Informatica®

Informatica® Mass Ingestion
January 2023

Getting Started

Informatica Mass Ingestion Getting Started
January 2023

Publication Date: 2023-01-16

Table of Contents

Preface

Read *Getting Started* for information about the steps you need to perform before you begin configuring ingestion tasks in Mass Ingestion. It assumes that you have a working knowledge of Informatica Intelligent Cloud Services.

CHAPTER 1

Getting Started with Mass Ingestion

Before you configure an ingestion task, verify that all prerequisite tasks have been completed.

Step 1. Check system requirements

Check the following items:

- For Mass Ingestion Databases minimum system requirements, see *Mass Ingestion Databases* > *Mass Ingestion Databases system requirements*.
- To determine the web browsers that are compatible with Mass Ingestion and the operating systems that are supported for the Secure Agent, check the Product Availability Matrix (PAM) for Informatica Intelligent Cloud Services (IICS) on the Informatica Network at https://network.informatica.com/community/informatica-network/product-availability-matrices/
- To determine the source and target types and versions that are supported for each ingestion type, check the KB article What are the supported sources and targets for IICS Cloud Mass Ingestion service?.

Step 2. Set up an organization.

If you are the administrator, set up an organization from the **Organization** page in Administrator. An organization is a secure area within the Informatica Intelligent Cloud Services repository that stores your licenses, user accounts, ingestion tasks, and information about jobs and security.

Then, configure users, user groups, and user role permissions for the organization.

If your organization has the Organization Hierarchy license, you can also create one or more sub-organizations within your organization. You can create sub-organizations to represent different business environments within your company. For example, you might create separate sub-organizations to represent your development, testing, and production environments.

For more information, see "Organizations" in the Administrator help.

Step 3. Download and install a Secure Agent

On the **Runtime Environments** page in Administrator, download a Secure Agent and install it. A Secure Agent is a lightweight program that runs tasks and enables secure communication across the firewall between your organization and Informatica Intelligent Cloud Services. Mass Ingestion does not support the Hosted Agent or serverless runtime environments for any ingestion type.

When you download and install a Secure Agent, a *Secure Agent group*, also called a runtime environment, is created. A Secure Agent group can contain one Secure Agent or multiple agents if your licensing allows it. If you use a Secure Agent group with multiple agents, an available agent will be picked from the active agent list to run your jobs. Mass Ingestion Applications and Mass Ingestion Databases support Secure Agent groups with multiple agents only for initial load jobs. For incremental load and combined initial and incremental load jobs, you must use a single Secure Agent.

Enabling services and connectors for the Secure Agent group downloads components and packages based on your selections and creates an associated runtime environment. For more information, see "Secure Agent Groups" in the Administrator help.

Step 4. Configure the runtime environment

On the **Runtime Environments** page in Administrator, select your runtime environment.

A runtime environment is the execution platform for running tasks. You must have at least one runtime environment in your organization for users to be able to run tasks. If you create another runtime environment, you must add an unassigned Secure Agent to it.

Under **System Configuration Details**, configure properties for the CMI Streaming Agent, Database Ingestion, or Mass Ingestion (file ingestion) service. For more information, see "Secure Agent services" on page 13.

Step 5. Configure connections

On the **Connections** page, configure connection properties for the source and target connectors that you want to use in ingestion tasks. For more information, see *Connectors and Connections* > *Mass Ingestion connectors* and *Connectors and Connections* > *Mass Ingestion connection properties*.

Step 6. Create your project

From the **Explore** page in Mass Ingestion service, create projects and project folders to organize your ingestion tasks. A project can contain multiple subfolders. See "Creating projects and project folders" on page 25.

Installing Secure Agents

You can install Secure Agents on Windows or Linux.

Secure Agent installation on Windows

On Windows, the Secure Agent runs as a Windows service. When you install the Secure Agent, you also install the Informatica Cloud Secure Agent Manager.

By default, the Secure Agent starts when you start Windows. You can stop and restart the Secure Agent using the Secure Agent Manager or Windows Services. If you install the Secure Agent on a different volume than you use to run the installation program, you must start and stop the Secure Agent from Windows Services.

You can also use the Secure Agent Manager to check the Secure Agent status and configure proxy information. The Secure Agent works with BASIC and DIGEST proxy authentication.

You can launch the Secure Agent Manager from the Start menu or desktop icon. When you close the Secure Agent Manager, it minimizes to the Windows taskbar notification area for quick access.

When you install a Secure Agent, you perform the following tasks:

1. Verify that the machine meets the minimum requirements.
2. Download the Secure Agent installer files.
3. Install and register the Secure Agent.

Secure Agent requirements on Windows

You can install the Secure Agent on any machine that has internet connectivity and can access Informatica Intelligent Cloud Services.

Verify the following requirements before you install the Secure Agent on Windows:

- The Secure Agent machine uses a supported operating system. For the list of supported operating systems for the Secure Agent, see the Product Availability Matrix (PAM) for Informatica Intelligent Cloud Services on the Knowledge Base.
- The Secure Agent machine has the Microsoft Visual C++ 2015 Redistributable.
- The Secure Agent machine has at least 4 CPU cores, 16 GB RAM, and at least 5 GB of free disk space.
- The Secure Agent machine is on a volume with at least 250GB disk space, with at least 5 GB free space or three times the size of the Secure Agent installation, whichever is greater.
- The account you use to install the Secure Agent has access to all remote directories that contain flat source or target files.
- No other Secure Agent is installed on the machine. If another Secure Agent is installed on the machine, uninstall it first.

For more information about Secure Agent requirements, see this article:
https://knowledge.informatica.com/s/article/526096

Configure the firewall

If your organization uses a protective firewall, include the Informatica Intelligent Cloud Services domain name or IP address ranges in the list of approved domain names or IP addresses. To ensure that the Secure Agent can perform all necessary tasks through the firewall, enable the port that the Secure Agent uses.

The Secure Agent uses port 443 (HTTPS) to connect to the internet. Configure your firewall to allow traffic to pass over port 443.

The allowlists of domains and IP addresses can vary according to your data center, which is also called a POD (Point of Deployment). You can identify your POD through the URL that appears when you open any service in Informatica Intelligent Cloud Services. The first few characters of the URL string identify the POD. For example, if the URL starts with `usw3.dm-us.informaticacloud.com`, your POD is USW3.

You can find the allowlists of Informatica Intelligent Cloud Services domains and IP addresses for different PODs in this Knowledge Base article on Informatica Network or by clicking the link at the top of the **Runtime Environments** page in Administrator.

Secure Agent permissions on Windows

A Secure Agent requires certain permissions to transfer data between sources and targets.

When you install a Secure Agent on Windows, the Secure Agent must be part of the local Administrators group.

Configure Windows settings

Before you use the Secure Agent on Windows, configure proxy settings and a Windows Secure Agent service login.

You can configure proxy settings in Secure Agent Manager. Configure a login for the Windows Secure Agent service on Windows.

Note: If you use the Secure Agent for Informatica Cloud Data Wizard, you do not need to configure proxy settings or a Windows service login for the Secure Agent.

Downloading and installing the Secure Agent on Windows

To install the Secure Agent on a Windows machine, you must download and run the Secure Agent installation program and then register the agent.

Secure Agent registration requires an install token. To get the install token, copy the token when you download the agent or use the **Generate Install Token** option in Administrator. The token expires after 24 hours.

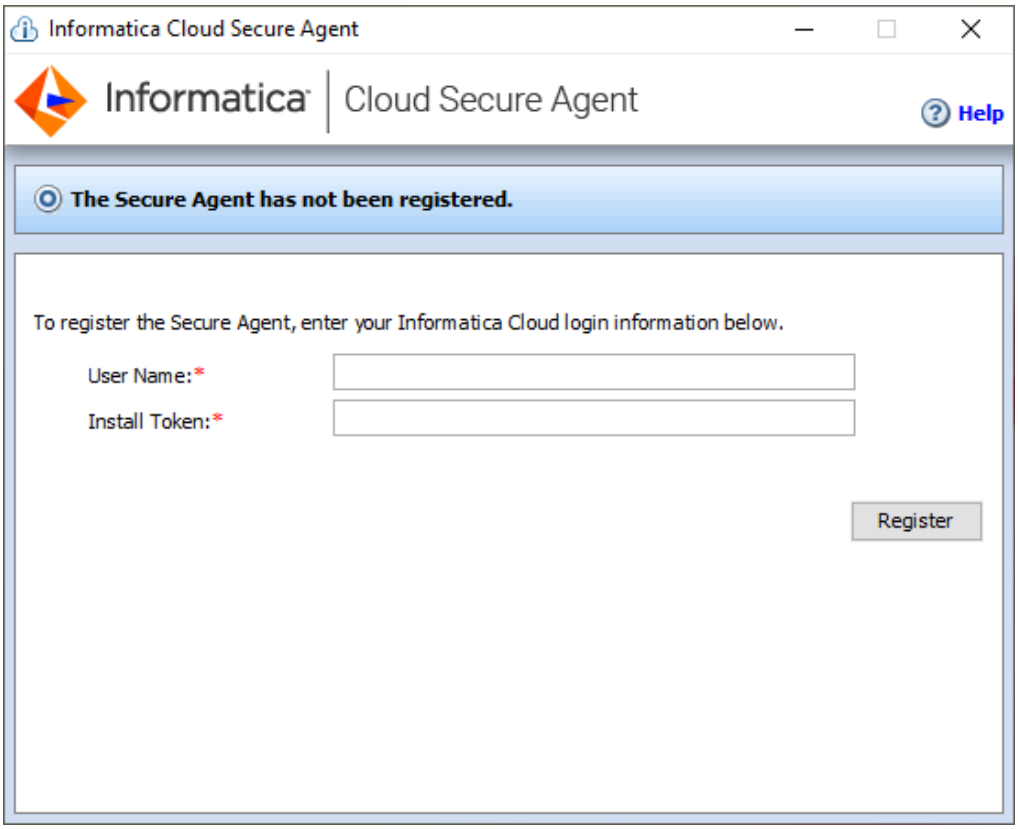
Before you download and install the Secure Agent, verify that no other Secure Agent is installed on the machine. If there is, you must uninstall it.

Tip: To verify the checksum of the Secure Agent installation program, use the agent REST API version 2 resource. For more information about the agent resource, see *REST API Reference*.

1. Open Administrator and select **Runtime Environments**.
2. On the **Runtime Environments** page, click **Download Secure Agent**.
3. Select the Windows 64-bit operating system platform, copy the install token, and then click **Download**.

   The installation program is downloaded to your machine. The name of the installation program is `agent64_install_ng_ext.<agent core version>.exe`.
4. Run the installation program as an Administrator:
   a. Specify the Secure Agent installation directory, and click **Next**.
   b. Click **Install** to install the agent.

   The Secure Agent Manager opens and prompts you to register the agent as shown in the following image:



5. If you did not copy the install token when you downloaded the agent, click **Generate Install Token** on the **Runtime Environments** page in Administrator, and copy the token.

6. In the Secure Agent Manager, enter the following information, and then click **Register**:

Option	Description
User Name	User name that you use to access Informatica Intelligent Cloud Services.
Install Token	Token that you copied.

The Secure Agent Manager displays the status of the Secure Agent. It takes a minute for all of the services to start.

7. If your organization uses an outgoing proxy server to connect to the internet, enter the proxy server information.
8. Close the Secure Agent Manager.

The Secure Agent Manager minimizes to the taskbar and continues to run as a service until stopped.

Configure the proxy settings on Windows

If your organization uses an outgoing proxy server to connect to the internet, the Secure Agent connects to Informatica Intelligent Cloud Services through the proxy server. The Secure Agent installer configures the proxy server settings for the Secure Agent based on settings configured in the browser. You can change the proxy server settings through the Secure Agent Manager.

Contact your network administrator for the correct proxy settings.

1. In the Secure Agent Manager, click **Proxy**.
2. To enter proxy server settings, click **Use a Proxy Server**.
3. Enter the following information:

Field	Description
Proxy Host	Required. Host name of the outgoing proxy server that the Secure Agent uses.
Proxy Port	Required. Port number of the outgoing proxy server.
User Name	User name to connect to the outgoing proxy server.
Password	Password to connect to the outgoing proxy server.

4. Click **OK**.

The Secure Agent Manager restarts the Secure Agent to apply the settings.

Configure a login for a Windows Secure Agent Service

On Windows, configure a network login for the Secure Agent service. The Secure Agent can access the network with the privileges and permissions associated with the login.

Configure a login for the Secure Agent machine to allow the Secure Agent to access directories to configure and run tasks. When you configure connections, configure tasks, and run tasks that use flat file or FTP/SFTP connection types, the Secure Agent might require read and write permissions on the related directories.

For example, to browse to a directory to configure a flat file or FTP/SFTP connection, the Secure Agent login might require permission to access the directory. Without a Secure Agent login with the appropriate permissions, Informatica Intelligent Cloud Services cannot display the directory in the **Browse for Directory** dialog box.

1. Go to the **Services** window from the Windows Administrative tools.
2. In the **Services** window, right-click the Informatica Cloud Secure Agent service and choose **Properties**.
3. In the **Properties** dialog box, click the **Log On** tab.
4. To configure a login, select **This Account**.
5. Enter an account and password.

   Use an account with the required privileges and permissions for the network security defined for the domain. By default, the account format is <domain name>\<user name>.
6. Click **OK**.
7. In the **Services** window, restart the Secure Agent service for the changes to take effect.

Secure Agent installation on Linux

On Linux, the Secure Agent runs as a process. You can use a shell command line to install, register, start, stop, and uninstall the Secure Agent.

You can also use the shell command line to check the Secure Agent status.

When you install a Secure Agent, you perform the following tasks:

1. Verify that the machine meets the minimum requirements.
2. Download the Secure Agent installer files.
3. Install and register the Secure Agent.

Consider the following guidelines:

- Do not install the Secure Agent as the root user. Instead, create a specific user profile to install the Secure Agent with full access to all folders from the Secure Agent installation directory.
- Do not install more than one Secure Agent on the same machine.
- Do not install the Secure Agent on the same machine that is running the Informatica PowerCenter Domain server.

For more information about Secure Agent requirements, see this article:
https://knowledge.informatica.com/s/article/526096

Secure Agent requirements on Linux

You can install the Secure Agent on any machine that has internet connectivity and can access Informatica Intelligent Cloud Services. Before you install the Secure Agent on Linux, verify the system requirements.

Verify the following requirements before you install the Secure Agent on Linux:

- Verify that the machine uses a supported operating system. For the list of supported operating systems for the Secure Agent, see the Product Availability Matrix (PAM) for Informatica Intelligent Cloud Services on the Knowledge Base.
- Verify that the machine has at least 5 GB of free disk space.
- Verify that the `libidn.x86_64` package is installed.
  If the package isn't present, install it using the following command: `sudo yum install libidn.x86_64`

  Note: The command to install the package might vary based on your Linux distribution.

- Verify that the environment variable LD_LIBRARY_PATH is set to the following location: `<Secure Agent installation directory>/apps/Data_Integration_Server/<version>/ICS/main/bin/rdtm`
- The account that you use to install the Secure Agent must have access to all remote directories that contain flat source or target files.
- If you use PowerCenter, install the Secure Agent using a different user account than the account you used to install PowerCenter.
  Informatica Intelligent Cloud Services and PowerCenter use some common environment variables. If the environment variables are not set correctly for Informatica Intelligent Cloud Services, your jobs might fail at run time.

For more information about Secure Agent requirements, contact Informatica Global Customer Support.

Configure the firewall

If your organization uses a protective firewall, include the Informatica Intelligent Cloud Services domain name or IP address ranges in the list of approved domain names or IP addresses. To ensure that the Secure Agent can perform all necessary tasks through the firewall, enable the port that the Secure Agent uses.

The Secure Agent uses port 443 (HTTPS) to connect to the internet. Configure your firewall to allow traffic to pass over port 443.

The allowlists of domains and IP addresses can vary according to your data center, which is also called a POD (Point of Deployment). You can identify your POD through the URL that appears when you open any service in Informatica Intelligent Cloud Services. The first few characters of the URL string identify the POD. For example, if the URL starts with `usw3.dm-us.informaticacloud.com`, your POD is USW3.

You can find the allowlists of Informatica Intelligent Cloud Services domains and IP addresses for different PODs in this Knowledge Base article on Informatica Network or by clicking the link at the top of the **Runtime Environments** page in Administrator.

Secure Agent permissions on Linux

A Secure Agent requires certain permissions to transfer data between sources and targets.

When you install a Secure Agent on Linux, the Secure Agent must have read/write/execute permissions for the installation directory.

Downloading and installing the Secure Agent on Linux

To install the Secure Agent on a Linux machine, you must download and run the Secure Agent installation program and then register the agent.

Secure Agent registration requires an install token. To get the install token, copy the token when you download the agent or use the **Generate Install Token** option in Administrator. The token expires after 24 hours.

When you register the agent, it is added to its own Secure Agent group by default. You can add the agent to a different Secure Agent group.

Before you download and install the Secure Agent, verify that no other Secure Agent is installed on the machine using the same Linux user account. If there is, you must uninstall it.

Tip: To verify the checksum of the Secure Agent installation program, use the agent REST API version 2 resource. For more information about the agent resource, see *REST API Reference*.

1. Open Administrator and select **Runtime Environments**.

2. On the **Runtime Environments** page, click **Download Secure Agent**.
3. Select the Linux 64-bit operating system platform, copy the install token, and then click **Download**.

   The installation program is downloaded to your machine. The name of the installation program is `agent64_install_ng_ext.<agent core version>.bin`.
4. Save the installation program to a directory on the machine where you want to run the Secure Agent.

   Note: If the file path contains spaces, the installation might fail.
5. From a shell command line, navigate to the directory where you downloaded the installation program and enter the following command:

   ```
   ./agent64_install_ng_ext.bin –i console
   ```
6. When the installer completes, navigate to the following directory:

   ```
   <Secure Agent installation directory>/apps/agentcore
   ```
7. To start the Secure Agent, enter the following command:

   ```
   ./infaagent startup
   ```

   The Secure Agent Manager starts. You must register the agent using the user name that you use to access Informatica Intelligent Cloud Services. You must also supply the install token.
8. If you did not copy the install token when you downloaded the agent, click **Generate Install Token** on the **Runtime Environments** page in Administrator, and copy the token.
9. To register the agent, in the `<Secure Agent installation directory>/apps/agentcore` directory, enter one of the following commands using your Informatica Intelligent Cloud Services user name and the token that you copied:
   - To add the agent to its own Secure Agent group, use the following command:

     ```
     ./consoleAgentManager.sh configureToken <user name> <install token>
     ```
   - To add the agent to an existing Secure Agent group, use the following command:

     ```
     ./consoleAgentManager.sh configureTokenWithRuntime <user name> <install token>
     <Secure Agent group name>
     ```

     Note: If the command includes a Secure Agent group name that doesn't exist, the Secure Agent is not assigned to a group. Be sure to use a valid Secure Agent group name.

   The following table lists the command options:

Option	Description
User Name	Required. Informatica Intelligent Cloud Services user name of the user installing the Secure Agent.
Install Token	Required. The install token that you copied.
Secure Agent group name	Optional. Include when you want to add the agent to an existing Secure Agent group instead. If this option isn't included in the command, the agent will be in its own Secure Agent group.

   You can check the registration status of a Secure Agent using the following command:

   ```
   ./consoleAgentManager.sh isConfigured
   ```

Configure the proxy settings on Linux

If your organization uses an outgoing proxy server to connect to the internet, the Secure Agent connects to Informatica Intelligent Cloud Services through the proxy server.

The Secure Agent installer configures the proxy server settings for the Secure Agent based on settings configured in the browser. You can update the proxy server settings defined for the Secure Agent from the command line.

To configure the proxy server settings for the Secure Agent on a Linux machine, use a shell command that updates the `proxy.ini` file. Contact the network administrator to determine the proxy settings.

1. Navigate to the following directory:

   ```
   <Secure Agent installation directory>/apps/agentcore
   ```

2. To update the `proxy.ini` file, enter the following command:

   ```
   ./consoleAgentManager.sh configureProxy <proxy host> <proxy port> <proxy user name>
   <proxy password>
   ```

3. Restart the Secure Agent.

Secure Agent services

Secure Agent services are pluggable microservices that the Secure Agent uses for data processing. Each Secure Agent service runs independently of the other services that run on the agent.

The independent services architecture provides the following benefits:

- The Secure Agent does not restart when you add a connector or package.
- Services are not impacted when another service restarts.
- Downtime during upgrades is minimized. The upgrade process installs a new version of the Secure Agent, updates connector packages, and applies configuration changes for the services. To minimize downtime, the old agent remains available and continues to run ingestion jobs during the upgrade. The new version of the Secure Agent runs jobs that start after the upgrade process completes.

The services that run on a Secure Agent vary based on your licenses and the Informatica Intelligent Cloud Services that your organization uses. For Mass Ingestion, the following Secure Agent services are available:

- Database Ingestion - for running application ingestion jobs and database ingestion jobs
- CMI Streaming Agent - for running streaming ingestion jobs
- Mass Ingestion - for running file ingestion jobs

Each service has a unique set of configuration properties. You might need to configure a service or change the service properties to optimize performance or if you are instructed to do so by Informatica Global Customer Support.

Database Ingestion service

Mass Ingestion Applications and Mass Ingestion Databases use the Database Ingestion agent service to run ingestion jobs.

After you download the Secure Agent to your runtime environment and enable the Database Ingestion service, the Database Ingestion packages are pushed to the on-premises system where the Secure Agent runs. You can then optionally configure properties for the Database Ingestion service that runs on the Secure Agent.

Database Ingestion service properties

To change or optimize the behavior of the Database Ingestion service that your Secure Agent group uses, configure Database Ingestion properties for your runtime environment.

To configure the properties, open your runtime environment and click **Edit**. Under **System Configuration Details**, select the **Database Ingestion** service and the **DBMI_AGENT_CONFIG** type.

The following table describes the Database Ingestion agent service properties:

Property	Description
maxTaskUnits	The maximum number application ingestion tasks and database ingestion tasks that can run concurrently on an on-premises machine where the Secure Agent is running. To calculate a reasonable number of task units for your Secure Agent machine, Informatica recommends that you divide the number of cores by 3 or 4. For example, if you have an 8-core machine, you could set this property to 2. Then monitor CPU usage and adjust the property value as needed to tune performance. During initial load processing, this property determines the number of tables that can be unloaded simultaneously. Remaining tables are queued and start unload processing when resources become available. **Note:** A single job can process many tables. The total number of tables that can be processed is limited only by available memory. On the average, 25 MB of RAM is required per table for an initial load task based on a 1 KB row size. During incremental load processing, this property determines the number of application ingestion and database ingestion jobs that can run simultaneously. Setting this property to a value greater than the number of cores on the Secure Agent machine can increase parallelism for task execution but also cause performance bottlenecks at task execution time.
serviceLogRetentionPeriod	The number of days to retain each internal Database Ingestion service log file after the last update is written to the file. When this retention period elapses, the log file is deleted. The default value is 7 days. Service logs are retained on the Secure Agent host where they are created: `<infaagent>/apps/Database_Ingestion/logs`. **Note:** This property is applicable to both Mass Ingestion Applications and Mass Ingestion Databases.
taskLogRetentionPeriod	The number of days to retain each job log file after the last update is written to the file. When this retention period elapses, the log file is deleted. The default value is 7 days.

Property	Description
ociPath	For Oracle sources and targets, the path to the Oracle Call Interface (OCI) file oci.dll or libclntsh.so. For a DBMI agent that is running, this value is appended to the path that is specified in the PATH environment variable on Windows or in the LD_LIBRARY_PATH environment variable on Linux. **Note:** This property is applicable only to Mass Ingestion Databases.
serviceUrl	The URL that the Database Ingestion service uses to connect to the Informatica Intelligent Cloud Services cloud. **Note:** This property is applicable to both Mass Ingestion Applications and Mass Ingestion Databases.
logLevel	The level of detail to include in the logs that the Database Ingestion service produces. Options are: - TRACE - DEBUG - INFO - WARN - ERROR The default value is TRACE. **Note:** This property is applicable to both Mass Ingestion Applications and Mass Ingestion Databases.
taskExecutionHeapSize	The maximum heap size, in gigabytes, for the Task Execution service. This value, in conjunction with maxTaskUnits property, affects the number of concurrent application ingestion and database ingestion tasks that can run on a Secure Agent. Try increasing the heap size to run more tasks concurrently. Enter this value followed by "g" for gigabytes, for example, '9g'. The default value is '8g'. **Note:** This property is applicable to both Mass Ingestion Applications and Mass Ingestion Databases.
useProxy	Set this property to true to enable the DBMI Agent to go through a proxy when connecting to or writing data to targets. The DBMI Agent then uses the proxy settings from the Secure Agent proxy configuration. By default, proxy settings are not used. **Note:** This property is applicable to both Mass Ingestion Applications and Mass Ingestion Databases.
intermediateStorageDirectory	For incremental load and combined initial and incremental load jobs, the local root directory under which intermediate files that contain data are stored when the **Enable Persistent Storage** option is selected in the associated task definitions. **Note:** This property is applicable only to Mass Ingestion Databases.

Property	Description
storageBackupDirectory	For incremental load and combined initial and incremental load jobs, the path to the directory that stores backup files when the **Enable Persistent Storage** option is selected in the associated task definitions. **Note:** This property is applicable only to Mass Ingestion Databases.
storageProperties	For incremental load and combined initial and incremental load jobs, a comma-separated list of key=value pairs that is used when the **Enable Persistent Storage** option is selected in the associated task definitions. Specify this property only at the direction of Informatica Global Customer Support. **Note:** This property is applicable only to Mass Ingestion Databases.
task_container.jvm.allowExceptionForInvalidEncodedData	If you receive transliteration errors that report invalid encoding to UTF-8, and you do not want to repair or correct the source data, set this property to false so that database ingestion jobs do not fail when trying to unload the data from the source. With this setting, the Database Ingestion service passes an equivalent Java property to the DataDirect JDBC driver to prevent the exception from occurring. After you set this property, you must restart the Database Ingestion service. **Note:** This property is applicable only to Mass Ingestion Databases.
testProperty	Do not set this property. It is intended for internal use by Informatica Global Customer Support and technical staff. This property appears only if you select DBMI_AGENT_ENV in the **Type** field.

Database Ingestion Agent environment variables

To change or optimize the behavior of the Database Ingestion Agent, define the following environment variables:

Environment Variable	Description
DBMI_REPLACE_UNSUPPORTED_CHARS	For Microsoft Azure Synapse Analytics targets, controls whether an application ingestion job or database ingestion job replaces characters in character data that the target cannot process correctly. To enable character replacement, set this environment variable to true. `DBMI_REPLACE_UNSUPPORTED_CHARS=true` Mass Ingestion Applications or Mass Ingestion Databases then uses the character that is specified in the DBMI_UNSUPPORTED_CHARS_REPLACEMENT environment variable to replace unsupported characters.
DBMI_UNSUPPORTED_CHARS_REPLACEMENT	If the DBMI_REPLACE_UNSUPPORTED_CHARS environment variable is set to true, specifies the character that replaces the characters in source data that a Microsoft Azure Synapse Analytics target cannot process correctly. Default value: ? (question mark) **Note:** Define this environment variable only for Mass Ingestion Databases.
DBMI_WRITER_CONN_POOL_SIZE	Indicates the number of connections that an application ingestion job or database ingestion job uses to propagate the change data to the target. The default value is 8. Valid values are 4 through 8.
DBMI_WRITER_RETRIES_MAX_COUNT	If a network issue occurs while a database ingestion job is loading source data to an Amazon S3 or Microsoft Azure Data Lake Storage Gen2 target, indicates the maximum number of times that the database ingestion job retries a request to continue the initial load or incremental load. If all of the retries fail, the job fails. The default value is 5.
DBMI_WRITER_RETRIES_INTERVAL_IN_MILLIS	Specifies the time interval, in milliseconds, that a database ingestion job waits before retrying the request to continue the initial load or incremental load to an Amazon S3 or Microsoft Azure Data Lake Storage Gen2 target if a network issue occurs. The default value is 1000.

Note: After you define or change an environment variable, restart the Database Ingestion Agent for the changes to take effect.

Mass Ingestion (Files)

To change or optimize the behavior of Mass Ingestion Files that your Secure Agent group uses, configure Mass Ingestion properties for your runtime environment in Administrator.

You can configure the following properties:

Type	Name	Description
AGENT_RUNTIME_SETTINGS	file-listener-snapshot-dir	A directory where the snapshots of a new file listener components are added. You can add the following directory paths: - A path relative to the `MassIngestionRuntime` directory. For example, `../data/monitor`. - The absolute path. For example, `<Secure agent installation directory>/apps/MassIngestionRuntime/data/monitor` where *Secure agent installation directory* is the name of the directory where the secure agent is installed. **Note:** Use the snapshot directory shared with all agents when multiple Secure Agents are present in a group.
AGENT_RUNTIME_SETTINGS	mi-task-workspace-dir	A directory in the agent that file ingestion tasks use as an intermediate staging area when transferring files to a target. The directory is a custom location in the agent. The path can be a shared location, mounted location, or a location apart from the default location in the agent.
AGENT_RUNTIME_SETTINGS	mi-task-quarantine-dir	A directory where the file ingestion task stores the infected files detected when you run a virus scan. The directory is the custom location in the agent. The path can be a shared location, mounted location, or a location apart from the default location in the agent. For example, userdata\quarantine **Note:** To automatically clean up the quarantine directory, set the agent property for the quarantine location to a system temporary files location such as `/tmp/informatica/fmi/quarantine`.
AGENT_RUNTIME_SETTINGS	file-listener-max-pool-size	The maximum number of threads to execute the file listener. Default is 20.
AGENT_RUNTIME_SETTINGS	file-listener-core-pool-size	The total number of threads. Default is 20.
AGENT_RUNTIME_SETTINGS	fmi-task-max-pool-size	The maximum number of threads to execute the file ingestion task. Default is 50.
AGENT_RUNTIME_SETTINGS	fmi-task-core-pool-size	The initial or minimum number of threads. Default is 20.
AGENT_RUNTIME_SETTINGS	ftp-receive-socket-buffer-size	The buffer size for FTP inbound packets. Default is 16 bytes.

Type	Name	Description
AGENT_RUNTIME_SETTINGS	ftp-send-socket-buffer-size	The buffer size for FTP outbound packets. Default is 16 bytes.
AGENT_RUNTIME_SETTINGS	http-client-timeout	The timeout duration in seconds for Agent requests to Informatica Intelligent Cloud Services. Default is 30 seconds.
PGP_SETTINGS	public-keyring-path	The directory to store the public key ring. You can add the following directory paths: - A path relative to the directory where mass ingestion is installed. For example, `../data/pubring.pkr` where *pubring.pkr* is the name of the file where you store the public key ring. - The absolute path. For example, `<Secure agent installation directory>/apps/MassIngestionRuntime/data/pubring.pkr` where *pubring.pkr* is the name of the file where you store the public key ring and *Secure agent installation directory* is the name of the directory where the agent is installed.
PGP_SETTINGS	secret-keyring-path	The directory to store the secret key ring. You can add the following directory paths: - A path relative to the directory where mass ingestion is installed. For example, `../data/secring.pkr` where *secring.pkr* is the name of the file where you store the secret key ring. - The absolute path. For example, `<Secure agent installation directory>/apps/MassIngestionRuntime/data/secring.pkr` where *secring.pkr* is the name of the file where you store the secret key ring and *Secure Agent installation directory* is the name of the directory where the agent is installed.
JVM_SETTINGS	app-heap-size	The minimum and maximum heap sizes of the Mass Ingestion Files application. Default is -Xms256m -Xmx2048m.
JVM_SETTINGS	lcm-heap-size	The minimum and maximum heap sizes of life-cycle management scripts. Default is -Xms32m -Xmx128m.

You can configure the following properties in the **Custom Configuration Details** area when you edit a Secure Agent:

Type	Name	Description
AGENT_RUNTIME_SETTINGS	ComplexFileDisableWriteChecksum	Set the value to **True** to ignore the `crc` file. The job runs successfully with Hadoop Files V2 as source and Snowflake Cloud Data Warehouse V2 as the target.

CMI Streaming Agent

Use the CMI Streaming Agent to define and deploy streaming ingestion tasks. You configure streaming ingestion tasks in the Mass Ingestion service.

A CMI Streaming Agent runs on an on-premise system and works in conjunction with the Mass Ingestion Streaming service. In an on-premise system, the CMI Streaming Agent runs the jobs deployed by Mass Ingestion Streaming. The agent provides status and statistics updates of each job.

On Linux, the CMI Streaming Agent does not start if the agent installation directory name contains a space. The agent returns a connection timeout status. After a few restart attempts, the agent goes into the error state.

Note: Prior to Spring 2020 April release of Informatica Intelligent Cloud Services Mass Ingestion service, CMI Streaming Agent was called Streaming Ingestion Agent.

CMI Streaming Agent properties

To change or optimize the behavior of the CMI Streaming Agent, configure agent properties for your run-time environment. Configure CMI Streaming Agent properties in the **System Configuration Details** area when you edit a Secure Agent.

You can configure Engine, Agent, and Script properties of a CMI Streaming Agent.
The following image shows some of the CMI Streaming Agent properties:

System Configuration Details

Service: CMI Streaming Agent

Type: All Types

Type	Name	Value
Engine	MaxLogFileSize	'5MB'
Engine	LogLevel	'DEBUG'
Agent	DataflowPullInterval	60
Agent	JVM	'-Xms256M -Xmx256M'
Agent	LogLevel	'DEBUG'
Agent	MaxLogFileSize	'10MB'
Agent	MaxNumberOfBackups	5
Scripts	LogLevel	'DEBUG'
Scripts	MaxFileSize	'5MB'
Scripts	MaxBackupIndex	5

You can configure the following CMI Streaming Agent properties:

Type	Property Name	Description
Engine	MaxLogFileSize	The maximum size of the log file that the engine can create. Default is 5 MB.
Engine	LogLevel	The log level for the engine.
Agent	DataflowPullInterval	The time interval after which the agent checks for updates in the task. Default is 60 seconds.
Agent	JVM	List of JVM properties for the agent. For example: `[-Xms256M -Xmx256M]`
Agent	LogLevel	The log level for the agent.
Agent	MaxLogFileSize	Maximum size of the log files that an agent can create. Default is 10 MB.
Agent	MaxNumberOfBackups	Maximum number of backup log files for the agent. Default is 5.
Scripts	LogLevel	The log level of the scripts.
Scripts	MaxFileSize	The maximum file size after which the log rolls over and creates a new file. Default is 10 MB.
Scripts	MaxBackupIndex	Maximum number of backup files maintained after rolling over. Default is 5.

Streaming Agent offline mode

You can run and monitor a streaming ingestion job when the CMI Streaming Agent is offline or not connected to the internet.

The Streaming Agent supports both online and offline modes of communication. In the offline mode, the streaming ingestion job continues to run even if the Streaming Agent does not communicate with the Informatica Intelligent Cloud Services for an extended period of time. The Streaming Agent continues to monitor the health and statistics of the ingestion tasks locally. When the Streaming Agent turns online and connects to the cloud services, it updates any configuration changes for the agent and tasks, as well as updates the health and statistics to the services.

To switch between the offline and online modes, you can use the command line utility provided by the Mass Ingestion Streaming service. Run the following command to start the command line utility:

```
<Informatica Secure Agent>/apps/Streaming_Ingestion_Agent/<version>/runcli.bat
```

The command line utility uses the command prompt `infa/stream>` and provides three groups of commands.

You can change the communication modes only through the command line utility. The Streaming Agent preserves the communication mode when the agent restarts.

The following table lists the commands of this command line utility:

Command	Description	Example

app-config	Shows the current configuration of the Streaming Agent application.	`infa/stream :>app-config`
		`deploy.pull.interval : 60`
		`health.poll.interval : 30`
		`minifi.ingester.file.location : ./conf`
		`siagent.communication.mode : Online`
		`siagent.monitoring.persist.dir : ../data`
		`siagent.statistics.post.batchsize : 720`
		`siagent.statistics.post.concurrency : 60`
		`siagent.status.persist.dir : ../data`
		`statistics.poll.interval : 30`

app-setconfig	Use this command to configure the following properties:
	- `siagent.communication.mode`. Use to configure offline or online communication mode.
	- `siagent.statistics.post.batchsize`. Use to define the number of snapshots in a batch.
	- `siagent.statistics.post.concurrency`. Use to define the number of worker threads to post statistics.
	The `--key` and `--value` tokens are optional.
		`infa/stream :>app-setconfig --key siagent.statistics.post.batchsize --value 20`
		or
		`infa/stream :>app-setconfig siagent.statistics.post.batchsize 20`

app-status	Shows the current status of the Streaming Agent.
	The `health status code` and `health error message` indicates the status of the agent (service) shown on the **Administrator**.
	`uptime` indicates the number of seconds since the Streaming Agent application is available.
		`infa/stream :>app-status`
		`health error message : No errors`
		`health status code : RUNNING(0)`
		`uptime : 67828`

app-statistics	Shows metadata and status of overall statistics collection in the Streaming Agent.
	- `collection interval`. Interval of statistics collection, in seconds.
	- `post interval`. Frequency of statistics posted or attempted post.
	- `max batch size`. Maximum number of snapshots posted in a single http post.
	- `last batch size`. Number of snapshots in the last http post.
	- `last time collected`. Timestamp when any statistics were last collected.
	- `last time posted`. Timestamp when any statistics were last posted.
		`infa/stream :>app-statistics`
		`collection interval : 30`
		`last batch size : 2`
		`last time collected : 7/3/20 10:19:03 AM IST`
		`last time posted : 7/3/20 10:18:53 AM IST`
		`max batch size : 20`
		`pending snapshots : 3`
		`post interval : 30`

Command	Description	Example
clear	Clears the screen.	-
exit, quit	Quits the application.	-
help	Shows a summary of all the commands available.	infa/stream :>help AVAILABLE COMMANDSAgent Application Commands app-config: Show agent application configuration app-setmode: Set the communication mode [Online/Offline] app-status: Show agent application statusBuilt-In Commands clear: Clear the shell screen. exit, quit: Exit the shell. help: Display help about available commands.Streaming Ingestion Task Commands task-health: Show streaming ingestion task health task-list: Show streaming ingestion task list task-metadata: Show streaming ingestion task metadata
task-list	Shows the list of streaming ingestion jobs currently deployed on the Streaming Agent.	infa/stream :>task-list 6e61e76f-2618-4292-ab3d-dd181f47ee91 ad5053c7-5ac2-493f-8cbb-a24900b61f71
task-health	Shows the health status of all streaming ingestion jobs in the Streaming Agent. Use the options `--name` or `--id` to specify a job. If none are specified, all jobs are listed.	infa/stream :>task-health --name aby_df4 processors : [{"id":"14a7a095-7fac-4fc3-ac5c-705369132516","status":"ERROR"},{"id":"821e6730-3aed-4d3f-b875-45f424b6b963","status":"RUNNING"}] status : ERROR timestamp : Sat May 09 06:04:08 IST 2020 infa/stream :>task-health 6e61e76f-2618-4292-ab3d-dd181f47ee91 processors : [{"id":"2a0b8715-aa7a-46c5-9d6a-6a356f5a0102","status":"ERROR"},{"id":"1172f3a8-35dd-41ef-be4b-bc0cf37e3794","status":"RUNNING"}] status : ERROR timestamp : Sat May 09 06:04:08 IST 2020 ad5053c7-5ac2-493f-8cbb-a24900b61f71 processors : [{"id":"14a7a095-7fac-4fc3-ac5c-705369132516","status":"ERROR"},{"id":"821e6730-3aed-4d3f-b875-45f424b6b963","status":"RUNNING"}] status : ERROR timestamp : Sat May 09 06:04:08 IST 2020

Command	Description	Example

task-metadata	Shows the metadata of all streaming ingestion jobs in the Streaming Agent.
	Use the options `--name` or `--id` to specify a job.
	If none are specified, all jobs are listed.
		`infa/stream :>task-metadata --name aby_df4`
		`id : ad5053c7-5ac2-493f-8cbb-a24900b61f71`
		`name : aby_df4`
		`runId : 9071 version : 1`
		`infa/stream :>task-metadata`
		`6e61e76f-2618-4292-ab3d-dd181f47ee91`
		`id : 6e61e76f-2618-4292-ab3d-dd181f47ee91`
		`name : aby_df2`
		`runId : 9069`
		`version : 8`
		`ad5053c7-5ac2-493f-8cbb-a24900b61f71`
		`id : ad5053c7-5ac2-493f-8cbb-a24900b61f71`
		`name : aby_df4`
		`runId : 9071`
		`version : 1`

task-statistics	Shows the statistics details of all streaming ingestion jobs in the Streaming Agent.
	Use the options `--name` or `--id` to specify a job.
	If none are specified, all jobs are listed.
		`infa/stream :>task-statistics --name aby_df1`
		`dataflow name : aby_df1`
		`last time collected : 1590861803731`
		`last time posted : 1590861806091`
		`infa/stream :>task-statistics`
		`7b7d3c09-df43-482f-b6c8-8dd80187e6d7`
		`dataflow name : aby_df2`
		`last time collected : 1590861770731`
		`last time posted : 1590861741132`
		`decfad0a-20df-4226-84f9-1ff1ab6ef96a`
		`dataflow name : aby_df1`
		`last time collected : 1590861768730`
		`last time posted : 1590861771054`

Online mode to Offline mode

By default, the Streaming Agent is in online mode.

To change the Streaming Agent to offline mode:

1. Launch the command line utility using the following command:
   In Windows:

   ```
   <Informatica_Secure_Agent>/apps/Streaming_Ingestion_Agent/<version>/runcli.bat
   ```

   In UNIX:

   ```
   <Informatica_Secure_Agent>/apps/Streaming_Ingestion_Agent/<version>/runcli.sh
   ```

2. Set the Streaming Agent to offline mode:

```
app-setconfig --key siagent.communication.mode --value Offline
```

or

```
app-setconfig siagent.communication.mode Offline
```

The Streaming Agent stops sending health updates and statistics of any streaming ingestion job.

Offline mode to Online mode

To change the Streaming Agent to online mode:

1. Launch the command line utility using the following command:
In Windows:

```
<Informatica_Secure_Agent>/apps/Streaming_Ingestion_Agent/<version>/runcli.bat
```

In UNIX:

```
<Informatica_Secure_Agent>/apps/Streaming_Ingestion_Agent/<version>/runcli.sh
```

2. Set the Streaming Agent to online mode:

```
app-setconfig --key siagent.communication.mode --value Online
```

or

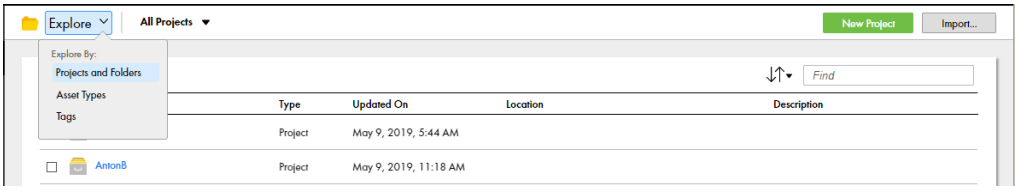
```
app-setconfig siagent.communication.mode Online
```

The Streaming Agent starts sending health updates of all the streaming ingestion jobs and the updates appear in the **Monitoring** page. It starts sending statistics of all the streaming ingestion jobs including the statistics backlog collected while it was offline to the service. It also synchronizes the updates to the streaming ingestion jobs or adds the new streaming ingestion job deployed while it was offline to the service.

Creating projects and project folders

Create a project to contain your ingestion task assets. You can create one or more folders under the project to logically organize your assets. However, you cannot create subfolders under a folder.

1. In the Mass Ingestion service, open the **Explore** page.
2. If the **Explore** page shows objects other than projects, select **Projects and Folders** in the **Explore** menu.



3. To create a project, click **New Project**.
4. In the **New Project Properties** dialog box, enter a project name up to 255 characters in length. You can also enter an optional description of the project.

   A project name cannot contain the following characters: `#?`|{}"^&[]/\.`
5. If you want to add a folder under the project, select the project and click **New Folder**.

6. In the **New Folder Properties** dialog box, enter a folder name up to 255 characters in length. You can also enter an optional description of the project.

   A folder name cannot contain the following characters: `#?`|{}"^&[]/\.`

When you define an ingestion task, you must specify the project or project folder location to contain the task definition.

Editing your user profile

Your user profile contains the details of your Informatica Intelligent Cloud Services user account.

You can update the following information in your profile:

- First and last name
- Job title
- Email address
- Phone number
- Time zone (used in the job execution time stamps on the **All Jobs**, **Running Jobs**, **My Jobs**, **Import/Export Logs**, and **My Import/Export Logs** pages)
- Password
- Security question and answer

Note: If you use SAML to sign on to Informatica Intelligent Cloud Services and your organization administrator has enabled SAML group and role mapping on the **SAML Setup** page in Administrator, you can only update the time zone. The other attributes are updated directly from your enterprise directory each time you log into Informatica Intelligent Cloud Services.

1. Click the **User** icon in the top right corner of the Informatica Intelligent Cloud Services window and then select **Profile**.
2. On the **Profile** page, add or edit personal information such as your name, job title, phone number, and time zone.
3. To update your email address, click **Update Email**.

   Informatica Intelligent Cloud Services sends a verification email to your new email address. The email contains a link that is valid for 24 hours. When you click the link in the email, the new address is verified, and it appears in your profile. If the link expires, you can resend the verification email.
4. Optionally, change your password or security question.
5. Click **Save**.

INDEX

T

U

W