Big Data Primer

Version 10.2.2

If you are new to one of our Big Data products, click through this primer to get basic information about each product, and the services, tools, documentation, and resources associated with your product.

Big Data Management
Overview

With Big Data Management®, you can process large amounts of diverse and fast-changing data on a compute cluster. Integrate and transform big data without writing or maintaining Apache code. Develop mappings and profiles, and choose the environment to process the data. Big Data Management supports the native environment of the Informatica domain and non-native environments Hadoop and Databricks.

When you push processing to a compute cluster, the Data Integration Service passes the job to one of the engines on the cluster. The Hadoop environment runs the Spark and Blaze engines. The Databricks environment runs the Databricks Spark engine. You can view the status of the big data jobs and view how the big data queries are performing in the native and Hadoop environments.

You can process hierarchical, semi-structured, and unstructured data and perform stateful computations. You can create ephemeral clusters to run jobs and save resources in a cloud platform. You can also use mass ingestion to ingest or replicate large amounts of data.

Big Data Management
Tools

You'll use the following tools with Big Data Management.

Developer Tool
A thick client application to create big data mappings and profiles without writing Apache code. You import sources and targets into the Model repository and use them in the mappings and profiles. You can run the mappings and profiles in the native or non-native environment. You can also view the mapping execution plan.

Mass Ingestion Tool
A web application where you design and implement mass ingestion specifications to define how you want to ingest data. Mass ingestion specifications replace the need to manually create and run mappings. Deploy the specification to perform the ingestion job, and monitor the job to review ingestion status and job statistics.

Administrator Tool
A web application for administrators to manage the Informatica domain. Use the Administrator tool to perform tasks like configuring high availability and setting up application services. Secure the domain through Kerberos or SAML authentication. Use native or LDAP security to authorize users and groups. Use the Administrator tool to integrate the Informatica domain with the Hadoop or Databricks environment. You can also tune the run-time engines for concurrency based on your big data processing requirements

Analyst Tool
A web application to analyze, cleanse, integrate, and standardize enterprise data. You can perform tasks in the Analyst tool based on your license. You can create profiles and scorecards and run them in the native or Hadoop environment.

Monitoring Tool
A web application to view statistics and log events for mappings that run in the native or Hadoop environment. The Monitoring tool displays current and historical information about mappings run on the Blaze and Spark engines.

Big Data Management
Services

The Informatica domain includes application services to support products that run on the domain. Big Data Management uses the following application services.

Required Services:

Model Repository Service
Manages the Model repository, which is a relational database that stores the design-time metadata for projects and folders. All requests to save or access Model repository metadata go through the Model repository.

Data Integration Service
Retrieves metadata from the Model repository to run mappings, workflows, and profiles and maintains run-time metadata. The Data Integration Service can process mappings in the native environment or push the mapping for processing in the Hadoop or Databricks environment. The Data Integration Service can concurrently process mappings and jobs to optimize performance.

Metadata Access Service
Provides all connection and cluster information to the Developer tool so you can import and preview metadata stored in Hadoop. HBase, HDFS, and Hive connections use the Metadata Access Service when you import an object from a Hadoop cluster. Not applicable for the Databricks environment.

Optional Services:

Analyst Service
Runs the Analyst tool in the Informatica domain. The Analyst Service manages the connections between the service components and users that have access to the Analyst tool.

REST Operations Hub Service
Gets mapping execution statistics for big data mappings that run in the native or Hadoop environment. The REST Operations Hub Service retrieves statistics through the REST API. You can view mapping statistics, mapping advanced statistics, mapping execution steps, and mapping execution plans.

Mass Ingestion Service
Manages and validates mass ingestion specifications that you create in the Mass Ingestion tool. The Mass Ingestion Service deploys the specification to the Data Integration Service that pushes the job to the Hadoop environment.

Search Service
Performs searches in the Analyst tool. The Search Service returns search results from the profiling warehouse and the Model repository, including data objects, mapping specifications, and scorecards. The Search Service can also return additional results like related assets, business terms, and policies.

Big Data Management
Administrator Documentation

Check out the following guides to get started with Big Data Management administration.

**Big Data Suite
Installation and Configuration Guide**

Learn how to install and configure the Informatica domain and application services. Answer a series of installer prompts to complete the installation. Then, configure the domain and application services.

**Big Data Management
Integration Guide**

Learn how to integrate the Informatica domain with the Hadoop or Databricks environment after you install and configure the Informatica services.

**Big Data Management
Administrator Guide**

Understand the Big Data Management architecture. Learn how to configure and manage security between the domain and the non-native environment. Learn how to tune services and engines for big data processing, and learn how to configure the domain to connect to the cluster.

**Informatica
Administrator Guide**

Learn how to log in to the Administrator tool and understand the user interface. Learn about domain architecture and management, including nodes, services, high availability, connections, and monitoring.

**Informatica
Application Service Guide**

Read about the application services in the Informatica domain and learn how to manage each service. You can also learn about management concepts and tasks including configuration, processing behavior, architecture, and tuning.

**Informatica
Security Guide**

Learn how to administer and manage security in the Informatica domain.

Continue to User Documentation

Big Data Management
User Documentation

Check out the following guides as you start using Big Data Management.

**Big Data Management
User Guide**

Read an introduction to Big Data Management. Understand how the engines in the Hadoop and Databricks environment process data. Learn how to configure mapping and profile properties to run in a non-native environment. Learn how to process hierarchical, semi-structured, and unstructured data and perform stateful computations.

**Informatica
Developer Tool Guide**

Learn how to log in to the Developer tool and understand the user interface. Learn about Model repository objects and how to create and import them.

**Informatica
Developer Mapping Guide**

Learn mapping concepts, including how to develop, run, and administer mappings. Learn how to use mapping parameters and dynamic mappings to reuse mapping logic for different sources. Optimize mappings through tuning and partitioning.

**Informatica
Developer Transformation Guide**

Learn about transformations that you can include in a mapping. Learn configuration, guidelines, usage, and run-time behavior.

**Informatica
Developer Workflow Guide**

Understand workflow concepts so you can run mappings and other tasks in a single operation. Learn how to develop, run, and administer a workflow. Also learn how to recover a workflow if it is interrupted or an error occurs.

**Informatica
Transformation Language Reference**

Learn how to use the transformation language that the Developer tool provides. Use the SQL-like functions in mappings to transform source data.

**Big Data Management
Mass Ingestion Guide**

Learn how to log in to the Mass Ingestion tool and understand the user interface. Understand the Mass Ingestion tool login and user interface. Learn how to create, deploy, run, and monitor mass ingestion jobs.

**Informatica
Analyst Guide**

Learn how to log in to the Analyst tool and how to navigate the workspaces. Learn how to search for data objects, mapping specifications, and profiles to view and analyze them in the Discovery workspace.

Big Data Management
Resources

Interested in exploring Big Data Management? Check out the following resources.

Product Availability Matrix
The Product Availability Matrix (PAM) publishes the support for third-party products for each release. Such products include operating systems, database systems, browsers, and other products provided by third-party vendors.

Product documentation
Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases.

How-to Library articles
How-to Library (H2L) articles are short documents intended to help a user complete a specific task, use a particular connector, or work most effectively with their configuration of Informatica products. H2L articles can be found in Informatica's Knowledge Base.

Knowledge Base articles
Knowledge Base (KB) articles are short documents with specific intentions, such as answering frequently asked questions, explaining error codes, or teaching users how to configure a setting or complete a task.

Videos
A variety of YouTube videos have been created to supplement the written documentation. Videos include overviews of products, explanations of system architecture, and instructions on how to complete a task.

Community discussions and blog posts
Informatica Network provides an active platform for users to post questions and get answers from their peers or Informatica professionals. Blog posts give informal insight into Informatica and its products.

Informatica University
This program offers role-based training programs to ensure that every user in an organization gets the most out of their investment in Informatica products. Informatica University offers video tutorials, on-demand courses, instructor-led training, and custom on-site training programs.

Big Data Streaming
Overview

Use Big Data Streaming to process streams of machine, device, and social media data in real time. You can collect the streamed data, build the business logic for the data, and push the logic to the Spark engine for processing. Big Data Streaming uses Spark Structured Streaming for stream processing. Spark Structured Streaming can handle late arrival of streaming events and process streaming data based on source timestamp.

Streaming sources can include messaging systems, such as Kafka and Amazon Kinesis. Targets can include batch targets, such as Azure Data Lake Store and HDFS, or streaming targets, such as Kafka and Amazon Kinesis Firehose. You can also perform stateful computing to store and retrieve state while evaluating expressions. Some source data objects support intelligent structure modeling to process any input type that the model can parse.

Use Informatica Edge Data streaming to ingest streaming and machine data from different sources to messaging systems and batch targets. You can then use the streamed data from the messaging systems as sources in Big Data Streaming.

Big Data Streaming
Tools

You'll use the following tools with Big Data Streaming.

Administrator Tool
A web application for administrators to manage the Informatica domain. Use the Administrator tool to perform tasks like configuring high availability and setting up application services. Secure the domain through Kerberos or SAML authentication. Use native or LDAP security to authorize users and groups. Use the Administrator tool to integrate the Informatica domain with the Hadoop environment. You can also tune the run-time engines.

Developer Tool
A thick client application to create and run streaming data mappings in the Hadoop environment. You import sources and targets into the Model repository and use them in mappings. You can also view the mapping execution plan that the Data Integration Service generates to run in the Hadoop environment.

Monitoring Tool
A web application to view statistics and log events for mappings that run in the Hadoop environment. The Monitoring tool displays current and historical information about mappings run on the Spark engine.

Big Data Streaming
Services

The Informatica domain includes application services to support each product that runs on the domain. Big Data Streaming uses the following application services.

Required Services:

Model Repository Service
Manages the Model repository, which is a relational database that stores the metadata for projects and folders. All requests to save or access Model repository metadata go through the Model repository.

Metadata Access Service
Provides all connection and cluster information to the Developer tool so you can import and preview Hadoop metadata at design time. HBase, HDFS, Hive, MapR Streams, and MapR-DB connections use the Metadata Access Service when you import an object from a Hadoop cluster.

Data Integration Service
Retrieves metadata from the Model repository when you run a Developer tool mapping or workflow from the Developer tool. The Data Integration Service pushes streaming mappings to the Spark engine in the Hadoop environment.

Required Services:

Edge Data Streaming Service
Stores information about data flows, license information, and data usage. The Edge Data Streaming Service manages data flows and monitors them.

Big Data Streaming
Administrator Documentation

Check out the following guides to get started with Big Data Streaming administration.

**Big Data Suite
Installation and Configuration Guide**

Learn how to install and configure the Informatica domain and application services. Answer a series of installer prompts to complete the installation. Then, configure the domain and application services.

**Big Data Management
Integration Guide**

Learn how to integrate the Informatica domain with the Hadoop environment after you install and configure the Informatica services.

**Big Data Management
Administrator Guide**

Understand the architecture of the Spark engine in the Hadoop environment. Learn how to configure and manage security between the domain and the Hadoop environment. Learn how to tune services and the Spark engine for big data processing, and learn how to configure the domain to connect to the cluster.

**Informatica
Administrator Guide**

Learn how to log in to the Administrator tool and understand the user interface. Learn about domain architecture and management, including nodes, services, high availability, connections, and monitoring.

**Informatica
Application Service Guide**

Read about the application services in the Informatica domain and learn how to manage each service. You can also learn about management concepts and tasks including configuration, processing behavior, architecture, and tuning.

**Informatica
Security Guide**

Learn how to administer and manage security in the Informatica domain.

Continue to User Documentation

Big Data Streaming
User Documentation

Check out the following guides as you start using Big Data Streaming.

**Big Data Streaming
User Guide**

Learn how to configure and run streaming mappings on the Spark engine in a Hadoop environment. Understand the sources and targets that you can use in a streaming mapping, how to configure stateful computing, and how to implement a Window transformation to partition streamed data.

**Big Data Management
User Guide**

Read an introduction to Big Data Management. Understand how different engines process data in the Hadoop environment. Learn how to configure mapping and profile properties for Hadoop processing. Learn how to process hierarchical, semi-structured, and unstructured data and perform stateful computations.

**Informatica
Developer Tool Guide**

Learn how to log in to the Developer tool and understand the user interface. Learn about Model repository objects and how to create and import them.

**Informatica
Developer Mapping Guide**

Learn mapping concepts, including how to develop, run, and administer mappings. Learn how to use mapping parameters and dynamic mappings to reuse mapping logic for different sources. Optimize mappings through tuning and partitioning.

**Informatica
Developer Transformation Guide**

Learn about transformations that you can include in a mapping. Learn configuration, guidelines, usage, and run-time behavior.

**Informatica
Edge Data Streaming User Guide**

Learn how to create and deploy a data flow. Understand the source and target services that you can use in a data flow.

Big Data Streaming
Resources

Interested in exploring Big Data Streaming? Check out the following resources.

Product Availability Matrix

The Product Availability Matrix (PAM) publishes the support for third-party products for each release. Such products include operating systems, database systems, browsers, and other products provided by third-party vendors.

Product documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases.

How-to Library articles

How-to Library (H2L) articles are short documents intended to help a user complete a specific task, use a particular connector, or work most effectively with their configuration of Informatica products. H2L articles can be found in Informatica's Knowledge Base.

Knowledge Base articles

Knowledge Base (KB) articles are short documents with specific intentions, such as answering frequently asked questions, explaining error codes, or teaching users how to configure a setting or complete a task.

Videos

A variety of YouTube videos have been created to supplement the written documentation. Videos include overviews of products, explanations of system architecture, and instructions on how to complete a task.

Community discussions and blog posts

Informatica Network provides an active platform for users to post questions and get answers from their peers or Informatica professionals. Blog posts give informal insight into Informatica and its products.

Informatica University

This program offers role-based training programs to ensure that every user in an organization gets the most out of their investment in Informatica products. Informatica University offers video tutorials, on-demand courses, instructor-led training, and custom on-site training programs.

Big Data Quality
Overview

With Big Data Quality, you can process large amounts of diverse and fast-changing data in the Hadoop environment. Perform data quality operations such as standardization, parsing, duplicate analysis, and address validation. You can also perform data discovery to learn more about the metadata of your source data. Create mappings and profiles, and choose to process the data in the Hadoop environment or in the native environment of the Informatica domain.

If you push processing to the Hadoop environment, the Data Integration Service generates an execution plan and passes it to one of the engines in the Hadoop environment: the Blaze engine, Spark engine, or Hive engine. You can view the status of the big data processing jobs and view how the big data queries are performing.

Big Data Quality
Tools

You'll use the following tools with Big Data Quality.

Developer Tool
A thick client application to create and run data quality mapplets, mappings, and profiles in the native or Hadoop environment. You can connect reference data objects to the data quality transformations that you add to mapplets and mappings. You can also access the business rules that you create in the Analyst tool and add them to mappings.

Monitoring Tool
A web application to view statistics and log events for mappings that run in the native or Hadoop environment. The Monitoring tool displays current and historical information about mappings that run on Blaze, Spark, and Hive engines.

Administrator Tool
A web application for administrators to manage the Informatica domain. Use the Administrator tool to configure the services that enable and run big data processes, including the Model Repository Service, Content Management Service, and the Data Integration Service. Use the Administrator tool to integrate the Informatica domain with the Hadoop environment. You can also tune the run-time engines for big data processing requirements such as concurrency.

Analyst Tool
A web application to analyze and verify data quality. Define business rules that you can run directly on enterprise data. Perform data discovery to discover the metadata of source systems. Create and run profiles to discover the content and structure of data sources. Create scorecards to track and measure data quality. You can choose to run profiles in the native environment or the Hadoop environment.

Big Data Quality
Services

The Informatica domain includes application services to support each product that runs on the domain. Big Data Quality uses the following application services.

Required Services:

Model Repository Service
Manages the Model repository. The Model repository stores metadata for objects that you create in the Developer tool or the Analyst tool, including profiles, data domains, mappings, and reference tables.

Data Integration Service
Retrieves metadata from the Model repository to run mappings, workflows, and profiles. The Data Integration Service can process mappings in the native environment or push mappings for processing in the Hadoop environment.

Metadata Access Service
Provides all connection and cluster information to the Developer tool so you can import and preview Hadoop metadata. HBase, HDFS, and Hive connections use the Metadata Access Service when you import an object from a Hadoop cluster.

Content Management Service
Manages reference data and rule specifications. The Content Management Service uses the Data Integration Service to transfer data between reference tables and external data sources. The Content Management Service also manages the rule specifications that represent business rules about your data.

Optional Services:

Search Service
Performs searches in the Analyst tool. The Search Service extracts information about content from the Model repository and the profiling warehouse.

Analyst Service
Runs the Analyst tool and manages the connections between service components and the users who have access to the Analyst tool. When you create the service, associate it with a Data Integration Service and the Model Repository Service.

Big Data Quality
Administrator Documentation

Check out the following guides to get started with Big Data Quality administration.

Big Data Suite
Installation and Configuration Guide

Learn how to install and configure the Informatica domain and application services. Answer a series of installer prompts to complete the installation. Then, configure the domain and application services.

Big Data Management
Integration Guide

Learn how to integrate the Informatica domain with the Hadoop environment after you install and configure the Informatica services.

Big Data Management
Administrator Guide

Understand the architecture of the run-time engines in the Hadoop environment. Learn how to configure and manage security between the domain and the Hadoop environment. Learn how to tune services and engines for big data processing, and learn how to configure the domain to connect to the cluster.

Informatica
Administrator Guide

Learn how to log in to the Administrator tool and understand the user interface. Learn about domain architecture and management, including nodes, services, high availability, connections, and monitoring.

Informatica
Application Service Guide

Read about the application services in the Informatica domain and learn how to manage each service. You can also learn about management concepts and tasks including configuration, processing behavior, architecture, and tuning.

Informatica
Security Guide

Learn how to administer and manage security in the Informatica domain.

Continue to User Documentation

Big Data Quality
User Documentation

Check out the following guides as you start using Big Data Quality.

**Data Quality
Getting Started Guide**

Understand the Informatica domain. Learn how to use the Developer tool and the Analyst tool to perform data quality tasks.

**Big Data Management
User Guide**

Read an introduction to Big Data Management. Learn the rules and guidelines for processing on different engines in the Hadoop environment. Learn how to configure mapping and profile properties for Hadoop processing.

**Informatica
Developer Tool Guide**

Learn how to log in to the Developer tool and understand the user interface. Learn about Model repository objects and how to create and import them.

**Informatica
Developer Mapping Guide**

Learn mapping concepts, including how to develop, run, and administer mappings. Learn how to use mapping parameters and dynamic mappings to reuse mapping logic for different sources. Optimize mappings through tuning and partitioning.

**Informatica
Developer Transformation Guide**

Learn about transformations that you can include in a mapping. Learn configuration, guidelines, usage, and run-time behavior.

**Informatica
Developer Workflow Guide**

Understand workflow concepts so you can run mappings and other tasks in a single operation. Learn how to develop, run, and administer a workflow. Also learn how to recover a workflow if it is interrupted or an error occurs.

Big Data Quality
User Documentation (continued)

Data Quality
Content Guide

Learn how to install, import, and configure the accelerators, address reference data files, and identity population files distributed by Informatica.

Informatica
Analyst Tool Guide

Learn how to log in to the Analyst tool and how to navigate the workspaces. Learn about the Connections workspace to manage connections to data objects and repositories. Learn about the Design workspace to create rule specifications and reference tables.

Informatica
Data Discovery Guide

Learn how you can use profiles to analyze the content, quality, and structure of data sources and to discover the metadata of source systems. Use profiles to discover data quality issues in a data source and to understand the relationships between columns in one or more data sources. You can perform data discovery in the Developer tool or the Analyst tool.

Informatica
Rule Specification Guide

Learn how to create and use rule specifications that represent the data requirements of a business rule. Create rule specifications in the Analyst tool and add them to mappings in the Developer tool.

Informatica
Reference Data Guide

Understand the reference data objects and files that you can use in the Developer tool and the Analyst tool. Informatica transformations can analyze and update your data using reference data such as reference tables, content sets, and address reference data files.

Informatica
Transformation Language Reference

Learn how to use the transformation language that the Developer tool provides. Use the SQL-like functions in mappings to transform source data.

Big Data Quality
Resources

Interested in exploring Big Data Quality? Check out the following resources.

Product Availability Matrix

The Product Availability Matrix (PAM) publishes the support for third-party products for each release. Such products include operating systems, database systems, browsers, and other products provided by third-party vendors.

Product documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases.

How-to Library articles

How-to Library (H2L) articles are short documents intended to help a user complete a specific task, use a particular connector, or work most effectively with their configuration of Informatica products. H2L articles can be found in Informatica's Knowledge Base.

Knowledge Base articles

Knowledge Base (KB) articles are short documents with specific intentions, such as answering frequently asked questions, explaining error codes, or teaching users how to configure a setting or complete a task.

Videos

A variety of YouTube videos have been created to supplement the written documentation. Videos include overviews of products, explanations of system architecture, and instructions on how to complete a task.

Community discussions and blog posts

Informatica Network provides an active platform for users to post questions and get answers from their peers or Informatica professionals. Blog posts give informal insight into Informatica and its products.

Informatica University

This program offers role-based training programs to ensure that every user in an organization gets the most out of their investment in Informatica products. Informatica University offers video tutorials, on-demand courses, instructor-led training, and custom on-site training programs.

Enterprise Data Catalog
Overview

Enterprise Data Catalog helps business users and IT users in your enterprise realize the power of data assets by providing a unified metadata view. The view includes technical metadata, business context, user annotations, asset relationships, and data quality.

Enterprise Data Catalog maintains a searchable catalog that serves as a centralized repository to store metadata. It extracts the metadata from external sources across your enterprise, such as databases, business intelligence sources, unstructured files, and file systems.

Enterprise Data Catalog is powered by Informatica CLAIRE™, an AI-based metadata engine, which is built over a distributed Hadoop cluster that makes it possible to store large volumes of metadata.

After you extract physical and operational metadata for structured, semi-structured, and unstructured data assets, you can organize the metadata based on business concepts, enrich data assets using attributes, and view the data lineage and relationship information for each object. You can search for data assets, refine your results, and perform metadata tasks. Search features are enhanced with the CLAIRE engine to deliver intelligent suggestions.

Enterprise Data Catalog
Tools

You'll use the following tools with Enterprise Data Catalog.

Administrator Tool
A web application for administrators to manage the domain, ensure domain security, and monitor jobs. Use the Administrator tool to create the Catalog Service, Informatica Cluster Service, and other application services. You can also manage user permissions.

Catalog Administrator
A web application to perform administrative tasks for Enterprise Data Catalog. Use the Catalog Administrator to manage resources, scheduling, metadata attributes, resource connections, profile configurations, and data domains.

Developer Tool
A thick client application to create, export, and import data domains. You can use the Developer tool to import metadata, create connections, and create data objects. Also use the Developer tool to run data domain discovery on metadata sources with data domains that use reference tables.

Enterprise Data Catalog application
A web application to search and analyze large volumes of metadata, understand the lineage and impact of data assets, and identify relationships between data assets.

Enterprise Data Catalog
Services

The Informatica domain includes application services to support each product that runs on the domain. Enterprise Data Catalog uses the following application services.

Required Services:

Data Integration Service
Performs data integration tasks for the Analyst tool, the Developer tool, and external clients. The Data Integration Service receives requests from Informatica client tools to run integration, profile, data domain discovery, and column similarity discovery, and it writes results to different databases.

Catalog Service
Runs Enterprise Data Catalog application services in the Informatica domain. The Catalog Service also manages the connections between service components and the users that have access to the Enterprise Data Catalog application and the Catalog Administrator.

Model Repository Service
Manages the Model repository. When you access a Model repository object from the Catalog Administrator 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. The data domains that you create in Enterprise Data Catalog are stored in the Model repository.

Informatica Cluster Service
Runs and manages the embedded Hadoop cluster where you deploy Enterprise Data Catalog. The Informatica Cluster Service launches the required Hadoop services on the hosts where the embedded cluster runs. Required if Enterprise Data Catalog uses an embedded cluster.

Enterprise Data Catalog
Services (continued)

Optional Services:

Analyst Service
Runs the Analyst tool in the Informatica domain. The Analyst Service manages the connections between the service components and the users who log in to the Analyst tool. Enterprise Data Catalog uses the Analyst Service for business glossary synchronization with the Analyst tool.

Content Management Service
Manages reference data and provides reference data information to the Data Integration Service, the Developer tool, and the Analyst tool. Enterprise Data Catalog uses the Content Management Service if you use Informatica Developer to import data domains into Model repository.

Enterprise Data Catalog
Administrator Documentation

Check out the following guides to get started with Enterprise Data Catalog administration. See the Enterprise Data Catalog documentation on Informatica Network.

**Enterprise Data Catalog
Installation and Configuration Guide**

Learn how to install and configure Enterprise Data Catalog. Answer a series of installer prompts to complete the installation. Then, configure the domain and application services.

**Informatica
Security Guide**

Learn how to administer and manage security in the Informatica domain.

**Enterprise Data Catalog
Upgrade Guide**

Learn how to upgrade Enterprise Data Catalog on an embedded cluster or an existing cluster by following a series of steps. The supported versions are 10.1 and later.

**Enterprise Data Catalog
REST API Reference**

Learn how to customize Enterprise Data Catalog using the REST APIs. Understand how to use the APIs using samples and use cases.

Enterprise Data Catalog
User Documentation

Check out the following guide as you start using Enterprise Data Catalog. See the Enterprise Data Catalog documentation on Informatica Network.

**Enterprise Data Catalog
User Guide**

Read an overview of Enterprise Data Catalog functionality. Understand how to search within Enterprise Data Catalog and use the views to see more details about each asset.

Enterprise Data Catalog
Resources

Interested in exploring Enterprise Data Catalog? Check out the following resources.

Product Availability Matrix

The Product Availability Matrix (PAM) publishes the support for third-party products for each release. Such products include operating systems, database systems, browsers, and other products provided by third-party vendors.

Product documentation

Use Informatica Network to explore an extensive library of documentation for current and recent product releases.

How-to Library articles

How-to Library (H2L) articles are short documents intended to help a user complete a specific task, use a particular connector, or work most effectively with their configuration of Informatica products. H2L articles can be found in Informatica's Knowledge Base.

Knowledge Base articles

Knowledge Base (KB) articles are short documents with specific intentions, such as answering frequently asked questions, explaining error codes, or teaching users how to configure a setting or complete a task.

Videos

A variety of YouTube videos have been created to supplement the written documentation. Videos include overviews of products, explanations of system architecture, and instructions on how to complete a task.

Community discussions and blog posts

Informatica Network provides an active platform for users to post questions and get answers from their peers or Informatica professionals. Blog posts give informal insight into Informatica and its products.

Informatica University

This program offers role-based training programs to ensure that every user in an organization gets the most out of their investment in Informatica products. Informatica University offers video tutorials, on-demand courses, instructor-led training, and custom on-site training programs.

Enterprise Data Lake
Overview

Enterprise Data Lake is a collaborative self-service big data discovery and preparation solution. Enterprise Data Lake helps you derive more value from your Hadoop-based data lake and make data available to all users in the organization.

As a data analyst or data scientist, you can spend more time on analysis and less time finding and preparing data. You can quickly and easily find and explore trusted data assets within and outside of the data lake using semantic search and smart recommendations. Then you can transform, cleanse, and enrich data in the data lake in a self-service manner, and publish data to share knowledge with the rest of the community.

As an IT or governance staff member, Enterprise Data Lake helps you to ensure quality, visibility, and governance of data in the data lake. You can track data lineage to verify that data is coming from the right sources and going to the right targets, and can enforce appropriate security and governance policies. You can also operationalize the work done by data analysts and data scientists into a repeatable data delivery process.

Enterprise Data Lake
Tools

You'll use the following tools with Enterprise Data Lake.

Catalog Administrator
A web application to create resources in Enterprise Data Catalog. Use the Catalog Administrator to manage resources that represent an external data source or a metadata repository. Scanners attached to a resource extract metadata from the resource and store it for discovery by Enterprise Data Lake users.

Enterprise Data Lake application
A web application to search for, discover, and prepare data assets. Combine, cleanse, transform, and structure the data to prepare for analysis. After you prepare the data, you publish the transformed data back to the data lake.

Administrator Tool
A web application for administrators to manage the application services that Enterprise Data Lake uses. Use the Administrator tool to administer the Informatica domain and security and to monitor mappings during the upload and publishing processes.

Developer Tool
A thick client application that administrators use to optimize and deploy the mappings created when analysts publish prepared data in the Enterprise Data Lake application. Use the Developer tool to create or import rules that cleanse and validate data.

Enterprise Data Lake
Services

The Informatica domain includes application services to support each product that runs on the domain. Enterprise Data Lake uses the following application services.

Required Services:

Enterprise Data Lake Service
Runs the Enterprise Data Lake application in the Informatica domain. When you upload data, the Enterprise Data Lake Service connects to the HDFS system in the Hadoop cluster to temporarily stage the data. When you preview data, the Enterprise Data Lake Service connects to the Hadoop cluster to read from the data lake.

Data Preparation Service
Provides the functionality that analysts use to prepare data for publication to the data lake. The Data Preparation Service connects to a database that serves as the Data Preparation repository to store metadata.

Catalog Service
Manages the metadata catalog. When a data analyst searches for assets in the Enterprise Data Lake application, the Enterprise Data Lake Service connects to the Catalog Service to return search results from the metadata stored in the catalog.

Data Integration Service
Performs data integration tasks for Enterprise Data Lake. When an analyst uploads data or publishes prepared data, the Enterprise Data Lake Service connects to the Data Integration Service to write the data to a Hive table in the Hadoop cluster.

Model Repository Service
Manages the Model repository associated with the Enterprise Data Lake Service and the Data Integration Service. The Enterprise Data Lake Service connects to the Model Repository Service to store the project metadata and converted mappings in the Model repository. The Data Integration Service connects to the Model Repository Service to run mappings stored in the Model repository.

Enterprise Data Lake
Administrator Documentation

Check out the following guides to get started with Enterprise Data Lake administration.

**Big Data Suite
Installation and Configuration Guide**

Learn how to install and configure the Informatica domain and application services. Answer a series of installer prompts to complete the installation. Then, configure the domain and application services.

**Enterprise Data Lake
Administrator Guide**

Understand how to configure and monitor the application, set up user accounts, manage schedules, and enable users to apply rules during data preparation.

**Informatica
Administrator Guide**

Learn how to use the Administrator tool to manage services associated with Enterprise Data Lake. Learn about domain architecture and management, including nodes, services, connections, and monitoring.

**Informatica
Application Service Guide**

Read about the application services associated with Enterprise Data Lake and learn how to manage each service. You can also learn about management concepts and tasks including configuration, processing behavior, architecture, and tuning.

Enterprise Data Lake
User Documentation

Check out the following guide as you start using Enterprise Data Lake.

**Enterprise Data Lake
User Guide**

Learn how to use the Enterprise Data Lake application to search for, discover, and prepare data for analysis. Learn how to manage projects and create visualizations for data assessment.

Enterprise Data Lake
Resources

Interested in exploring Enterprise Data Lake? Check out the following resources.

Product Availability Matrix

The Product Availability Matrix (PAM) publishes the support for third-party products for each release. Such products include operating systems, database systems, browsers, and other products provided by third-party vendors.

Product documentation

Use the Informatica Documentation Portal to explore an extensive library of documentation for current and recent product releases.

How-to Library articles

How-to Library (H2L) articles are short documents intended to help a user complete a specific task, use a particular connector, or work most effectively with their configuration of Informatica products. H2L articles can be found in Informatica's Knowledge Base.

Knowledge Base articles

Knowledge Base (KB) articles are short documents with specific intentions, such as answering frequently asked questions, explaining error codes, or teaching users how to configure a setting or complete a task.

Videos

A variety of YouTube videos have been created to supplement the written documentation. Videos include overviews of products, explanations of system architecture, and instructions on how to complete a task.

Community discussions and blog posts

Informatica Network provides an active platform for users to post questions and get answers from their peers or Informatica professionals. Blog posts give informal insight into Informatica and its products.

Informatica University

This program offers role-based training programs to ensure that every user in an organization gets the most out of their investment in Informatica products. Informatica University offers video tutorials, on-demand courses, instructor-led training, and custom on-site training programs.