog box, click **Browse** and select the Model Repository Service.
10. Click **OK**.
11. Click **Next**.
12. Enter a user name and password.
13. Click **Finish**.

    The Developer tool adds the Model repository to the Object Explorer view. When you run the Developer tool the next time, you can connect to the same repository.

Part VI: Uninstallation

This part contains the following chapter:

CHAPTER 21

Uninstallation

This chapter includes the following topics:

Informatica Uninstallation Overview

Uninstall Informatica to remove the Informatica server or clients from a machine.

The Informatica uninstallation process deletes all Informatica files and clears all Informatica configurations from a machine. The uninstallation process does not delete files that are not installed with Informatica. For example, the installation process creates temporary directories. The uninstaller does not keep a record of these directories and therefore cannot delete them. You must manually delete these directories for a clean uninstallation.

When you install the Informatica server or Informatica clients, the installer creates an uninstaller. The uninstaller is stored in the uninstallation directory.

The following table lists the uninstallation directory for each type of installation:

Installation	Uninstallation Directory Name
Informatica Server	`<Informatica installation directory>/Uninstaller_Server`
Informatica Clients	`<Informatica installation directory>/Uninstaller_Client`

To uninstall Informatica, use the uninstaller created during the installation. On Linux, uninstall Informatica from the command line. On Windows, uninstall Informatica from the Windows Start menu or Control Panel.

Rules and Guidelines for Uninstallation

Use the following rules and guidelines when you uninstall Informatica components:

- The Informatica server uninstallation mode depends on the mode you use to install Informatica server. For example, you install Informatica server in console mode. When you run the uninstaller, it runs in console mode. The Informatica clients uninstallation mode does not depend on the mode you use to install Informatica clients. For example, you install Informatica clients in silent mode. When you run the uninstaller, it can run in graphical or silent mode.
- Uninstalling Informatica does not affect the Informatica repositories. The uninstaller removes the Informatica files. It does not remove repositories from the database. If you need to move the repositories, you can back them up and restore them to another database.
- Uninstalling Informatica does not remove the metadata tables from the domain configuration database. If you install Informatica again using the same domain configuration database and user account, you must manually remove the tables or choose to overwrite the tables. You can use the infasetup BackupDomain command to back up the domain configuration database before you overwrite the metadata tables. To remove the metadata tables manually, use the infasetup DeleteDomain command before you run the uninstaller.
- Uninstalling Informatica removes all installation files and subdirectories from the Informatica installation directory. Before you uninstall Informatica, stop all Informatica services and processes and verify that all of the files in the installation directory are closed. At the end of the uninstallation process, the uninstaller displays the names of the files and directories that could not be removed.
- The Informatica server installation creates the following folder for the files and libraries required by third party adapters built using the Informatica Development Platform APIs:
  `<Informatica installation directory>/services/shared/extensions`

  Uninstalling the Informatica server deletes this folder and any subfolders created under it. If you have adapter files stored in the /extensions folder, back up the folder before you start uninstallation.
- If you perform the uninstallation on a machine, you must back up the ODBC folder before you uninstall. Restore the folder after the uninstallation completes.

Uninstalling the Informatica Server in Console Mode

If you installed the Informatica server in console mode, uninstall the Informatica server in console mode.

Before you run the uninstaller, stop all Informatica services and processes and verify that all files in the installation directory are closed. The uninstallation process cannot remove files that are open or are being used by a service or process that is running.

1. Go to the following directory:

   `<Informatica installation directory>/Uninstaller_Server`
2. Type the following command to run the uninstaller:

   ```
   ./uninstaller
   ```

   If you installed the Informatica server in console mode, the uninstaller launches in console mode.

Uninstalling Informatica Server in Silent Mode

If you installed the Informatica server in silent mode, uninstall the Informatica server in silent mode.

Before you run the uninstaller, stop all Informatica services and processes and verify that all files in the installation directory are closed. The uninstallation process cannot remove files that are open or are being used by a service or process that is running.

1. Go to the following directory:

   ```
   <Informatica installation directory>/Uninstaller_Server
   ```

2. Type the following command to run the silent uninstaller:

   ```
   ./uninstaller.sh
   ```

   If you installed the Informatica server in silent mode, the uninstaller launches in silent mode. The silent uninstaller runs in the background. The process can take a while. The silent uninstallation fails if the installation directory is not accessible.

After you uninstall the the Informatica server, delete any remaining folders and files from the Informatica installation directory. For example:

- Informatica _<Version>_Services_InstallLog.log file
- Informatica_<Version>_Services_<timestamp>.log file

Informatica Developer Tool Uninstallation

You can uninstall the Informatica Developer tool in graphical mode and silent mode on Windows.

When you uninstall Informatica Developer tool, the installer does not remove the environment variables, INFA_TRUSTSTORE, that it creates during installation. When you install a later version of Informatica Developer tool, you must edit the environment variable to point to the new value of the SSL certificate.

Uninstalling Informatica Clients in Graphical Mode

If you installed the Informatica clients in graphical mode, uninstall the Informatica clients in graphical mode.

1. Click **Start** > **Program Files** > **Informatica [Version]** > **Client** > **Uninstaller**.

   The **Uninstallation** page appears.

2. Click **Next**.

   The **Application Client Uninstall Selection** page appears.

3. Select the client applications you want to uninstall and click **Uninstall**.

4. Click **Done** to close the uninstaller.

   After the uninstallation is complete, the **Post-Uninstallation Summary** page appears, displaying the results of the uninstallation.

After you uninstall the Informatica clients, delete any remaining folders and files from the Informatica installation directory. For example:

- Informatica _<Version>_Client_InstallLog.log file
- Informatica_<Version>_Client.log file

Log out of the machine and log back in. Then clear the Informatica-specific CLASSPATH and PATH environment variables.

Uninstalling Informatica Clients in Silent Mode

If you installed the Informatica clients in silent mode, uninstall the Informatica clients in silent mode.

Creating the Properties File

Informatica provides a sample properties file that includes the properties required by the installer.

Customize the sample properties file to create a properties file and specify the options for your uninstallation. Then run the silent uninstallation.

1. Go to `<Informatica installation directory>/Uninstaller_Client`.
2. Locate the sample `SilentInput.properties` file.
3. Create a backup copy of the SilentInput.properties file.
4. Use a text editor to open and modify the values of the properties file.

   The following table describes the properties that you can modify:

Property Name	Description
DXT_COMP	Indicates whether to uninstall Informatica Developer. If the value is 1, the Developer tool will be uninstalled. If the value is 0, the Developer tool will not be uninstalled. Default is 1.

5. Save the `SilentInput.properties` file.

Running the Silent Uninstaller

After you configure the properties file, run the silent uninstallation.

1. Go to `<Informatica installation directory>/Uninstaller_Client`.
2. To run the silent installation, double-click the `uninstaller.bat` or `uninstaller.exe` file.

   The silent uninstaller runs in the background. The process can take a while. The silent uninstallation fails if you incorrectly configure the properties file or if the installation directory is not accessible.

After you uninstall the Informatica clients, delete any remaining folders and files from the Informatica installation directory. For example:

- Informatica _<Version>_Client_InstallLog.log file
- Informatica_<Version>_Client.log file

Log out of the machine and log back in. Then clear the Informatica-specific CLASSPATH and PATH environment variables.

APPENDIX A

Starting and Stopping Informatica Services

This appendix includes the following topics:

Starting and Stopping Informatica Services Overview

On each node where you install Informatica, the installer creates a Linux daemon to run Informatica. When the installation completes successfully, the installer starts the Informatica daemon on Linux.

The Informatica service runs the Service Manager on the node. The Service Manager manages all domain functions and starts application services configured to run on the node. The method you use to start or stop Informatica depends on the operating system. You can use Informatica Administrator to shut down a node. When you shut down a node, you stop Informatica on the node.

You can configure the behavior of the Informatica service.

The Informatica service also runs Informatica Administrator. You use Informatica Administrator to administer the Informatica domain objects and user accounts. Log in to Informatica Administrator to create the user accounts for users of Informatica and to create and configure the application services in the domain.

Starting and Stopping the Informatica Services

On Linux, run infaservice.sh to start and stop the Informatica daemon. By default, infaservice.sh is installed in the following directory:

```
<Informatica installation directory>/tomcat/bin
```

1. Go to the directory where infaservice.sh is located.

2. At the command prompt, enter the following command to start the daemon:

   ```
   infaservice.sh startup
   ```

   Enter the following command to stop the daemon:

   ```
   infaservice.sh shutdown
   ```

   Note: If you use a softlink to specify the location of infaservice.sh, set the INFA_HOME environment variable to the location of the Informatica installation directory.

Stopping Informatica in Informatica Administrator

When you shut down a node using Informatica Administrator, you stop the Informatica service on that node.

You can abort the processes that are running or allow them to complete before the service shuts down. If you shut down a node and abort the repository service processes running on the node, you can lose changes that have not yet been written to the repository. If you abort a node running integration service processes, the workflows will abort.

1. Log in to Informatica Administrator.
2. In the Navigator, select the node to shut down.
3. On the Domain tab **Actions** menu, select **Shutdown Node**.

Rules and Guidelines for Starting or Stopping Informatica

Consider the following rules and guidelines when starting and stopping Informatica on a node:

- When you shut down a node, the node is unavailable to the domain. If you shut down a gateway node and do not have another gateway node in the domain, the domain is unavailable.
- When you start Informatica, verify that the port used by the service on the node is available. For example, if you stop Informatica on a node, verify that the port is not used by any other process on the machine before you restart Informatica. If the port is not available, Informatica will fail to start.
- If you do not use Informatica Administrator to shut down a node, any process running on the node will be aborted. If you want to wait for all processes to complete before shutting down a node, use Informatica Administrator.
- If you have two nodes in a domain with one node configured as a primary node for an application service and the other node configured as a backup node, start Informatica on the primary node before you start the backup node. Otherwise, the application service will run on the backup node and not the primary node.

APPENDIX B

Connecting to Databases

This appendix includes the following topics:

Connecting to Databases from Linux Overview

To use native connectivity, you must install and configure the database client software for the database that you want to access. To ensure compatibility between the application service and the database, install a client software that is compatible with the database version and use the appropriate database client libraries. To increase performance, use native connectivity.

The Informatica installation includes DataDirect ODBC drivers. If you have existing ODBC data sources created with an earlier version of the drivers, you must create new ODBC data sources using the new drivers. Configure ODBC connections using the DataDirect ODBC drivers provided by Informatica or third party ODBC drivers that are Level 2 compliant or higher.

You must configure a database connection for the following services in the Informatica domain:

- Model Repository Service
- Data Integration Service
- Analyst Service

When you connect to databases from Linux, use native drivers to connect to IBM DB2, Oracle, or Sybase ASE databases. You can use ODBC to connect to other sources and targets.

Connecting to an IBM DB2 Universal Database

For native connectivity, install the version of IBM DB2 Client Application Enabler (CAE) appropriate for the IBM DB2 database server version. To ensure compatibility between Informatica and databases, use the appropriate database client libraries.

Configuring Native Connectivity

You can configure native connectivity to an IBM DB2 database to increase performance.

The following steps provide a guideline for configuring native connectivity. For specific instructions, see the database documentation.

1. To configure connectivity on the machine where the Data Integration Service process runs, log in to the machine as a user who can start a service process.
2. Set the DB2INSTANCE, INSTHOME, DB2DIR, and PATH environment variables.

   The UNIX IBM DB2 software always has an associated user login, often db2admin, which serves as a holder for database configurations. This user holds the instance for DB2.

   DB2INSTANCE. The name of the instance holder.

   Using a Bourne shell:

   ```
   $ DB2INSTANCE=db2admin; export DB2INSTANCE
   ```

   Using a C shell:

   ```
   $ setenv DB2INSTANCE db2admin
   ```

   INSTHOME. This is db2admin home directory path.

   Using a Bourne shell:

   ```
   $ INSTHOME=~db2admin
   ```

   Using a C shell:

   ```
   $ setenv INSTHOME ~db2admin>
   ```

   DB2DIR. Set the variable to point to the IBM DB2 CAE installation directory. For example, if the client is installed in the /opt/IBM/db2/V9.7 directory:

   Using a Bourne shell:

   ```
   $ DB2DIR=/opt/IBM/db2/V9.7; export DB2DIR
   ```

   Using a C shell:

   ```
   $ setenv DB2DIR /opt/IBM/db2/V9.7
   ```

   PATH. To run the IBM DB2 command line programs, set the variable to include the DB2 bin directory.

   Using a Bourne shell:

   ```
   $ PATH=${PATH}:$DB2DIR/bin; export PATH
   ```

   Using a C shell:

   ```
   $ setenv PATH ${PATH}:$DB2DIR/bin
   ```

3. Set the shared library variable to include the DB2 lib directory.

   The IBM DB2 client software contains a number of shared library components that the Data Integration Service processes load dynamically. Set the shared library environment variable so that the services can find the shared libraries at run time.

   The shared library path must also include the Informatica installation directory *(server_dir)*.

   Set the shared library environment variable based on the operating system.

The following table describes the shared library variables for each operating system:

Operating System	Variable
Solaris	LD_LIBRARY_PATH
Linux	LD_LIBRARY_PATH
AIX	LIBPATH

For example, use the following syntax for Solaris and Linux:

- Using a Bourne shell:

```
$ LD_LIBRARY_PATH=${LD_LIBRARY_PATH}:$HOME/server_dir:$DB2DIR/lib; export
LD_LIBRARY_PATH
```

- Using a C shell:

```
$ setenv LD_LIBRARY_PATH ${LD_LIBRARY_PATH}:$HOME/server_dir:$DB2DIR/lib
```

For AIX:

- Using a Bourne shell:

```
$ LIBPATH=${LIBPATH}:$HOME/server_dir:$DB2DIR/lib; export LIBPATH
```

- Using a C shell:

```
$ setenv LIBPATH ${LIBPATH}:$HOME/server_dir:$DB2DIR/lib
```

4. Edit the .cshrc or .profile to include the complete set of shell commands. Save the file and either log out and log in again or run the source command.

   Using a Bourne shell:

   ```
   $ source .profile
   ```

   Using a C shell:

   ```
   $ source .cshrc
   ```

5. If the DB2 database resides on the same machine on which the Data Integration Service process runs, configure the DB2 instance as a remote instance.

   Run the following command to verify if there is a remote entry for the database:

   ```
   DB2 LIST DATABASE DIRECTORY
   ```

   The command lists all the databases that the DB2 client can access and their configuration properties. If this command lists an entry for “Directory entry type” of “Remote,” skip to Step 7on page 329.

6. If the database is not configured as remote, run the following command to verify whether a TCP/IP node is cataloged for the host:

   ```
   DB2 LIST NODE DIRECTORY
   ```

   If the node name is empty, you can create one when you set up a remote database. Use the following command to set up a remote database and, if needed, create a node:

   ```
   db2 CATALOG TCPIP NODE <nodename> REMOTE <hostname_or_address> SERVER <port number>
   ```

   Run the following command to catalog the database:

   ```
   db2 CATALOG DATABASE <dbname> as <dbalias> at NODE <nodename>
   ```

   For more information about these commands, see the database documentation.

7. Verify that you can connect to the DB2 database. Run the DB2 Command Line Processor and run the command:

   ```
   CONNECT TO <dbalias> USER <username> USING <password>
   ```

   If the connection is successful, clean up with the `CONNECT RESET` or `TERMINATE` command.

Connecting to Microsoft SQL Server

Use the Microsoft SQL Server connection to connect to a Microsoft SQL Server database from a Linux machine.

Configuring Native Connectivity

You must choose ODBC or OLEDB as the provider type while configuring a Microsoft SQL Server connection.

The server name and database name are retrieved from the connect string if you enable the Use DSN option. The connect string is the DSN configured in the odbc.ini file. If you do not enable the Use DSN option, you must specify the server name and database name in the connection properties. If you cannot to connect to the database, verify that you correctly entered all of the connectivity information.

After you upgrade, the Microsoft SQL Server connection is set to the OLEDB provider type by default. It is recommended that you upgrade all your Microsoft SQL Server connections to use the ODBC provider type. You can upgrade all your Microsoft SQL Server connections to the ODBC provider type by using the following commands:

- If you are using the Informatica platform, run the following command: infacmd.sh isp upgradeSQLSConnection

After you run the upgrade command, you must set the environment variable on each machine that hosts the Developer tool and on the machine that hosts Informatica services in the following format:

```
ODBCINST=<INFA_HOME>/ODBC7.1/odbcinst.ini
```

After you set the environment variable, you must restart the node that hosts the Informatica services.

For specific connectivity instructions, see the database documentation.

Rules and Guidelines for Microsoft SQL Server

Consider the following rules and guidelines when you configure ODBC connectivity to a Microsoft SQL Server database on Windows:

- If you want to use a Microsoft SQL Server connection without using a Data Source Name (DSN less connection), you must configure the odbcinst.ini environment variable.
- If you are using a DSN connection, you must add the entry "EnableQuotedIdentifiers=1" to the ODBC DSN. If you do not add the entry, data preview and mapping run fail.
- When you use a DSN connection, you can configure the DataDirect specific properties. For more information about how to configure and use the Data Direct specific properties, see the DataDirect documentation.
- You can use the Microsoft SQL Server NTLM authentication on a DSN less Microsoft SQL Server connection on the Microsoft Windows platform.
- If the Microsoft SQL Server table contains a UUID data type and if you are reading data from an SQL table and writing data to a flat file, the data format might not be consistent between the OLE DB and ODBC connection types.
- You cannot use SSL connection on a DSN less connection. If you want to use SSL, you must use the DSN connection. Enable the Use DSN option and configure the SSL options in the odbc.ini file.
- If the Microsoft SQL Server uses Kerberos authentication, you must set the GSSClient property to point to the Informatica Kerberos libraries. Use the following path and filename: `<Informatica installation directory>/server/bin/libgssapi_krb5.so.2`.Create an entry for the GSSClient property in the DSN

entries section in `odbc.ini` for a DSN connection or in the SQL Server wire protocol section in `odbcinst.ini` for a connection that does not use DSN.

- If you use the DataDirect ODBC driver to connect to Microsoft SQL Server, the Decimal data rounds off within the target database based on the scale values in the database tables. For example, if the scale is 5, the target Decimal data round-off occurs after the fifth digit that follows the decimal point. When the scale is 5, the input value 12.3456789 rounds off to a target Decimal value of 12.34568.
- If you use Microsoft SQL Sever Native client to configure native connectivity to Microsoft SQL Server databases, the Decimal data truncates based on the specified scale in the target database tables. For example, if the scale is 5, the Decimal data truncation occurs after the fifth digit that follows the decimal point. When the scale is 5, the input value 12.3456789 truncates to a target Decimal value of 12.34567.

Configuring SSL Authentication through ODBC

You can configure SSL authentication for Microsoft SQL Server through ODBC using the DataDirect New SQL Server Wire Protocol driver.

1. Open the odbc.ini file and add an entry for the ODBC data source and DataDirect New SQL Server Wire Protocol driver under the section [ODBC Data Sources].
2. Add the attributes in the odbc.ini file for configuring SSL.

   The following table lists the attributes that you must add to the odbc.ini file when you configure SSL authentication:

Attribute	Description
EncryptionMethod	The method that the driver uses to encrypt the data sent between the driver and the database server. Set the value to 1 to encrypt data using SSL.
ValidateServerCertificate	Determines whether the driver validates the certificate sent by the database server when SSL encryption is enabled. Set the value to 1 for the driver to validate the server certificate.
TrustStore	The location and name of the trust store file. The trust store file contains a list of Certificate Authorities (CAs) that the driver uses for SSL server authentication.
TrustStorePassword	The password to access the contents of the trust store file.
HostNameInCertificate	Optional. The host name that is established by the SSL administrator for the driver to validate the host name contained in the certificate.

Connecting to a Netezza Database

Install and configure Netezza ODBC driver on the machine where the Data Integration Service process runs. Use the DataDirect Driver Manager in the DataDirect driver package shipped with the Informatica product to configure the Netezza data source details in the odbc.ini file.

Configuring ODBC Connectivity

You can configure ODBC connectivity to a Netezza database.

The following steps provide a guideline for configuring ODBC connectivity. For specific instructions, see the database documentation.

1. To configure connectivity for the integration service process, log in to the machine as a user who can start a service process.
2. Set the ODBCHOME, NZ_ODBC_INI_PATH, and PATH environment variables.

   ODBCHOME. Set the variable to the ODBC installation directory. For example:

   Using a Bourne shell:

   ```
   $ ODBCHOME=<Informatica server home>/ODBC7.1; export ODBCHOME
   ```

   Using a C shell:

   ```
   $ setenv ODBCHOME =<Informatica server home>/ODBC7.1
   ```

   PATH. Set the variable to the ODBCHOME/bin directory. For example:

   Using a Bourne shell:

   ```
   PATH="${PATH}:$ODBCHOME/bin"
   ```

   Using a C shell:

   ```
   $ setenv PATH ${PATH}:$ODBCHOME/bin
   ```

   NZ_ODBC_INI_PATH. Set the variable to point to the directory that contains the odbc.ini file. For example, if the odbc.ini file is in the $ODBCHOME directory:

   Using a Bourne shell:

   ```
   NZ_ODBC_INI_PATH=$ODBCHOME; export NZ_ODBC_INI_PATH
   ```

   Using a C shell:

   ```
   $ setenv NZ_ODBC_INI_PATH $ODBCHOME
   ```

3. Set the shared library environment variable.

   The shared library path must contain the ODBC libraries. It must also include the Informatica services installation directory `(server_dir)`.

   Set the shared library environment variable based on the operating system. Set the Netezza library folder to <NetezzaInstallationDir>/lib64.

   The following table describes the shared library variables for each operating system:

Operating System	Variable
Solaris	LD_LIBRARY_PATH
Linux	LD_LIBRARY_PATH
AIX	LIBPATH

   For example, use the following syntax for Solaris and Linux:

   - Using a Bourne shell:

     ```
     $ LD_LIBRARY_PATH="${LD_LIBRARY_PATH}:$HOME/server_dir:$ODBCHOME/
     lib:<NetezzaInstallationDir>/lib64”
     export LD_LIBRARY_PATH
     ```

- Using a C shell:

```
$ setenv LD_LIBRARY_PATH "${LD_LIBRARY_PATH}:$HOME/server_dir:$ODBCHOME/
lib:<NetezzaInstallationDir>/lib64"
```

For AIX

- Using a Bourne shell:

```
$ LIBPATH=${LIBPATH}:$HOME/server_dir:$ODBCHOME/lib:<NetezzaInstallationDir>/
lib64; export LIBPATH
```

- Using a C shell:

```
$ setenv LIBPATH ${LIBPATH}:$HOME/server_dir:$ODBCHOME/
lib:<NetezzaInstallationDir>/lib64
```

4. Edit the existing odbc.ini file or copy the odbc.ini file to the home directory and edit it.

   This file exists in $ODBCHOME directory.

   ```
   $ cp $ODBCHOME/odbc.ini $HOME/.odbc.ini
   ```

   Add an entry for the Netezza data source under the section [ODBC Data Sources] and configure the data source.

   For example:

   ```
   [NZSQL]
   Driver = /export/home/appsqa/thirdparty/netezza/lib64/libnzodbc.so
   Description = NetezzaSQL ODBC
   Servername = netezza1.informatica.com
   Port = 5480
   Database = infa
   Username = admin
   Password = password
   Debuglogging = true
   StripCRLF = false
   PreFetch = 256
   Protocol = 7.0
   ReadOnly = false
   ShowSystemTables = false
   Socket = 16384
   DateFormat = 1
   TranslationDLL =
   TranslationName =
   TranslationOption =
   NumericAsChar = false
   ```

   For more information about Netezza connectivity, see the Netezza ODBC driver documentation.

5. Verify that the last entry in the odbc.ini file is InstallDir and set it to the ODBC installation directory.

   For example:

   ```
   InstallDir=<Informatica install directory>/<ODBCHOME directory>
   ```

6. Edit the .cshrc or .profile file to include the complete set of shell commands.
7. Restart the Informatica services.

Connecting to an Oracle Database

For native connectivity, install the version of Oracle client appropriate for the Oracle database server version. To ensure compatibility between Informatica and databases, use the appropriate database client libraries.

You must install compatible versions of the Oracle client and Oracle database server. You must also install the same version of the Oracle client on all machines that require it. To verify compatibility, contact Oracle.

Configuring Native Connectivity

You can configure native connectivity to an Oracle database to increase performance.

The following steps provide a guideline for configuring native connectivity through Oracle Net Services or Net8. For specific instructions, see the database documentation.

1. To configure connectivity for the Data Integration Service process, log in to the machine as a user who can start the server process.
2. Set the ORACLE_HOME, NLS_LANG, TNS_ADMIN, and PATH environment variables.

   ORACLE_HOME. Set the variable to the Oracle client installation directory. For example, if the client is installed in the /HOME2/oracle directory. set the variable as follows:

   Using a Bourne shell:

   ```
   $ ORACLE_HOME=/HOME2/oracle; export ORACLE_HOME
   ```

   Using a C shell:

   ```
   $ setenv ORACLE_HOME /HOME2/oracle
   ```

   NLS_LANG. Set the variable to the locale (language, territory, and character set) you want the database client and server to use with the login. The value of this variable depends on the configuration. For example, if the value is american_america.UTF8, set the variable as follows:

   Using a Bourne shell:

   ```
   $ NLS_LANG=american_america.UTF8; export NLS_LANG
   ```

   Using a C shell:

   ```
   $ NLS_LANG american_america.UTF8
   ```

   To determine the value of this variable, contact the administrator.

   ORA_SDTZ. To set the default session time zone when the Data Integration Service reads or writes the Timestamp with Local Time Zone data, specify the ORA_SDTZ environment variable.

   You can set the ORA_SDTZ environment variable to any of the following values:

   - Operating system local time zone ('OS_TZ')
   - Database time zone ('DB_TZ')
   - Absolute offset from UTC (for example, '-05:00')
   - Time zone region name (for example, 'America/Los_Angeles')

   You can set the environment variable at the machine where Informatica server runs.

   TNS_ADMIN. If the tnsnames.ora file is not in the same location as the Oracle client installation location, set the TNS_ADMIN environment variable to the directory where the tnsnames.ora file resides. For example, if the file is in the /HOME2/oracle/files directory, set the variable as follows:

   Using a Bourne shell:

   ```
   $ TNS_ADMIN=$HOME2/oracle/files; export TNS_ADMIN
   ```

   Using a C shell:

   ```
   $ setenv TNS_ADMIN=$HOME2/oracle/files
   ```

   Note: By default, the `tnsnames.ora` file is stored in the following directory: `$ORACLE_HOME/network/admin`.

   PATH. To run the Oracle command line programs, set the variable to include the Oracle bin directory.

   Using a Bourne shell:

   ```
   $ PATH=${PATH}:$ORACLE_HOME/bin; export PATH
   ```

   Using a C shell:

   ```
   $ setenv PATH ${PATH}:ORACLE_HOME/bin
   ```

3. Set the shared library environment variable.

   The Oracle client software contains a number of shared library components that the Data Integration Service processes load dynamically. To locate the shared libraries during run time, set the shared library environment variable.

   The shared library path must also include the Informatica installation directory `(server_dir)`.

   Set the shared library environment variable to LD_LIBRARY_PATH.

   For example, use the following syntax:

   - Using a Bourne shell:

     ```
     $ LD_LIBRARY_PATH=${LD_LIBRARY_PATH}:$HOME/server_dir:$ORACLE_HOME/lib; export
     LD_LIBRARY_PATH
     ```

   - Using a C shell:

     ```
     $ setenv LD_LIBRARY_PATH ${LD_LIBRARY_PATH}:$HOME/server_dir:$ORACLE_HOME/lib
     ```

4. Edit the .cshrc or .profile to include the complete set of shell commands. Save the file and either log out and log in again, or run the source command.

   Using a Bourne shell:

   ```
   $ source .profile
   ```

   Using a C shell:

   ```
   $ source .cshrc
   ```

5. Verify that the Oracle client is configured to access the database.

   Use the SQL*Net Easy Configuration Utility or copy an existing `tnsnames.ora` file to the home directory and modify it.

   The `tnsnames.ora` file is stored in the following directory: `$ORACLE_HOME/network/admin`.

   Enter the correct syntax for the Oracle connect string, typically `databasename.world`.

   Here is a sample `tnsnames.ora` file. Enter the information for the database.

   ```
   mydatabase.world =
      (DESCRIPTION
      (ADDRESS_LIST =
           (ADDRESS =
              (COMMUNITY = mycompany.world
              (PROTOCOL = TCP)
              (Host = mymachine)
              (Port = 1521)
           )
      )
      (CONNECT_DATA =
         (SID = MYORA7)
         (GLOBAL_NAMES = mydatabase.world)
   ```

6. Verify that you can connect to the Oracle database.

   To connect to the Oracle database, launch SQL*Plus and enter the connectivity information. If you fail to connect to the database, verify that you correctly entered all of the connectivity information.

   Enter the user name and connect string as defined in the `tnsnames.ora` file.

Connecting to a Teradata Database

Install and configure native client software on the machines where the Data Integration Service process runs. To ensure compatibility between Informatica and databases, use the appropriate database client libraries.

Install the Teradata client, the Teradata ODBC driver, and any other Teradata client software that you might need on the machine where the Data Integration Service runs. You must also configure ODBC connectivity.

Note: Based on a recommendation from Teradata, Informatica uses ODBC to connect to Teradata. ODBC is a native interface for Teradata.

Configuring ODBC Connectivity

You can configure ODBC connectivity to a Teradata database.

The following steps provide a guideline for configuring ODBC connectivity. For specific instructions, see the database documentation.

1. To configure connectivity for the integration service process, log in to the machine as a user who can start a service process.
2. Set the TERADATA_HOME, ODBCHOME, and PATH environment variables.

   TERADATA_HOME. Set the variable to the Teradata driver installation directory. The defaults are as follows:

   Using a Bourne shell:

   ```
   $ TERADATA_HOME=/opt/teradata/client/<version>; export TERADATA_HOME
   ```

   Using a C shell:

   ```
   $ setenv TERADATA_HOME /opt/teradata/client/<version>
   ```

   ODBCHOME. Set the variable to the ODBC installation directory. For example:

   Using a Bourne shell:

   ```
   $ ODBCHOME=$INFA_HOME/ODBC<version>; export ODBCHOME
   ```

   Using a C shell:

   ```
   $ setenv ODBCHOME $INFA_HOME/ODBC<version>
   ```

   PATH. To run the *ddtestlib* utility, to verify that the DataDirect ODBC driver manager can load the driver files, set the variable as follows:

   Using a Bourne shell:

   ```
   PATH="${PATH}:$ODBCHOME/bin:$TERADATA_HOME/bin"
   ```

   Using a C shell:

   ```
   $ setenv PATH ${PATH}:$ODBCHOME/bin:$TERADATA_HOME/bin
   ```

3. Set the shared library environment variable.

   The Teradata software contains multiple shared library components that the integration service process loads dynamically. Set the shared library environment variable so that the services can find the shared libraries at run time.

   The shared library path must also include installation directory of the Informatica service *(server_dir)*.

   Set the shared library environment variable based on the operating system.

The following table describes the shared library variables for each operating system:

Operating System	Variable
Solaris	LD_LIBRARY_PATH
Linux	LD_LIBRARY_PATH
AIX	LIBPATH

For example, use the following syntax for Solaris and Linux:

- Using a Bourne shell:

```
$ LD_LIBRARY_PATH="${LD_LIBRARY_PATH}:$HOME/server_dir:$ODBCHOME/lib:

$TERADATA_HOME/lib64:$TERADATA_HOME/odbc_64/lib";

export LD_LIBRARY_PATH
```

- Using a C shell:

```
$ setenv LD_LIBRARY_PATH "${LD_LIBRARY_PATH}:$HOME/server_dir:$ODBCHOME/lib:
$TERADATA_HOME/lib64:

$TERADATA_HOME/odbc_64/lib"
```

For AIX

- Using a Bourne shell:

```
$ LIBPATH=${LIBPATH}:$HOME/server_dir:$ODBCHOME/lib:$TERADATA_HOME/
lib64:$TERADATA_HOME/odbc_64/lib; export LIBPATH
```

- Using a C shell:

```
$ setenv LIBPATH ${LIBPATH}:$HOME/server_dir:$ODBCHOME/lib:$TERADATA_HOME/lib64:

$TERADATA_HOME/odbc_64/lib
```

4. Edit the existing odbc.ini file or copy the odbc.ini file to the home directory and edit it.

   This file exists in $ODBCHOME directory.

   ```
   $ cp $ODBCHOME/odbc.ini $HOME/.odbc.ini
   ```

   Add an entry for the Teradata data source under the section [ODBC Data Sources] and configure the data source.

   For example:

   ```
   MY_TERADATA_SOURCE=Teradata Driver
   [MY_TERADATA_SOURCE]
   Driver=/u01/app/teradata/td-tuf611/odbc/drivers/tdata.so
   Description=NCR 3600 running Teradata V1R5.2
   DBCName=208.199.59.208
   DateTimeFormat=AAA
   SessionMode=ANSI
   DefaultDatabase=
   Username=
   Password=
   ```

5. Set the DateTimeFormat to AAA in the Teradata data ODBC configuration.
6. Optionally, set the SessionMode to ANSI. When you use ANSI session mode, Teradata does not roll back the transaction when it encounters a row error.

   If you choose Teradata session mode, Teradata rolls back the transaction when it encounters a row error. In Teradata mode, the integration service process cannot detect the rollback, and does not report this in the session log.

7. To configure connection to a single Teradata database, enter the DefaultDatabase name. To create a single connection to the default database, enter the user name and password. To connect to multiple databases, using the same ODBC DSN, leave the DefaultDatabase field empty.

   For more information about Teradata connectivity, see the Teradata ODBC driver documentation.

8. Verify that the last entry in the odbc.ini is InstallDir and set it to the odbc installation directory.

   For example:

   ```
   InstallDir=<Informatica installation directory>/ODBC<version>
   ```

9. Edit the .cshrc or .profile to include the complete set of shell commands.
10. Save the file and either log out and log in again, or run the source command.

    Using a Bourne shell:

    ```
    $ source .profile
    ```

    Using a C shell:

    ```
    $ source .cshrc
    ```

11. For each data source you use, make a note of the file name under the Driver=<parameter> in the data source entry in odbc.ini. Use the *ddtestlib* utility to verify that the DataDirect ODBC driver manager can load the driver file.

    For example, if you have the driver entry:

    ```
    Driver=/u01/app/teradata/td-tuf611/odbc/drivers/tdata.so
    ```

    run the following command:

    ```
    ddtestlib /u01/app/teradata/td-tuf611/odbc/drivers/tdata.so
    ```

12. Test the connection using BTEQ or another Teradata client tool.

Connecting to an ODBC Data Source

Install and configure native client software on the machine where the Data Integration Service run. Also install and configure any underlying client access software required by the ODBC driver. To ensure compatibility between Informatica and the databases, use the appropriate database client libraries.

The Informatica installation includes DataDirect ODBC drivers. If the odbc.ini file contains connections that use earlier versions of the ODBC driver, update the connection information to use the new drivers. Use the System DSN to specify an ODBC data source.

1. On the machine where the application service runs, log in as a user who can start a service process.
2. Set the ODBCHOME and PATH environment variables.

   ODBCHOME. Set to the DataDirect ODBC installation directory. For example, if the install directory is `/export/home/Informatica/10.0.0/ODBC7.1`.

   Using a Bourne shell:

   ```
   $ ODBCHOME=/export/home/Informatica/10.0.0/ODBC7.1; export ODBCHOME
   ```

   Using a C shell:

   ```
   $ setenv ODBCHOME /export/home/Informatica/10.0.0/ODBC7.1
   ```

   PATH. To run the ODBC command line programs, like *ddtestlib*, set the variable to include the odbc bin directory.

Using a Bourne shell:

```
$ PATH=${PATH}:$ODBCHOME/bin; export PATH
```

Using a C shell:

```
$ setenv PATH ${PATH}:$ODBCHOME/bin
```

Run the *ddtestlib* utility to verify that the DataDirect ODBC driver manager can load the driver files.

3. Set the shared library environment variable.

   The ODBC software contains a number of shared library components that the service processes load dynamically. Set the shared library environment variable so that the services can find the shared libraries at run time.

   The shared library path must also include the Informatica installation directory *(server_dir)*.

   Set the shared library environment variable based on the operating system.

   The following table describes the shared library variables for each operating system:

Operating System	Variable
Linux	LD_LIBRARY_PATH

   For example, use the following syntax for Linux:

   - Using a Bourne shell:

     ```
     $ LD_LIBRARY_PATH=${LD_LIBRARY_PATH}:$HOME/server_dir:$ODBCHOME/lib; export
     LD_LIBRARY_PATH
     ```

   - Using a C shell:

     ```
     $ setenv LD_LIBRARY_PATH $HOME/server_dir:$ODBCHOME:${LD_LIBRARY_PATH}
     ```

4. Edit the existing odbc.ini file or copy the odbc.ini file to the home directory and edit it.

   This file exists in $ODBCHOME directory.

   ```
   $ cp $ODBCHOME/odbc.ini $HOME/.odbc.ini
   ```

   Add an entry for the ODBC data source under the section [ODBC Data Sources] and configure the data source.

   For example:

   ```
   MY_MSSQLSERVER_ODBC_SOURCE=<Driver name or data source description>
   [MY_SQLSERVER_ODBC_SOURCE]
   Driver=<path to ODBC drivers>
   Description=DataDirect 7.1 SQL Server Wire Protocol
   Database=<SQLServer_database_name>
   LogonID=<username>
   Password=<password>
   Address=<TCP/IP address>,<port number>
   QuoteId=No
   AnsiNPW=No
   ApplicationsUsingThreads=1
   ```

   This file might already exist if you have configured one or more ODBC data sources.

5. Verify that the last entry in the odbc.ini is InstallDir and set it to the odbc installation directory.

   For example:

   ```
   InstallDir=/export/home/Informatica/10.0.0/ODBC7.1
   ```

6. If you use the odbc.ini file in the home directory, set the ODBCINI environment variable.

   Using a Bourne shell:

   ```
   $ ODBCINI=/$HOME/.odbc.ini; export ODBCINI
   ```

Using a C shell:

```
$ setenv ODBCINI $HOME/.odbc.ini
```

7. Edit the .cshrc or .profile to include the complete set of shell commands. Save the file and either log out and log in again, or run the source command.

   Using a Bourne shell:

   ```
   $ source .profile
   ```

   Using a C shell:

   ```
   $ source .cshrc
   ```

8. Use the *ddtestlib* utility to verify that the DataDirect ODBC driver manager can load the driver file you specified for the data source in the odbc.ini file.

   For example, if you have the driver entry:

   ```
   Driver = /export/home/Informatica/10.0.0/ODBC7.1/lib/DWxxxxnn.so
   ```

   run the following command:

   ```
   ddtestlib /export/home/Informatica/10.0.0/ODBC7.1/lib/DWxxxxnn.so
   ```

9. Install and configure any underlying client access software needed by the ODBC driver.

   Note: While some ODBC drivers are self-contained and have all information inside the .odbc.ini file, most are not. For example, if you want to use an ODBC driver to access Sybase IQ, you must install the Sybase IQ network client software and set the appropriate environment variables.

   To use the Informatica ODBC drivers (DWxxxxnn.so), manually set the PATH and shared library path environment variables. Alternatively, run the odbc.sh or odbc.csh script in the $ODBCHOME folder. This script will set the required PATH and shared library path environment variables for the ODBC drivers provided by Informatica.

Connecting to a JDBC Data Source

To enable the the Data Integration Service to write to relational targets, download JDBC driver .jar files to the Data Integration Service host and to all client machines that run mappings that have relational targets.

Obtain the driver .jar file from the database vendor. For example, to access an Oracle database, download the file `ojdbc.jar` from the Oracle website.

1. Place the JDBC driver .jar file in the following directory on the Data Integration Service machine `<Informatica installation directory>/externaljdbcjars`. Then recycle the Data Integration Service.
2. Place the JDBC driver .jar file in the following directory on machines that host the Developer tool: `<Informatica installation directory>/clients/externaljdbcjars`. Then recycle the Developer tool.

Sample odbc.ini File

The following sample shows the entries for the ODBC drivers in the ODBC.ini file:

```
[ODBC Data Sources]
SQL Server Legacy Wire Protocol=DataDirect 7.1 SQL Server Legacy Wire Protocol
DB2 Wire Protocol=DataDirect 7.1 DB2 Wire Protocol
```

```
Informix Wire Protocol=DataDirect 7.1 Informix Wire Protocol
Oracle Wire Protocol=DataDirect 8.0 Oracle Wire Protocol
Sybase Wire Protocol=DataDirect 7.1 Sybase Wire Protocol
SQL Server Wire Protocol=DataDirect 8.0 SQL Server Wire Protocol
MySQL Wire Protocol=DataDirect 7.1 MySQL Wire Protocol
PostgreSQL Wire Protocol=DataDirect 7.1 PostgreSQL Wire Protocol
Greenplum Wire Protocol=DataDirect 7.1 Greenplum Wire Protocol

[ODBC]
IANAAppCodePage=4
InstallDir=/<Informatica installation directory>/ODBC7.1
Trace=0
TraceFile=odbctrace.out
TraceDll=/<Informatica installation directory>/ODBC7.1/lib/DWtrc27.so

[DB2 Wire Protocol]
Driver=/<Informatica installation directory>/ODBC7.1/lib/DWdb227.so
Description=DataDirect 7.1 DB2 Wire Protocol
AccountingInfo=
AddStringToCreateTable=
AlternateID=
AlternateServers=
ApplicationName=
ApplicationUsingThreads=1
AuthenticationMethod=0
BulkBinaryThreshold=32
BulkCharacterThreshold=-1
BulkLoadBatchSize=1024
BulkLoadFieldDelimiter=
BulkLoadRecordDelimiter=
CatalogSchema=
CharsetFor65535=0
ClientHostName=
ClientUser=
#Collection applies to z/OS and iSeries only
Collection=
ConcurrentAccessResolution=0
ConnectionReset=0
ConnectionRetryCount=0
ConnectionRetryDelay=3
CurrentFuncPath=
#Database applies to DB2 UDB only
Database=<database_name>
DefaultIsolationLevel=1
DynamicSections=1000
EnableBulkLoad=0
EncryptionMethod=0
FailoverGranularity=0
FailoverMode=0
FailoverPreconnect=0
GrantAuthid=PUBLIC
GrantExecute=1
GSSClient=native
HostNameInCertificate=
IpAddress=<DB2_server_host>
KeyPassword=
KeyStore=
KeyStorePassword=
LoadBalanceTimeout=0
LoadBalancing=0
#Location applies to z/OS and iSeries only
Location=<location_name>
LogonID=
MaxPoolSize=100
MinPoolSize=0
Password=
PackageCollection=NULLID
PackageNamePrefix=DD
PackageOwner=
Pooling=0
ProgramID=
```

```
QueryTimeout=0
ReportCodePageConversionErrors=0
TcpPort=50000
TrustStore=
TrustStorePassword=
UseCurrentSchema=0
ValidateServerCertificate=1
WithHold=1
XMLDescribeType=-10

[Informix Wire Protocol]
Driver=/<Informatica installation directory>/ODBC7.1/lib/DWifcl27.so
Description=DataDirect 7.1 Informix Wire Protocol
AlternateServers=
ApplicationUsingThreads=1
CancelDetectInterval=0
ConnectionRetryCount=0
ConnectionRetryDelay=3
Database=<database_name>
HostName=<Informix_host>
LoadBalancing=0
LogonID=
Password=
PortNumber=<Informix_server_port>
ServerName=<Informix_server>
TrimBlankFromIndexName=1
UseDelimitedIdentifiers=0

[Oracle Wire Protocol]
Driver=/<Informatica installation directory>/ODBC7.1/lib/DWora28.so
Description=DataDirect 8.0 Oracle Wire Protocol
AlternateServers=
ApplicationUsingThreads=1
AccountingInfo=
Action=
ApplicationName=
ArraySize=60000
AuthenticationMethod=1
BulkBinaryThreshold=32
BulkCharacterThreshold=-1
BulkLoadBatchSize=1024
BulkLoadFieldDelimiter=
BulkLoadRecordDelimiter=
CachedCursorLimit=32
CachedDescLimit=0
CatalogIncludesSynonyms=1
CatalogOptions=0
ClientHostName=
ClientID=
ClientUser=
ConnectionReset=0
ConnectionRetryCount=0
ConnectionRetryDelay=3
DataIntegrityLevel=0
DataIntegrityTypes=MD5,SHA1
DefaultLongDataBuffLen=1024
DescribeAtPrepare=0
EditionName=
EnableBulkLoad=0
EnableDescribeParam=0
EnableNcharSupport=0
EnableScrollableCursors=1
EnableStaticCursorsForLongData=0
EnableTimestampWithTimeZone=0
EncryptionLevel=0
EncryptionMethod=0
EncryptionTypes=AES128,AES192,AES256,DES,3DES112,3DES168,RC4_40,RC4_56,RC4_128,
RC4_256
FailoverGranularity=0
FailoverMode=0
FailoverPreconnect=0
```

```
FetchTSWTZasTimestamp=0
GSSClient=native
HostName=<Oracle_server>
HostNameInCertificate=
InitializationString=
KeyPassword=
KeyStore=
KeyStorePassword=
LoadBalanceTimeout=0
LoadBalancing=0
LocalTimeZoneOffset=
LockTimeOut=-1
LoginTimeout=15
LogonID=
MaxPoolSize=100
MinPoolSize=0
Module=
Password=
Pooling=0
PortNumber=<Oracle_server_port>
ProcedureRetResults=0
ProgramID=
QueryTimeout=0
ReportCodePageConversionErrors=0
ReportRecycleBin=0
ServerName=<server_name in tnsnames.ora>
ServerType=0
ServiceName=
SID=<Oracle_System_Identifier>
TimestampeEscapeMapping=0
TNSNamesFile=<tnsnames.ora_filename>
TrustStore=
TrustStorePassword=
UseCurrentSchema=1
ValidateServerCertificate=1
WireProtocolMode=2

[Sybase Wire Protocol]
Driver=/<Informatica installation directory>/ODBC7.1/lib/DWase27.so
Description=DataDirect 7.1 Sybase Wire Protocol
AlternateServers=
ApplicationName=
ApplicationUsingThreads=1
ArraySize=50
AuthenticationMethod=0
BulkBinaryThreshold=32
BulkCharacterThreshold=-1
BulkLoadBatchSize=1024
BulkLoadFieldDelimiter=
BulkLoadRecordDelimiter=
Charset=
ConnectionReset=0
ConnectionRetryCount=0
ConnectionRetryDelay=3
CursorCacheSize=1
Database=<database_name>
DefaultLongDataBuffLen=1024
EnableBulkLoad=0
EnableDescribeParam=0
EnableQuotedIdentifiers=0
EncryptionMethod=0
FailoverGranularity=0
FailoverMode=0
FailoverPreconnect=0
GSSClient=native
HostNameInCertificate=
InitializationString=
Language=
LoadBalancing=0
LoadBalanceTimeout=0
LoginTimeout=15
```

```
LogonID=
MaxPoolSize=100
MinPoolSize=0
NetworkAddress=<Sybase_host,Sybase_server_port>
OptimizePrepare=1
PacketSize=0
Password=
Pooling=0
QueryTimeout=0
RaiseErrorPositionBehavior=0
ReportCodePageConversionErrors=0
SelectMethod=0
ServicePrincipalName=
TruncateTimeTypeFractions=0
TrustStore=
TrustStorePassword=
ValidateServerCertificate=1
WorkStationID=

[SQL Server Wire Protocol]
Driver=/<Informatica installation directory>/ODBC7.1/lib/DWsqls28.so
Description=DataDirect 8.0 SQL Server Wire Protocol
AlternateServers=
AlwaysReportTriggerResults=0
AnsiNPW=1
ApplicationName=
ApplicationUsingThreads=1
AuthenticationMethod=1
BulkBinaryThreshold=32
BulkCharacterThreshold=-1
BulkLoadBatchSize=1024
BulkLoadOptions=2
ConnectionReset=0
ConnectionRetryCount=0
ConnectionRetryDelay=3
Database=<database_name>
EnableBulkLoad=0
EnableQuotedIdentifiers=0
EncryptionMethod=0
FailoverGranularity=0
FailoverMode=0
FailoverPreconnect=0
FetchTSWTZasTimestamp=0
FetchTWFSasTime=1
GSSClient=native
HostName=<SQL_Server_host>
HostNameInCertificate=
InitializationString=
Language=
LoadBalanceTimeout=0
LoadBalancing=0
LoginTimeout=15
LogonID=
MaxPoolSize=100
MinPoolSize=0
PacketSize=-1
Password=
Pooling=0
PortNumber=<SQL_Server_server_port>
QueryTimeout=0
ReportCodePageConversionErrors=0
SnapshotSerializable=0
TrustStore=
TrustStorePassword=
ValidateServerCertificate=1
WorkStationID=
XML Describe Type=-10

[MySQL Wire Protocol]
Driver=/<Informatica installation directory>/ODBC7.1/lib/DWmysql27.so
Description=DataDirect 7.1 MySQL Wire Protocol
```

```
AlternateServers=
ApplicationUsingThreads=1
ConnectionReset=0
ConnectionRetryCount=0
ConnectionRetryDelay=3
Database=<database_name>
DefaultLongDataBuffLen=1024
EnableDescribeParam=0
EncryptionMethod=0
FailoverGranularity=0
FailoverMode=0
FailoverPreconnect=0
HostName=<MySQL_host>
HostNameInCertificate=
InteractiveClient=0
LicenseNotice=You must purchase commercially licensed MySQL database software or
a MySQL Enterprise subscription in order to use the DataDirect Connect for ODBC
for MySQL Enterprise driver with MySQL software.
KeyStore=
KeyStorePassword=
LoadBalanceTimeout=0
LoadBalancing=0
LogonID=
LoginTimeout=15
MaxPoolSize=100
MinPoolSize=0
Password=
Pooling=0
PortNumber=<MySQL_server_port>
QueryTimeout=0
ReportCodepageConversionErrors=0
TreatBinaryAsChar=0
TrustStore=
TrustStorePassword=
ValidateServerCertificate=1

[PostgreSQL Wire Protocol]
Driver=/<Informatica installation directory>/ODBC7.1/lib/DWpsql27.so
Description=DataDirect 7.1 PostgreSQL Wire Protocol
AlternateServers=
ApplicationUsingThreads=1
ConnectionReset=0
ConnectionRetryCount=0
ConnectionRetryDelay=3
Database=<database_name>
DefaultLongDataBuffLen=2048
EnableDescribeParam=1
EncryptionMethod=0
ExtendedColumnMetadata=0
FailoverGranularity=0
FailoverMode=0
FailoverPreconnect=0
FetchTSWTZasTimestamp=0
FetchTWFSasTime=0
HostName=<PostgreSQL_host>
HostNameInCertificate=
InitializationString=
KeyPassword=
KeyStore=
KeyStorePassword=
LoadBalanceTimeout=0
LoadBalancing=0
LoginTimeout=15
LogonID=
MaxPoolSize=100
MinPoolSize=0
Password=
Pooling=0
PortNumber=<PostgreSQL_server_port>
QueryTimeout=0
ReportCodepageConversionErrors=0
```

```
TransactionErrorBehavior=1
TrustStore=
TrustStorePassword=
ValidateServerCertificate=1
XMLDescribeType=-10

[Greenplum Wire Protocol]
Driver=/<Informatica installation directory>/ODBC7.1/lib/DWgplm27.so
Description=DataDirect 7.1 Greenplum Wire Protocol
AlternateServers=
ApplicationUsingThreads=1
ConnectionReset=0
ConnectionRetryCount=0
ConnectionRetryDelay=3
Database=<database_name>
DefaultLongDataBuffLen=2048
EnableDescribeParam=0
EnableKeysetCursors=0
EncryptionMethod=0
ExtendedColumnMetadata=0
FailoverGranularity=0
FailoverMode=0
FailoverPreconnect=0
FetchTSWTZasTimestamp=0
FetchTWFSasTime=0
HostName=<Greenplum_host>
InitializationString=
KeyPassword=
KeysetCursorOptions=0
KeyStore=
KeyStorePassword=
LoadBalanceTimeout=0
LoadBalancing=0
LoginTimeout=15
LogonID=
MaxPoolSize=100
MinPoolSize=0
Password=
Pooling=0
PortNumber=<Greenplum_server_port>
QueryTimeout=0
ReportCodepageConversionErrors=0
TransactionErrorBehavior=1
XMLDescribeType=-10

[SQL Server Legacy Wire Protocol]
Driver=/<Informatica installation directory>/ODBC7.1/lib/DWmsss27.so
Description=DataDirect 7.1 SQL Server Legacy Wire Protocol
Address=<SQLServer_host, SQLServer_server_port>
AlternateServers=
AnsiNPW=Yes
ConnectionRetryCount=0
ConnectionRetryDelay=3
Database=<database_name>
FetchTSWTZasTimestamp=0
FetchTWFSasTime=0
LoadBalancing=0
LogonID=
Password=
QuotedId=No
ReportCodepageConversionErrors=0
SnapshotSerializable=0
```

Note: You might have to customize the DSN entries in the `ODBC.ini` file based on the third-party driver that you use. For more information about the DSN entries, see the corresponding third-party driver documentation.

APPENDIX C

Updating the DynamicSections Parameter of a DB2 Database

This appendix includes the following topics:

DynamicSections Parameter Overview

IBM DB2 packages contain the SQL statements to be executed on the database server. The DynamicSections parameter of a DB2 database determines the maximum number of executable statements that the database driver can have in a package. You can raise the value of the DynamicSections parameter to allow a larger number of executable statements in a DB2 package. To modify the DynamicSections parameter, connect to the database using a system administrator user account with BINDADD authority.

Setting the DynamicSections Parameter

Use the DataDirect Connect for JDBC utility to raise the value of the DynamicSections parameter in the DB2 database.

To use the DataDirect Connect for JDBC utility to update the DynamicSections parameter, complete the following tasks:

- Download and install the DataDirect Connect for JDBC utility.
- Run the Test for JDBC tool.

Downloading and Installng the DDconnect JDBC Utility

Download the DataDirect Connect for JDBC utility from the DataDirect download web site to a machine that has access to the DB2 database server. Extract the contents of the utility file and run the installer.

1. Go to the DataDirect download site: http://www.datadirect.com/support/product-documentation/downloads
2. Choose the Connect for JDBC driver for an IBM DB2 data source.

3. Register to download the DataDirect Connect for JDBC Utility.
4. Download the utility to a machine that has access to the DB2 database server.
5. Extract the contents of the utility file to a temporary directory.
6. In the directory where you extracted the file, run the installer.

The installation program creates a folder named testforjdbc in the installation directory.

Running the Test for JDBC Tool

After you install the DataDirect Connect for JDBC Utility, run the Test for JDBC tool to connect to the DB2 database. You must use a system administrator user account with the BINDADD authority to connect to the database.

1. In the DB2 database, set up a system adminstrator user account with the BINDADD authority.
2. In the directory where you installed the DataDirect Connect for JDBC Utility, run the Test for JDBC tool (testforjdbc).
3. On the Test for JDBC Tool window, click Press Here to Continue.
4. Click Connection > Connect to DB.
5. In the Database field, enter the following text:

   ```
   jdbc:datadirect:db2://
   HostName:PortNumber;databaseName=DatabaseName;CreateDefaultPackage=TRUE;ReplacePackag
   e=TRUE;DynamicSections=3000
   ```

   HostName is the name of the machine hosting the DB2 database server.

   PortNumber is the port number of the database.

   DatabaseName is the name of the DB2 database.
6. In the User Name and Password fields, enter the system administrator user name and password you use to connect to the DB2 database.
7. Click Connect, and then close the window.

INDEX

E

G

H

I

J

K

L

M

N

O

P

R

S

T

U

W