Informatica®
Current Release

Operational Insights Online Help

Informatica Operational Insights Online Help
Current Release
February 2019

Publication Date: 2019-02-14

Table of Contents

CHAPTER 1

Introducing Operational Insights

Operational Insights is an Informatica Intelligent Cloud Services service that gives you visibility into the performance and operational efficiency of Informatica assets across your enterprise.

The rich analytics-driven features Operational Insights provide you with the following capabilities:

- Comprehensive monitoring statistics provide an overview of the health of all Informatica assets across the enterprise, including Big Data and PowerCenter domains, nodes, grids and services.
- Data processing analytics based on Big Data domain job and PowerCenter workflow statistics enable you to assess usage of your investment in Informatica services.
- Run-time job execution, workflow, and task metrics allow you to troubleshoot performance degradation and execution failures.
- Resource consumption and job and workflow trend analytics help you predict when you need to increase capacity or reallocate resources.
- Recommendations help you identify and fix the most commonly occurring errors within a domain.
- Email notifications alert you about problems within a Big Data or PowerCenter domain or with a Secure Agent.
- Dynamic auto-scaling of PowerCenter grids running in Amazon EC2 by adding elastic nodes in the cloud to increase processing capacity when needed.

Operational Insights provides you with a single pane of glass for monitoring Informatica domains running in on-premise sites as well as in an Amazon Web Services cloud. The service is aware of all domains, nodes, grids, and services deployed within your enterprise. Run-time statistics and asset configuration metadata is uploaded from domains to the service on configurable schedules, providing you with an accurate, up-to-date overview of your Informatica deployments.

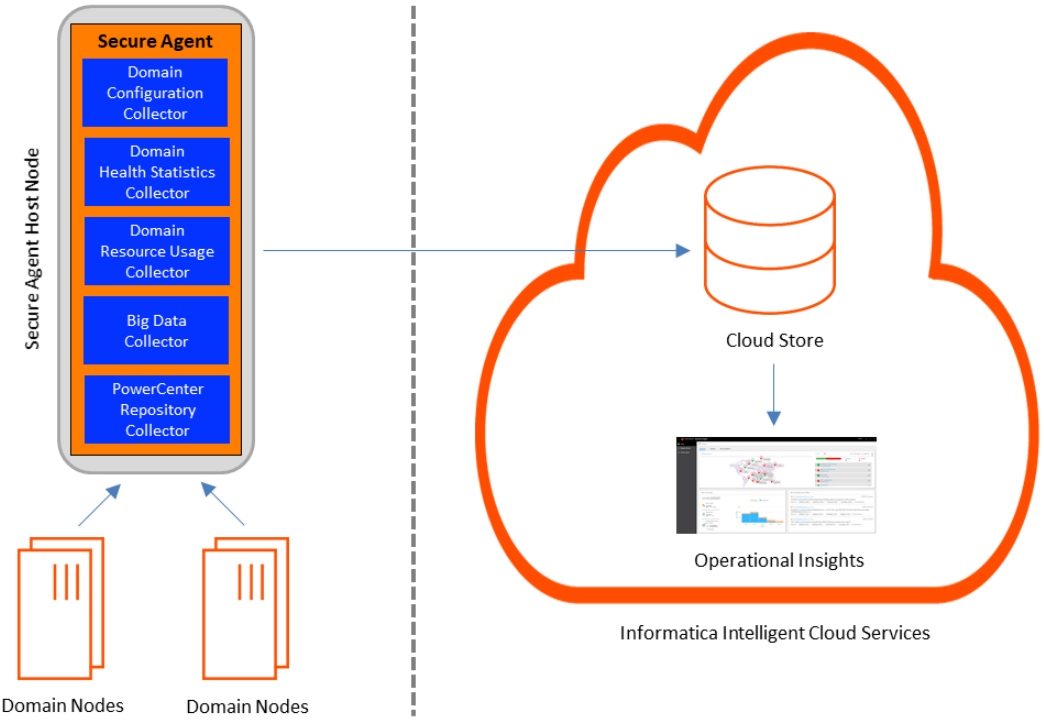
Understanding the architecture

The Secure Agent required by Informatica Intelligent Cloud Services provides the connection between an Informatica domain and Operational Insights. The Secure Agent manages the collection of metrics, performance statistics, and configuration metadata from assets within the domain. The Secure Agent also uploads the collected operational data and domain-related metadata to the Informatica Intelligent Cloud Services for use by Operational Insights.

You must configure each domain that Operational Insights monitors to communicate with a Secure Agent. You can configure multiple domains to communicate with a single Secure Agent, or you can download and install a Secure Agent on a node within a domain.

Components known as collectors running within the Secure Agent perform the actual collection and uploading of statistics and configuration metadata from domain assets. You can configure the collection schedule for each collector.

The following image provides a high-level overview of the Operational Insights infrastructure:



After the Secure Agent is enabled, it detects all assets within the domain, including nodes, application services, PowerCenter repositories, and Model repositories. When you add a new asset to the domain, the Secure Agent automatically recognizes the asset. The Secure Agent also detects when a PowerCenter Repository Service or a Model Repository Service is added to the domain; you only need to provide the connection details for the PowerCenter repository database or the Model repository database to enable statistics collection.

Understanding collectors

Collectors are components that run within the Secure Agent that communicates with an Informatica domain. You do not need to configure a collector to collect data from a specific asset; the Secure Agent is aware of all assets within the domain, and each enabled collector collects operational data and metadata from the relevant assets.

The collectors can collect data from Informatica release 9.6.1 domains and from all Informatica release 10.x domains. If needed, you can disable any collector except for the Domain Configuration Collector. You can also change the default collection schedule for each collector.

The collectors begin collecting and uploading data to Informatica Intelligent Cloud Services after you click **Save** as the final step in the domain registration process. Data is uploaded to Informatica Intelligent Cloud Services at the time it is collected.

The following collectors are deployed with the Secure Agent:

Collector	Description	Default Collection Frequency	Metadata Collected
Domain Configuration Collector	Collects and uploads configuration metadata for the domain and all domain assets, including nodes and services.	Every 24 hours	- Domain details: Domain name, list of nodes in the domain, details for grids within the domain. - Node details: Name, HTTP port, logs directory, maximum processes, list of services running. - System configuration for each node: Operating system details, number of CPU cores and CPU speed, physical memory details.
Domain Health Statistics Collector	Collects and uploads availability statistics for domain assets, including nodes and application services.	Every 5 minutes	- Availability statistics: Domains, nodes, services, and grids.
Domain Resource Usage Statistics Collector	Collects and uploads CPU and memory consumption statistics for all nodes within the domain. The collector collects the statistics from the Model repository managed by the monitoring Model Repository Service specified in the Monitoring Configuration for the domain.	Every 1 hour	- CPU utilization: Informatica processes and all processes running on each node. - Memory utilization: Informatica processes and all processes running on each node.

Collector	Description	Default Collection Frequency	Metadata Collected
Big Data Collector	Collects and uploads statistics on Big Data jobs run on Hadoop. You must specify the cluster configuration that contains the configuration information for each Hadoop cluster you want to collect statistics from.	Every 1 hour	- Hadoop cluster configuration: Number of clusters and nodes by Hadoop distribution. - Hadoop cluster resource usage: CPU and memory utilization for Informatica processes and all processes running on each cluster node. - Job execution statistics: Type of job, how many rows inserted/ rejected, how many mappings failed/succeeded, volume of data processed, start and end time, Data Integration Service that submitted the job. - Runtime metrics: Number and type of unique mappings and workflows that ran on each cluster or on all clusters.
PowerCenter Repository Collector	Collects and uploads runtime workflow and session metrics from PowerCenter repositories within the domain. You must provide the JDBC connection details for the repository database for each PowerCenter repository you want to collect statistics from.	Every 1 hour	- Workflow details: Workflow name, start and end time, status (succeeded, failed, etc.), PowerCenter Integration Service than submitted the workflow. - Session tasks: Task ID and type, start and end time, rows read, rows written, node the task ran on. - Repository folder details: Folder ID and name.

CHAPTER 2

Creating accounts

You must create the following accounts to enable users in your organization to use Operational Insights:

1. Create an Informatica Intelligent Cloud Services account for your organization, if you do not already have one.
2. Create a user account for each Operational Insights user.

Creating an Informatica Intelligent Cloud Services user

All Operational Insights users must have an Informatica Intelligent Cloud Services user account. You use the Informatica Intelligent Cloud Services Administer service to create user accounts.

For additional details on creating and managing users, see *Administrator* in the Cloud Data Integration online help.

1. Log in to Informatica Intelligent Cloud Services at the following link:

   https://dm-us.informaticacloud.com/identity-service/home
2. Click **Administrator**.
3. Click **Users** in the navigation pane.
4. Click **Add User**.
5. Provide details and assign roles to the user.
6. Click **Save** when finished.

CHAPTER 3

Installing and Configuring a Secure Agent

You must configure every domain that Operational Insights monitors to communicate with a Secure Agent.

If the domain can access an existing Secure Agent, you can configure the domain to use the Secure Agent when you register the domain with Operational Insights. You can configure multiple domains within your organization to use the same Secure Agent.

If the domain doesn't have access to a Secure Agent, you must download and install a Secure Agent on a node within the domain. The Secure Agent must be installed on a machine that has access to the internet.

Note that if the Operational Insights service shuts down for more than 12 hours, the Secure Agent connection to the service times out. Manually restart the Secure Agent on the node to re-establish the connection.

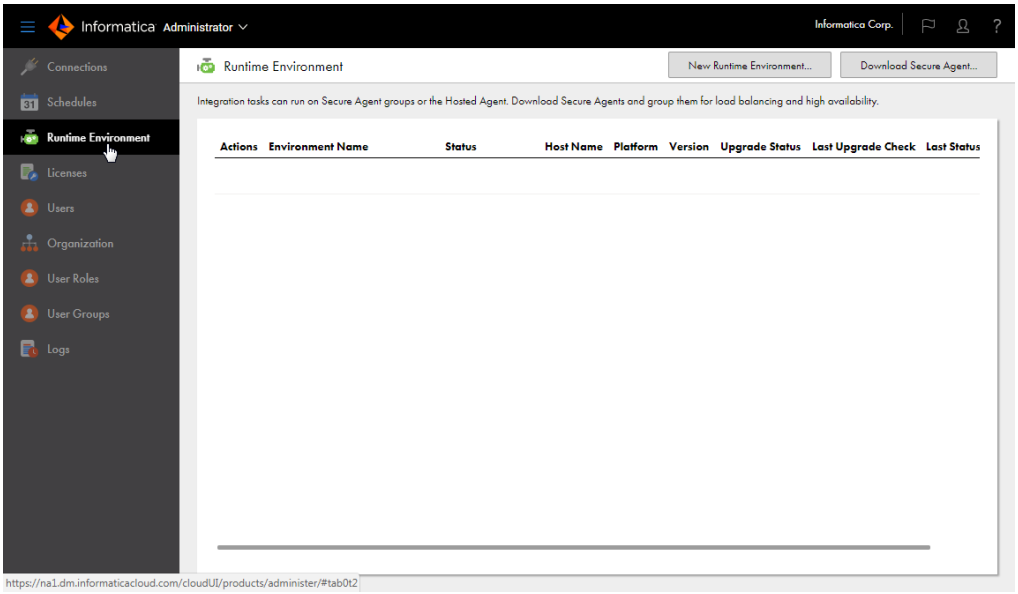
Downloading a Secure Agent

If the Informatica domain cannot access a Secure Agent, download a Secure Agent to a machine that the domain can access.

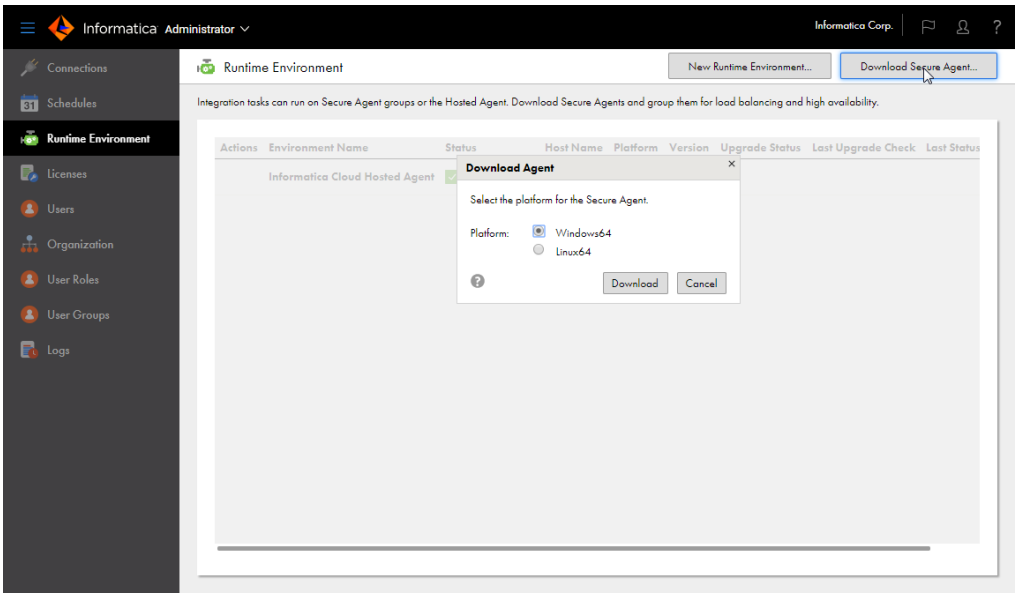
1. Log in to Informatica Intelligent Cloud Services at the following link:

   https://dm-us.informaticacloud.com/identity-service/home
2. Click **Administrator**.

3. Click **Runtime Environment** in the navigation panel.



4. Click **Download Secure Agent**.



5. Select the operating system on which you will run the Secure Agent, and then click **Download**.
6. Save the Secure Agent installer to the machine where you want to run the Secure Agent.

Secure Agent Install Prerequisites

To use the Secure Agent with Informatica Intelligent Cloud Services, perform the following tasks before you download and install the Secure Agent:

- On Windows, you must log in as a non-administrative user. On Linux, you must log in as a non-root user. If you install the Secure Agent as a user with administrative rights or with root access, the underlying Secure Agent database does not start.
- (Optional) If you want use JDBC and SAP connector third party libraries with the Process Server, copy the third party libraries to the following location:
  `<Secure Agent Installation Directory>/apps/process-engine/ext`

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. Use the Secure Agent Manager to stop and restart the Secure Agent. You can also use the Secure Agent Manager to check the Secure Agent status and configure proxy information.

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 where you install the Secure Agent 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:

- Verify that the machine on which you install the Secure Agent 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 Informatica Network:

  https://network.informatica.com/community/informatica-network/product-availability-matrices/overview
- Verify that the account you use to install the Secure Agent has access to all remote directories that contain flat source or target files.
- Verify that no other Secure Agent is installed on the machine. If another Secure Agent is installed on the machine, you must uninstall it first.

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 and Operational Insights domain name or IP address ranges in the list of approved domain names or IP addresses. You should also enable the port that the Secure Agent uses. This ensures that the Secure Agent can perform all necessary tasks through the firewall.

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

The whitelists 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 must add the IP address ranges required by both Informatica Intelligent Cloud Services and Operational Insights to your list of approved domain names or IP addresses.

You can find the whitelists of Informatica Intelligent Cloud Services domains and IP addresses for different PODs in the following KB article on Informatica Network:

https://kb.informatica.com/faq/7/Pages/20/524982.aspx

You can find the whitelist of Operational Insights IP address ranges in the following article:

https://kb.informatica.com/faq/7/Pages/21/532624.aspx

Secure Agent Permissions

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.

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

Configuring 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.

Download the Windows Secure Agent Installer

Download the Secure Agent installer from Informatica Intelligent Cloud Services. You can download and run the Secure Agent on any machine that meets the minimum requirements.

1. Open Administrator and select **Runtime Environments**.
2. On the **Runtime Environments** page, click **Download Secure Agent**.
3. Select the operating system on which you will run the Secure Agent, and then click **Download**.
4. Save the installation wizard to the machine where you want to run the Secure Agent.

   The name of the Windows Secure Agent installation file is `agent64_install.exe`

Install and Register the Secure Agent on Windows

After you download the Secure Agent installer, you can install it. You can install one Secure Agent on a machine. After the installation completes, you can register the Secure Agent. After you complete the registration, the Secure Agent Manager displays the Secure Agent status.

Verify that no other Secure Agent is installed on the machine. If there is, you must uninstall it.

1. Open the directory where you downloaded `agent64_install.exe`, and double-click the file.
2. Choose the installation folder and click **Next**.
3. Review the pre-installation summary and click **Install**.
4. After the installer completes, click **Done**.

   A registration page appears.
5. Enter your Informatica Intelligent Cloud Services user name and password and click **Register**.

   The Secure Agent starts, and the **Informatica Cloud Secure Agent** window displays the status of the Secure Agent. You can stop and start the agent and change the proxy settings in this window.
6. 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 machine on which the Secure Agent is installed 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 where you install the Secure Agent meets the minimum requirements.
2. Download the Secure Agent installer files.
3. Install and register the Secure Agent.

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 on which you install the Secure Agent 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 Informatica Network:

  https://network.informatica.com/community/informatica-network/product-availability-matrices/overview
- Verify that the machine where you install the Secure Agent has at least 500 MB of free disk space.

- 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 and Operational Insights domain name or IP address ranges in the list of approved domain names or IP addresses. You should also enable the port that the Secure Agent uses. This ensures that the Secure Agent can perform all necessary tasks through the firewall.

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

The whitelists 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 must add the IP address ranges required by both Informatica Intelligent Cloud Services and Operational Insights to your list of approved domain names or IP addresses.

You can find the whitelists of Informatica Intelligent Cloud Services domains and IP addresses for different PODs in the following KB article on Informatica Network:

https://kb.informatica.com/faq/7/Pages/20/524982.aspx

You can find the whitelist of Operational Insights IP address ranges in the following article:

https://kb.informatica.com/faq/7/Pages/21/532624.aspx

Secure Agent Permissions

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.

Download the Linux Secure Agent Installer

Download the Secure Agent installer from Informatica Intelligent Cloud Services. You can download and run the Secure Agent on any machine that meets the minimum requirements.

1. Open Administrator and select **Runtime Environments**.
2. On the **Runtime Environments** page, click **Download Secure Agent**.
3. Select the operating system on which you will run the Secure Agent, and then click **Download**.
4. Save the installer to a path on the machine where you want to run the Secure Agent.

   Note: If the path contains spaces, the installation might fail.

   The name of the Linux Secure Agent installation file is `agent64_install.bin`.

Install and Register the Secure Agent on Linux

You can install one Secure Agent on a machine. You must uninstall the existing Secure Agent from the machine before you install a Secure Agent. After the installation completes, start the Secure Agent, then use your Informatica Intelligent Cloud Services user name and password to register the Secure Agent.

1. From a shell command line, navigate to the directory where you downloaded the file and enter the following command:

   ```
   agent64_install.bin -i console
   ```

   Note: If you install the Secure Agent on Linux, do not include spaces in the directory path. If you include spaces, the Secure Agent installation might stop responding.
2. When the installer completes, navigate to the following directory:

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

   ```
   infaagent startup
   ```
4. After the Secure Agent starts, register the Secure Agent. In the same directory, enter the following command using your Informatica Intelligent Cloud Services user name and password:

   ```
   consoleAgentManager.sh configure <username> <password>
   ```

   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.bat configureProxy <proxy host> <proxy port> <proxy user name>
   <proxy password>
   ```
3. Restart the Secure Agent.

CHAPTER 4

Registering and managing domains

You must register each Informatica domain that you want to monitor using Operational Insights. A wizard guides you through the process.

The following are the key steps in the registration process:

1. Configure the connection to the domain.
2. Supply details that enable users to more easily locate the domain.
3. Configure the collectors.

After you complete the registration process and click **Save** as the final step, the collectors begin collecting and uploading data to the Informatica Intelligent Cloud Services for use by Operational Insights.

Enabling the monitoring Model Repository Service

You must configure the monitoring Model Repository Service and the associated Model Repository in a domain before you register a domain with Operational Insights.

Operational Insights extracts CPU and memory consumption metrics for domain nodes from the Model repository specified in the Monitoring Configuration for the domain. The Model Repository Service that manages with the Model repository is known as the monitoring Model Repository Service.

You can create a monitoring Model Repository Service when you run the installer to create a domain. For more information, see the "Prepare for Application Services and Databases" chapter in the *Informatica Big Data Suite Installation and Configuration Guide*.

You can also use the Administrator tool to configure the monitoring Model Repository Service in a domain. For more information, see the "Configuring the Monitoring Model Repository Service" section in the *Informatica Application Service Guide*.

Configuring the domain connection

Enter the information required to enable Operational Insights to connect to the domain.

1. Click **Register Domain** in the navigation bar on the left side of the page.

2. Enter the following general properties:

Property	Description
Domain Display Name	Domain name to display in the Operational Insights user interface. You can assign any name you like, but it must be unique within Operational Insights.
Domain Name	Domain name displayed in Informatica Administrator (the Administrator tool).
Master Gateway Host	Host name of the master gateway node machine. Enter the value exactly as shown in the General Properties > Host Name property for the node in the Administrator tool. To find this value in the Administrator tool: - Select the **Services and Nodes** view, and then select the node in the Domain Navigator. - Under General Properties, locate the Host Name property.
Gateway Node Port	HTTP port used by the gateway node.
Domain Version	Informatica release installed in the domain. Operational Insights can monitor assets within an Informatica release 9.6.1 and all Informatica release 10.x domains.
Products	Informatica products to monitor using Operational Insights. Products are selected based on the domain version.

3. Select the Secure Agent that collects and uploads data from the domain to the Informatica Intelligent Cloud Services.

Property	Description
Secure Agent Group	Group the Secure Agent installed in the domain belongs to.
Secure Agent Name	Name of the Secure Agent installed in the domain.

4. Enter the following domain security details. Operational Insights must be able to connect to the domain as an Informatica administrator.

Property	Description
Security Domain	Select the security domain used by the domain.
Administrator User Name	User name for the Informatica domain administrator account.
Administrator Password	Password for the Informatica domain administrator account.
TLS Enabled	Select if the domain is secured with the Transport Layer Security (TLS) protocol

Property	Description
Truststore Path	If the domain is secured with TLS, copy the `infa_truststore.jks` file from a domain node to the Secure Agent host, and then specify the path and file name for the file on the Secure Agent host. By default, the file is installed in the following directory on each domain node: `<Informatica installation directory>\services\shared\security`
Truststore Password	If the domain uses a custom truststore file, specify the encrypted truststore password.

5. Enter the following domain auto-scaling details to enable elastic nodes added to the grid to communicate with the domain.

Property	Description
Sitekey Path	Copy the `sitekey` file from a domain node to the Secure Agent host, and then specify the path and file name for the file on the Secure Agent host. By default, the file is installed in the following directory on each domain node: `<Informatica installation directory>\isp\config\keys`
Keystore Path	If the domain is secured with TLS, copy the `infa_keystore.jks` file from a domain node to the Secure Agent host, and then specify the path and file name for the file on the Secure Agent host. By default, the file is installed in the following directory on each domain node: `<Informatica installation directory>\services\shared\security`
Keystore Password	If the domain uses a custom keystore file, specify the encrypted keystore password.

6. Click **Test Connection** to test the connection to the master gateway node.

Entering the domain details

Enter details to help users find the domain within Operational Insights.

You can use tags to categorize domains. Users can use the tags to search for the domains. You also select the domain type to indicate how the domain is used.

You can also organize domains by geographical location on an interactive map. The map is displayed on the Operational Insights home page. Assigning domains to locations helps you analyze performance and determine capacity and processing capabilities across the enterprise.

1. Assign tags to the domain under Domain Details. You can assign multiple tags to a domain.
   - To assign an existing tag, select it from the list.
   - To add a new tag, type the tag in the entry field, then click the **Enter** key on your keyboard.
2. Select the domain type that best matches how the domain is used within the organization..

3. Specify the location of the domain on the map.
   - To assign the domain to an existing location, click the location on the map.
   - To assign the domain to a new location, click where you want to add the location on the map, then type in the location name.
4. Click **Next** to save your entries.

Configuring the Domain Configuration Collector

Configure the Domain Configuration Collector, which collects and uploads configuration metadata for the domain and all domain assets.

The default collection frequency is every 24 hours. You can create a custom schedule if needed to better suit your requirements.

The collector is enabled by default. You cannot disable the collector.

Configuring the collector schedule

You can configure a custom schedule for the collector. The schedule you create overrides the default collector schedule.

Enter the following properties:

Property	Description
Repeats	The interval at which to repeat collection.
Repeats Frequency	The frequency at which to perform collection. The frequency is based on the repetition value you select. For example, to collect data every two hours, select Hourly as the repetition, and then set the frequency value to 2.
Starts on	The date and time the custom schedule takes effect.
Timezone	The timezone the schedule is based on.

Configuring the Domain Health Statistics Collector

Configure the Domain Health Statistics Collector, which collects and uploads availability statistics for domain assets.

The default collection frequency is every 5 minutes. You can create a custom schedule if needed to better suit your requirements.

The collector is enabled by default. Clear the **Enabled** checkbox to diable the collector.

Configuring the collector schedule

You can configure a custom schedule for the collector. The schedule you create overrides the default collector schedule.

Enter the following properties:

Property	Description
Repeats	The interval at which to repeat collection.
Repeats Frequency	The frequency at which to perform collection. The frequency is based on the repetition value you select. For example, to collect data every two hours, select Hourly as the repetition, and then set the frequency value to 2.
Starts on	The date and time the custom schedule takes effect.
Timezone	The timezone the schedule is based on.

Configuring the Domain Resource Usage Statistics Collector

Configure the Domain Resource Usage Statistics Collector, which collects and uploads CPU and memory consumption metrics for all nodes within the domain.

The metrics are extracted from the Model repository specified in the Monitoring Configuration for the domain. You must configure the connection to the Model repository.

The default collection frequency is every 1 hour. You can create a custom schedule if needed to better suit your requirements.

The domain is enabled by default. Clear the **Enabled** checkbox to disable the collector.

Collecting historical data

You can configure the collector to populate Operational Insights with up to 60 days of historical data.

Historical data is collected for the previous 30 days by default. However you can specify any number of days between 1 and 60.

Data collection begins at the time the domain is added to Operational Insights. Roughly 24 hours worth of data is collected every hour, meaning that approximately 30 hours are required to populate Operational Insights with data for the prior month.

Historical data collection is enabled by default. Clear the **Collect Historical Data** checkbox to disable historical data collection.

Connecting to the Monitoring Statistics Model repository

Enter the information required to collect CPU and memory consumption metrics for domain nodes from the Model repository associated with the monitoring Model Repository.

You only need to enter the connection information when configuring either the Domain Resource Usage Statistics Collector or the Big Data Collector. The same information is used by both collectors.

Use the Administrator tool that you use to manage the domain to locate the required property values:

- Select the **Services and Nodes** view.
- Click the **Monitoring Configuration** tab, and then note the name of the Model Repository Service.
- Select the Model Repository Service instance in the Domain Navigator, and then note the properties listed under Repository Database Properties.

After you locate the required property values, complete the following steps configure the connection to the Model repository.

1. Enter the following required properties:

Property	Description
Database Type	Model repository database type.
Username	User name for the Model repository database.
Password	Password for the Model repository database.
JDBC Connection String	JDBC connection string used to connect to the Model repository database.

2. Enter the following optional properties:

Property	Description
Secure JDBC Parameters	Secure database parameters if the Model repository database is secured with the SSL protocol.
Schema Name	Schema name in the Model repository database that contains monitoring data.
Tablespace Name	If the Model repository database is an IBM DB2 database, you can specify the name of the tablespace that contains monitoring data.

3. Click **Test Connection** to verify the connection settings.

Configuring the collector schedule

You can configure a custom schedule for the collector. The schedule you create overrides the default collector schedule.

Enter the following properties:

Property	Description
Repeats	The interval at which to repeat collection.
Repeats Frequency	The frequency at which to perform collection. The frequency is based on the repetition value you select. For example, to collect data every two hours, select Hourly as the repetition, and then set the frequency value to 2.
Starts on	The date and time the custom schedule takes effect.
Timezone	The timezone the schedule is based on.

Configuring the PowerCenter Repository Collector

Configure the PowerCenter Repository Collector, which collects and uploads runtime workflow and session metrics from PowerCenter repositories within the domain. You must select PowerCenter in the list of products used by the domain on the Domain Connection panel to configure the collector.

You must configure the connection to each PowerCenter repository database in the domain.

The default collection frequency is every 1 hour. You can create a custom schedule if needed to better suit your requirements.

The collector is disabled by default. Select the **Enabled** checkbox to enable the collector.

Collecting historical data

You can configure the collector to populate Operational Insights with up to 60 days of historical data.

Historical data is collected for the previous 30 days by default. However you can specify any number of days between 1 and 60.

Data collection begins at the time the domain is added to Operational Insights. Roughly 24 hours worth of data is collected every hour, meaning that approximately 30 hours are required to populate Operational Insights with data for the prior month.

Historical data collection is enabled by default. Clear the **Collect Historical Data** checkbox to disable historical data collection.

Adding a PowerCenter repository

If the domain is a PowerCenter domain, configure a connection to each PowerCenter repository database within the domain. Supply the JDBC connection string used to connect to the PowerCenter repository database. You can optionally provide parameters required to connect to a secure database.

1. Click **Add PowerCenter Repository**.

2. Enter the following required properties:

Property	Description
Database Type	PowerCenter repository database type.
Service Name	Name of the PowerCenter Repository Service that manages the PowerCenter repository database.
Username	User name for the PowerCenter repository database.
Password	Password for the PowerCenter repository database.
JDBC Connection String	JDBC connection string used to connect to the PowerCenter repository database.

3. Enter the following optional properties:

Property	Description
Secure JDBC Parameters	Secure database parameters if the PowerCenter repository database is secured with the SSL protocol.
Schema Name	Schema name in the PowerCenter repository database that contains monitoring data. To find this value in the Administrator tool, select the Services and Nodes view, and then select the PowerCenter Repository Service instance in the Domain Navigator.
Table Name	Table in the PowerCenter repository database that contains monitoring data. To find this value in the Administrator tool, select the Services and Nodes view, and then select the PowerCenter Repository Service instance in the Domain Navigator.

4. Select the **Enable Repository** checkbox to enable the collector to collect data from the repository.
5. Click **Test Connection** to verify the connection configuration.
6. Click **Save** to save the connection details.
7. Repeat this process for each PowerCenter repository in the domain.

Configuring the collector schedule

You can configure a custom schedule for the collector. The schedule you create overrides the default collector schedule.

Enter the following properties:

Property	Description
Repeats	The interval at which to repeat collection.
Repeats Frequency	The frequency at which to perform collection. The frequency is based on the repetition value you select. For example, to collect data every two hours, select Hourly as the repetition, and then set the frequency value to 2.

Property	Description
Starts on	The date and time the custom schedule takes effect.
Timezone	The timezone the schedule is based on.

Configuring the Big Data Collector

Configure the Big Data collector to collect and upload statistics on Hadoop clusters used by the domain, including statistics on jobs run on the clusters. You must select Big Data in the list of products used by the domain on the Domain Connection panel to configure the collector.

The default collection frequency is every 1 hour. You can create a custom schedule to suit your requirements.

The collector is enabled by default. Clear the **Enabled** checkbox to disable the collector.

Click **Finish** when you finish configuring the collector.

Collecting historical data

You can configure the collector to populate Operational Insights with up to 60 days of historical data.

Historical data is collected for the previous 30 days by default. However you can specify any number of days between 1 and 60.

Data collection begins at the time the domain is added to Operational Insights. Roughly 24 hours worth of data is collected every hour, meaning that approximately 30 hours are required to populate Operational Insights with data for the prior month.

Historical data collection is enabled by default. Clear the **Collect Historical Data** checkbox to disable historical data collection.

Selecting the cluster configuration

If the domain is a Big Data domain, select the cluster configuration used by the domain to connect to the Hadoop cluster. The Big Data collector uses the cluster configuration to gather job execution statistics and operational metrics for the cluster.

You can view the cluster configurations created in the domain in the **Connections** tab in Informatica Administrator (the Administrator tool).

1. Click **Select Cluster Configuration**.
2. Select the cluster configuration to use to connect to the Hadoop cluster from the menu.
3. Select the **Enable Cluster Configuration** checkbox to enable the collector to collect data from the cluster.
4. To connect to a secure cluster, click **TLS Enabled**, and then specify the path and password for the cluster truststore file.
5. Click **Save** to save the configuration.

Configuring the collector schedule

You can configure a custom schedule for the collector. The schedule you create overrides the default collector schedule.

Enter the following properties:

Property	Description
Repeats	The interval at which to repeat collection.
Repeats Frequency	The frequency at which to perform collection. The frequency is based on the repetition value you select. For example, to collect data every two hours, select Hourly as the repetition, and then set the frequency value to 2.
Starts on	The date and time the custom schedule takes effect.
Timezone	The timezone the schedule is based on.

Finalizing the on-boarding configuration

After you complete the on-boarding process, the collectors begin collecting and uploading data to the Informatica Intelligent Cloud Services for use by Operational Insights.

Click **Save** to complete the domain on-boarding process and begin collecting data.

Searching for domains

You can search for specific domains using attributes or tags assigned to domains as search parameters.

Specify domain attributes as key:value pairs. For tags, just supply the tag value. Separate multiple parameters with a comma.

For example, enter this query to search for a domain in London that is assigned the tag "Production":

```
loc:London,Production
```

The domains matching your search criteria dynamically appear in the page.

Viewing and editing domain configuration details

Use the Details page to view and edit the configuration details for a domain, including the Secure Agent and auto-scaling configurations. You can also edit collector configurations, view collection history, and start an on demand collection.

To view or edit the configuration details for a domain, select the domain, and then click the **Details** tab. You can perform the following tasks from the Details page:

Task	Description
Edit the domain connection and configuration details.	Select **Edit** from the menu in the Properties panel to modify the registration metadata for a domain.
Create or edit an auto-scaling configuration.	Click **Configure Now** to enable auto-scaling, or select **Edit** from the menu in the Auto-scaling panel to modify an existing auto-scaling configuration. See Chapter 9, "Auto-scaling PowerCenter grids in the cloud" on page 64 for details.
Edit the configuration for a collector.	Select **Edit** from the menu in the collector panel.
View recent collection activity for a collector.	Select **View History** from the menu in the collector panel.
Start an on demand data collection.	Select **Collect Now** from the menu in the collector panel.

Unregistering a domain

You can unregister a domain from Operational Insights.

When you unregister a domain, all of the collected operational data is deleted and cannot be recovered.

If auto-scaling is enabled for the domain, the auto-scaling configuration is also deleted. However, any elastic nodes running in the cloud are not removed. You must manually remove the elastic nodes from the cloud.

1. Click the **Overview** tab.
2. Click a domain.

   You might need to first select a location, then select a domain within the location.
3. Click the **Analytics** tab.
4. From the edit menu, select **Unregister Domain**.

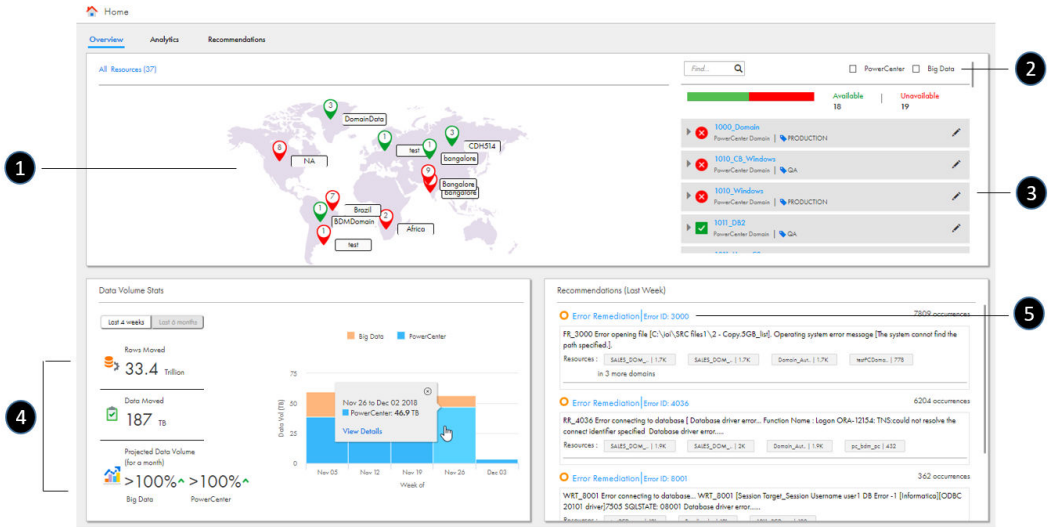
CHAPTER 5

Monitoring the enterprise

This chapter includes the following topics:

Viewing enterprise health

Use the Overview page to assess overall usage and health of your Informatica assets. The page provides you with a comprehensive overview of your Informatica deployments.



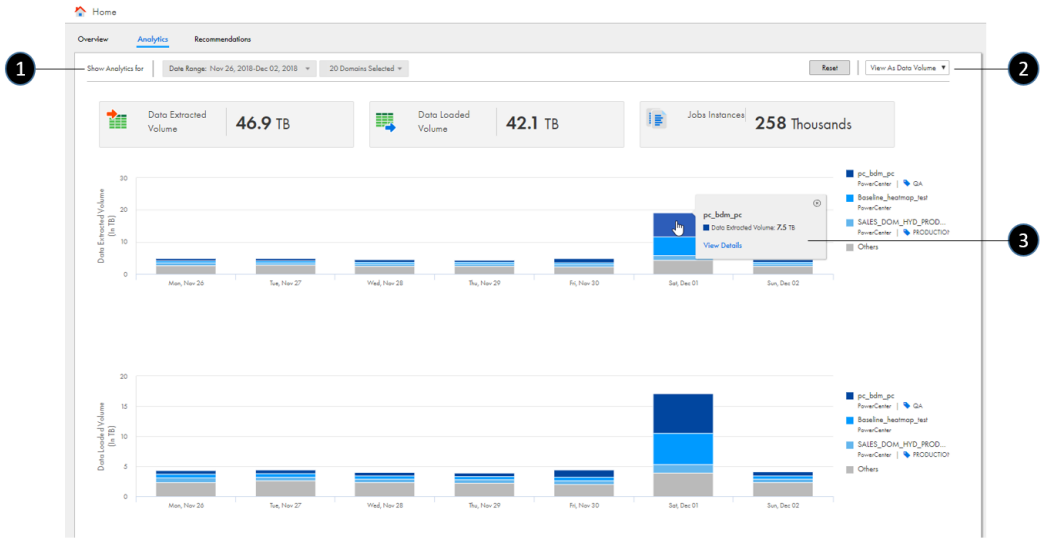
You can perform the following tasks from this page:

Task	Description
1	Click a location to view the status of the domains within the location. The domains within the location appear in the panel to the right.
2	You can search for specific domains using attributes or tags assigned to domains as search parameters. See "Searching for domains" on page 28 for details.
3	Expand a domain to view its details, and then click the domain name to view data processing and workflow statistics for the nodes in the domain. See "Viewing Big Data domain analytics" on page 37 or "Viewing PowerCenter domain analytics" on page 45 for details. Click the pencil icon to edit the domain registration data or to modify a collector configuration.
4	Review data movement statistics for the last four weeks or the last six months for all domains in the enterprise. You can also see the projected data volume for the next 7 days, which is calculated based on recent processing trends. Click the segment in the bar chart corresponding to the product domains and time period you want to view analytics for. To view detailed analytics, click the **View Details** link to view the Analytics page for the selected domain type and time period. Use this data to help you assess usage of your Informatica deployments and to determine future capacity requirements. See "Viewing data processing statistics" on page 32 for details.
5	Scroll through the top 10 recommendations generated for all domains across the enterprise. The recommendations shown are for the most frequently occurring errors over the past 7 days. Click a recommendation title to view the complete recommendation. See "Viewing recommendations" on page 32 for details.

Viewing data processing statistics

Use the Analytics page to view summary data processing statistics for all domains in the enterprise. Use this data to assess usage of your Informatica investments and determine if your current processing capacity is adequate.

Each bar in a chart represents a 24-hour period, based on the date range you select. The bar includes segments for the top three Big Data or PowerCenter domains across the enterprise, based on data processing and job instance analytics. The remaining domains are shown in the Others segment in the bar.



On the Home page, then do one of the following to view the Analytics page:

- Click the **Analytics** tab.
- Click the segment in the bar chart corresponding to the product domains and time period you want to view analytics for.

The following table lists the tasks you can perform from this page:

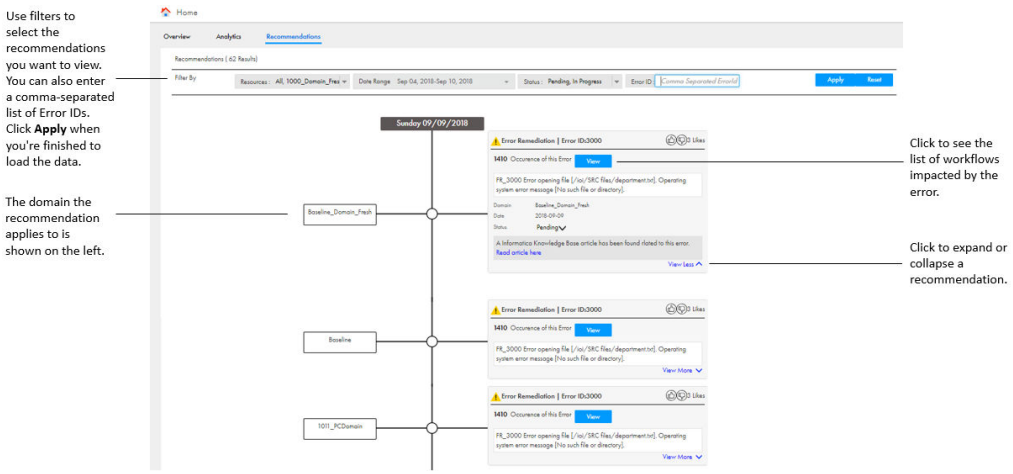
Task	Description
1	Select the date range and the domains you want to view data for. If you access the page from the bar chart on the home page, the page displays data for the selected bar chart segment.
2	Choose to view statistics based on data volume processed or on table rows moved.
3	Click a segment in a bar to view the total data processed or total job instances for the domain within the time period. To view detailed analytics, click the **View Details** link to view the Analytics page for the selected domain. See "Viewing Big Data domain analytics" on page 37 or "Viewing PowerCenter domain analytics" on page 45 for details.

Viewing recommendations

Use recommendations to improve performance, resolve errors, and avoid potential issues within your Informatica domains.

Error remediation recommendations are generated daily for all domains across the enterprise. The recommendations shown are for the most frequently occurring errors over the past 7 days. Recommendations include links to Informatica Knowledge Base articles related to the error code reported in the recommendation.

1. Click the **Recommendations** tab.



2. Use filters to select the domain, date range and status you want to view recommendations for. You can also enter one or more comma-separated Error ID values to filter by error code.

   The page displays up to 25 recommendations. Click **View More** to view the next set of recommendations.

3. Click **View** in a recommendation card to view the workflows impacted by the error.

   A dialog box listing the workflows opens. Click a workflow to view details.

4. Click **View More** within a recommendation card to view additional details. You can expand multiple recommendations at the same time.

5. Rate the recommendation.

   - If you click the thumbs up icon, the number of Likes is increased by 1.
   - If you click the thumbs down icon, a comment dialog opens so you can explain why you gave the recommendation a poor rating.
     Your feedback is not visible to other Operational Insights users. It is used by Informatica to improve the quality or usefulness of the recommendation.

6. Indicate the status of the recommendation to help you track your progress in resolving the error.

   If a recommendation is not applicable to you, select the **Dismiss** status to remove the recommendation from the list.

7. Click **Read article here** to open an Informatica Knowledge Base article related to the error code in a new browser.

Configuring alerts

You can configure Operational Insights to send email notifications when an issue occurs within a domain or with a Secure Agent.

You can configure alerts for the following events:

- The domain or the Secure Agent is unavailable.
- A collector, service, or node running within the domain is unavailable.
- CPU or memory consumption by a domain node, a domain service, or the Secure Agent host has crossed a configurable threshold.

You can enable or disable individual alerts or all alerts for each domain or Secure Agent that Operational Insights monitors. You also specify the email users or groups that receive alert notifications.

1. Click **Alert Settings** in the left navigation bar.
2. Select a domain or a Secure Agent.
3. Configure each alert you want to enable.
4. Enter the user names or group names that receive email notifications when an issue occurs.

Viewing Secure Agent statistics

Check the status of each Secure Agent used by your organization, as well as the status of the services that use the Secure Agent. You can also view the memory consumption and CPU usage by the processes that run on the Secure Agent host machine, so that you can take action before the machine runs out of memory or CPU capacity.

Secure Agent statistics are updated every 5 minutes. You can choose to view statistics for the last 24 hours, the last week, or the last month.

1. Click **Secure Agents** in the left navigation bar.
2. Select the Secure Agent you want to view statistics for.

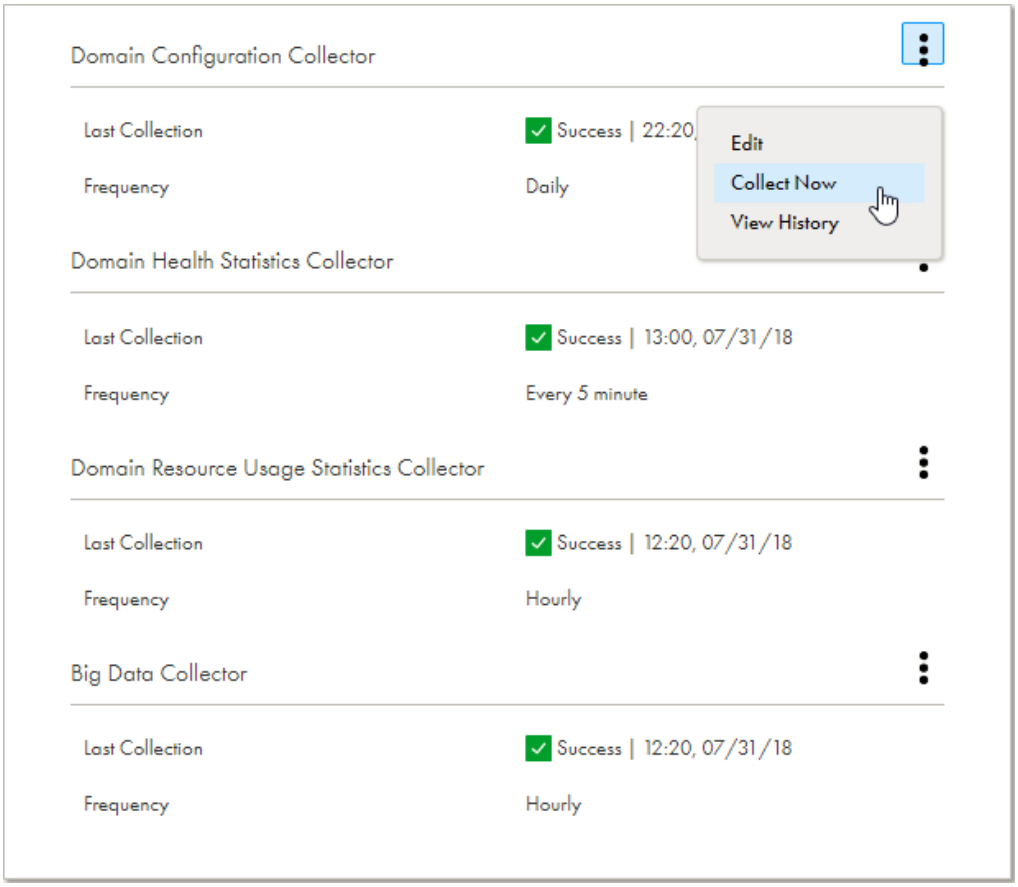
   To edit the configuration or view log data for the Secure Agent, click the **Manage** link.
3. Select the time period to view statistics for.
4. Move your cursor across the resource utilization graphs to view CPU usage and memory consumption details for specific points in time.

Managing collectors

For each collector configured for a domain, you can edit the collector configuration, view collector logs, and trigger an on-demand data collection.

1. Click the **Overview** tab.
2. Click a domain.
3. Click the **Details** tab.
4. From the menu for a collector, select the following options:

- Select **Edit** to modify the collector configuration.
- Select **Collect Now** to trigger an on-demand data collection.
- Click **View History**, the click **View Logs** to view log data for a specific data collection.



Troubleshooting collectors

Review the log file for a collector to assess performance or troubleshoot issues.

The log file contains log data for all of the collectors running within the Secure Agent used by the domain. The file is generated in the following directory on the Secure Agent host:

```
<Secure Agent installation directory>\apps\OpsInsightsDataCollector\logs
```

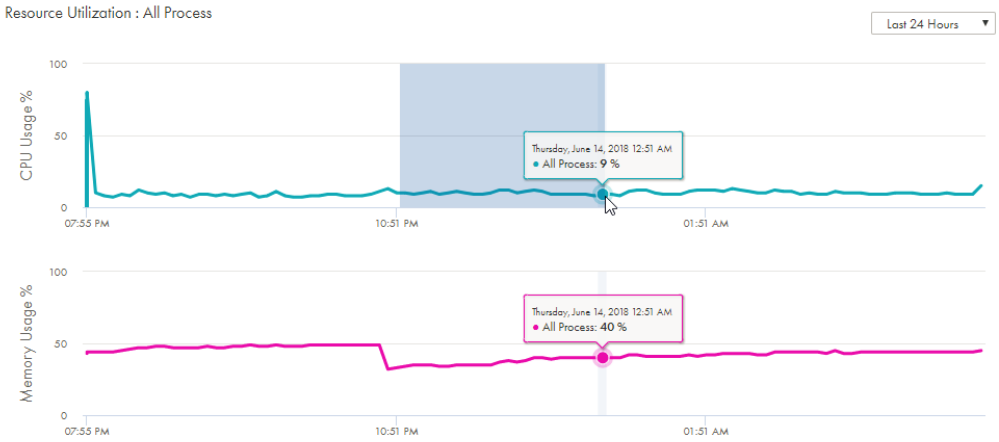
Zooming in on graph details

You can zoom in on resource utilization graphs to view details for a specific time frame. You select the start and end times to zoom in on in the graphs.

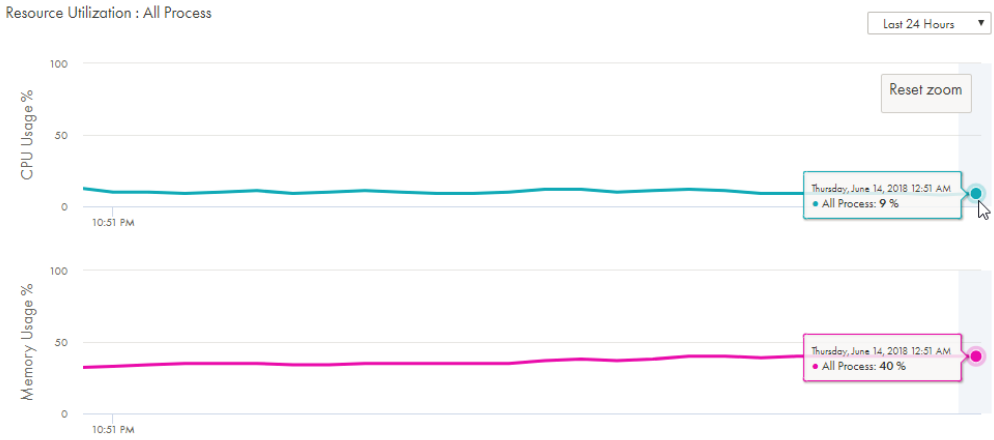
1. Place the cursor on the time frame start point in a graph.

2. Left-click your mouse.
3. Drag the cursor to the time frame end point in the graph.

   The following image shows the time frame from 10:51 p.m. to 12:51 p.m. selected:



   The resource utilization graphs update to display data only for the specified time frame, as shown in the following image:



4. Click **Reset zoom** to return the graphs to the original state.

CHAPTER 6

Monitoring Big Data domains

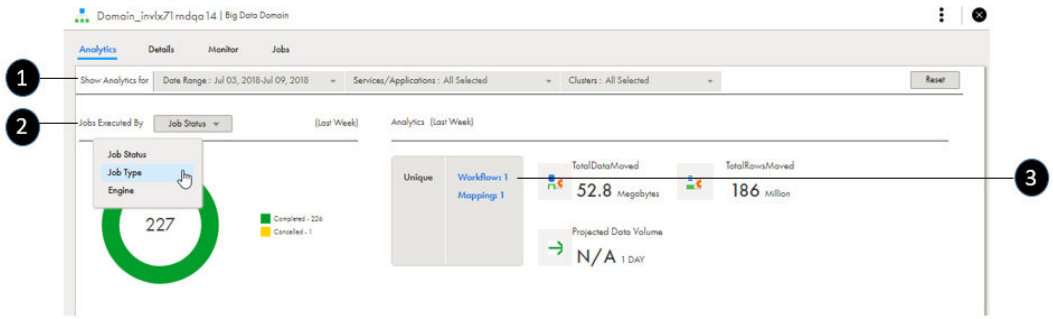
This chapter includes the following topics:

Viewing Big Data domain analytics

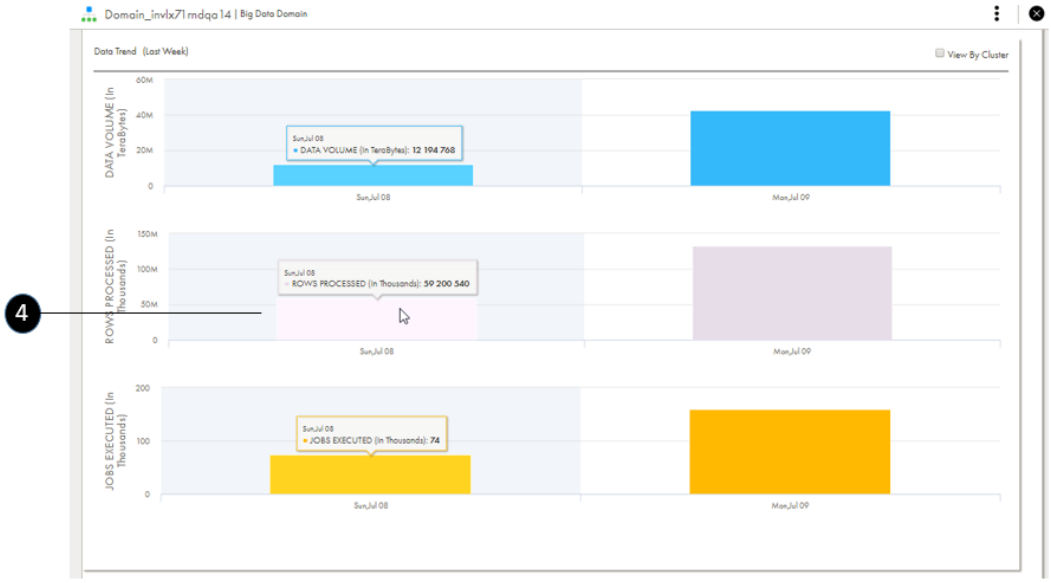
Use the Domain Analytics page to view job execution statistics and data processing trends for a specific Big Data domain.

Select a Big Data domain, then click the **Analytics** tab to view the page.

The upper part of the page displays job execution and data processing statistics for the selected date range, Data Integration Service instances, and clusters. By default, the table displays data for mapping and workflow jobs run by Data Integration Service instances submitting jobs to registered clusters over the last seven days.



The lower portion of the page displays data processing trends for the same selections:



The following table lists the tasks you can perform from this page:

Task	Description
1	Use filters to select the date range, Data Integration Service instances, and Hadoop clusters to view data for. The data displayed in the page is updated based on your filter settings. By default, data for the last seven days is shown for all Data Integration Service instances and Hadoop clusters in the domain.
2	Choose whether to view jobs by job completion status, job type, or execution engine. Click on **Completed** or **Canceled** to view job details, and then click the pie chart to view the selected jobs.
3	Click a link to view details for unique workflows and mappings for the selected date range, services and clusters. See "Viewing Big Data job analytics" on page 38 for additional details.
4	Move your cursor across the chart to view details on the total amount of data moved during the time range specified in the filter. Select the **View by Cluster** checkbox to view statistics for each cluster.

Viewing Big Data job analytics

Use Big Data job execution statistics to assess job execution performance, identify failed and long-running jobs, and troubleshoot issues.

Use filters to drill down on the job execution data you want to view. The data displayed in each page is based on the combined filters you set.

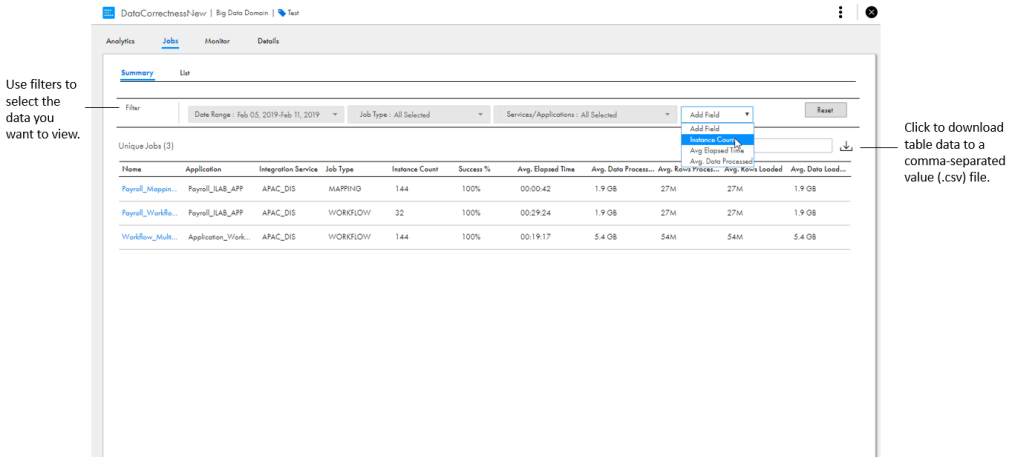
Viewing job execution summary data

Use summary data collected on job run executions to gain insight into job run performance. You can view the overall performance of jobs for specified time ranges, by Data Integration Service, and by job type.

1. Click the **Overview** tab.
2. Click a domain.

   You might need to first select a location, then select a domain within the location.
3. Click the **Jobs** tab.
4. Click the **Summary** tab.
5. Use filters to select the date range, Data Integration Service that runs jobs, or job type to view data for. You can select additional columns to filter on from the **Add Field** menu.

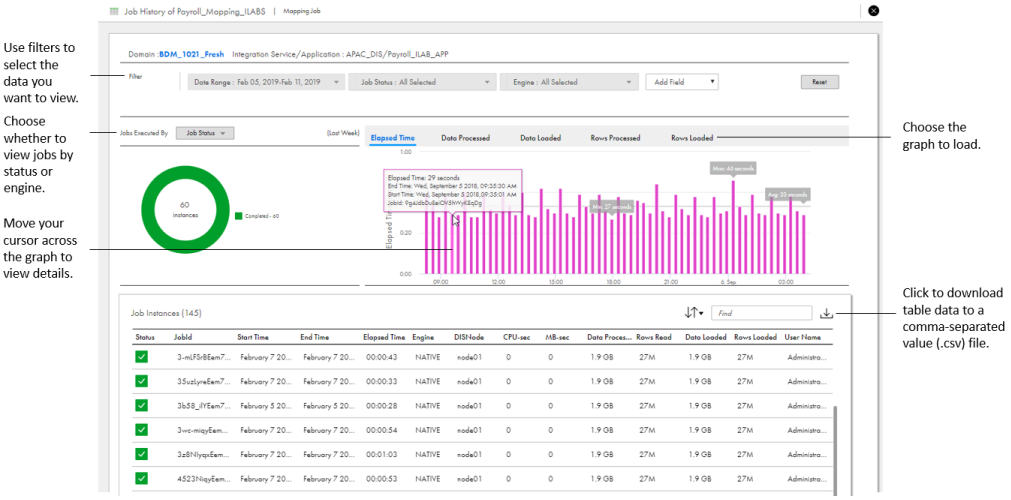
   The table updates according to your filter settings. By default, the table displays data for mapping and workflow jobs run by Data Integration Service instances submitting jobs to registered clusters over the last seven days. You can download the table data to a comma-separated value (.csv) file.



6. Click a job in the table to view details.

   A page displaying statistics for individual job instances loads. Use filters to select the data you want to view. You can select additional columns to filter on from the **Add Field** menu.

7. A graph displaying job run statistics loads in the page. You can view statistics by elapsed run time, amount of data processed, the amount of data read from source rows read, and the amount of data written to target rows. The charts display data based on the filters you set.



Move your cursor across the graph to view specific details. You can zoom in on resource utilization graphs to view details for a specific time frame. See "Zooming in on graph details" on page 35 for details.

Viewing job instance performance data

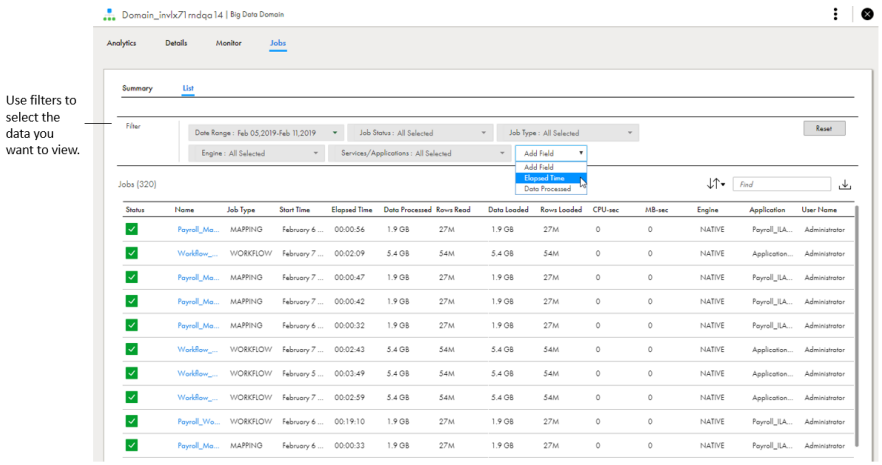
Use execution data for job instances that run in the domain to uncover performance issues. Data collected for each job includes whether the job succeeded or failed, job run start and end times, and total data processed.

By default, the table displays data for mapping and workflow jobs run by Data Integration Service instances submitting jobs to execution engines on all registered clusters over the last seven days. You can download the table data to a comma-separated value (.csv) file.

1. Click the **Overview** tab.
2. Click a Big Data domain.

   You might need to first select a location, then select a domain within the location.
3. Click the **Jobs** tab.
4. Click the **List** tab.
5. Use filters to select jobs to view data for. To add additional filters to the table, click **Add Field**, then select the filter column to add.

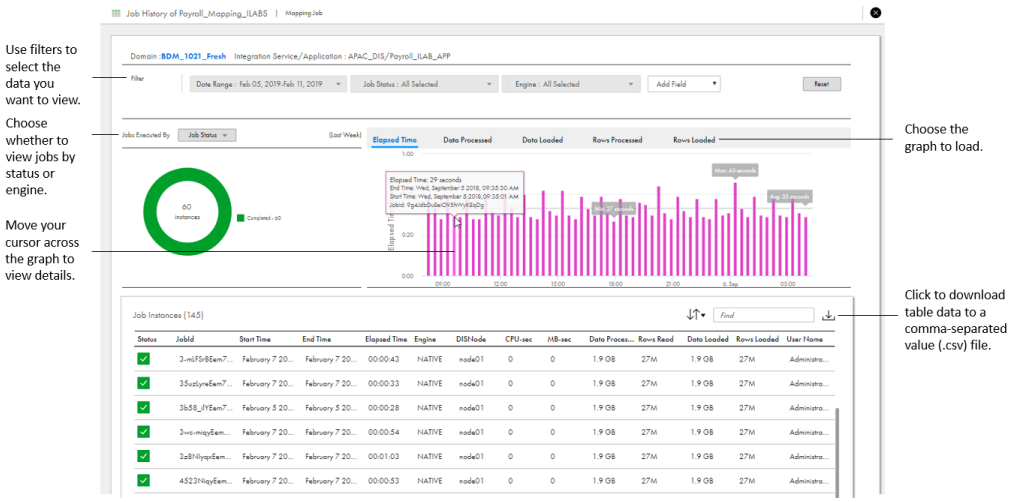
The table updates according to your filter settings.



6. Click a job in the table.

   A page displaying statistics for individual job instances loads. Use filters to select the data you want to view. You can select additional columns to filter on from the **Add Field** menu.

7. A graph displaying job run statistics loads in the page. You can view statistics by elapsed run time, amount of data processed, the amount of data read from source rows read, and the amount of data written to target rows. The charts display data based on the filters you set.



   Move your cursor across the graph to view specific details. You can zoom in on resource utilization graphs to view details for a specific time frame. See "Zooming in on graph details" on page 35 for details.

Viewing Big Data domain resource usage analytics

Use CPU and memory consumption metrics to trace job execution performance issues.

Viewing resource utilization for nodes, application services, and grids in a Big Data domain

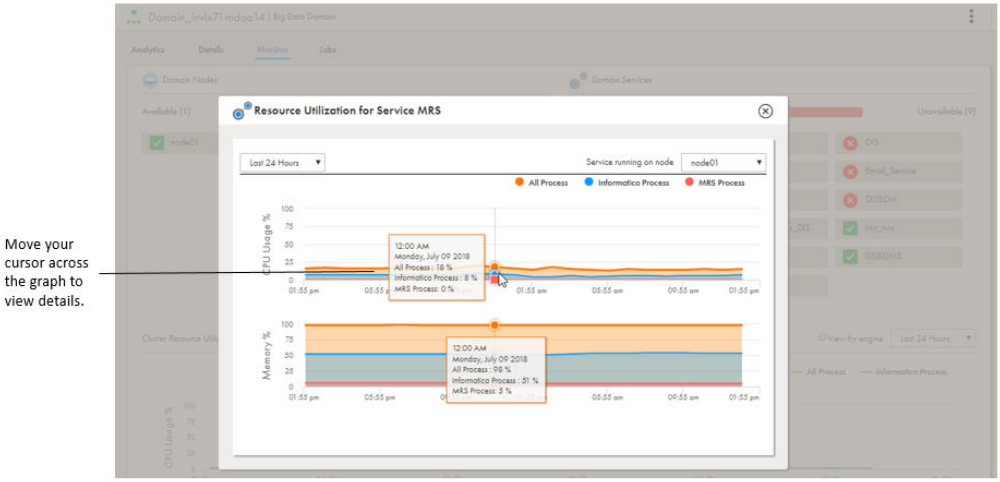
Use the Domain Resource Utilization graph to view memory and CPU usage statistics for individual nodes within a domain, as well as for application services running a specific node. You can also view statistics for Data Integration Service grids running within the domain.

You might want to consider adding more computing resources within a domain if CPU or memory utilization on a node exceeds 75%.

1. Click the **Overview** tab.
2. Click a Big Data domain.

   You might need to first select a location, then select a domain within the location.
3. Click the **Monitor** tab.
4. Click an Informatica node or an application service.

   A graph displaying CPU and memory consumption for the selected node, application service, or domain grid loads.



   When you select an application, consumption data is shown for all processes running on the node, including the application service.
5. Select the time period to view details for in the graph. You can choose to view statistics for the last 24 hours, the last week, or the last month.
6. Move your cursor across the graph to view specific details.

   You can zoom in on resource utilization graphs to view details for a specific time frame. See "Zooming in on graph details" on page 35 for details.

Viewing resource utilization for clusters

View resource consumption for processes that run on Hadoop cluster nodes. You can also view usage statistics for the execution engines that run on the cluster.

1. Click the **Overview** tab.
2. Click a Big Data domain.

   You might need to first select a location, then select a domain within the location.
3. Click the **Monitor** tab.

4. Select the cluster configuration for the cluster from the Cluster Resource Utilization menu.

5. To show utilization by execution engine, select the **View by Engine** checkbox.

   The graph updates to show data for each execution engine.

6. Select the time period to view details for.
7. The CPU Usage % graph displays by default. To add graphs showing memory utilization and the number of cluster nodes, select the graph to add, then click **Add Insight**.

   To view a graph displaying data based on the number of Hadoop cluster nodes that jobs ran on, select **Nodes** from the menu.

8. Move your cursor across the graph to view specific details.

   You can zoom in on resource utilization graphs to view details for a specific time frame. See "Zooming in on graph details" on page 35 for details.

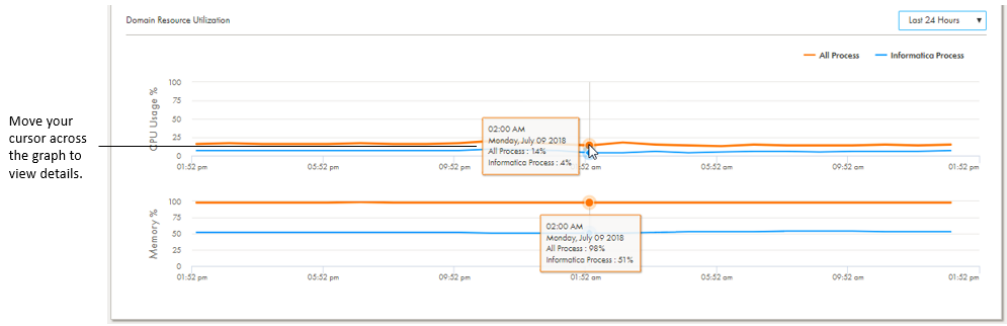
Viewing resource utilization for a Big Data domain

Use the Domain Resource Utilization graph to view memory and CPU usage for all processes and Informatica processes running on node within the domain.

1. Click the **Overview** tab.
2. Click a Big Data domain.

   You might need to first select a location, then select a domain within the location.

3. Click the **Monitor** tab.
4. Scroll to the Domain Resource Utilization graph at the bottom of the page.



5. Select the time period to view details for. You can choose to view statistics for the last 24 hours, the last week, or the last month.
6. Move your cursor across the graph to view specific details.

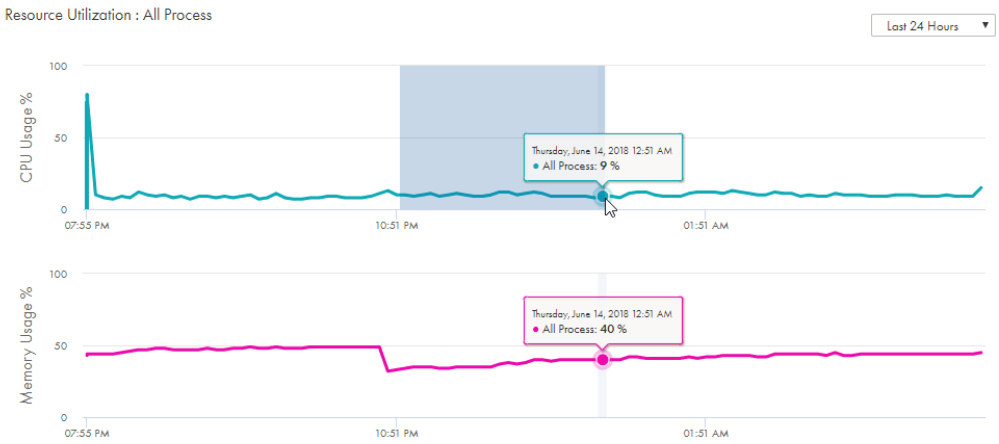
   You can zoom in on resource utilization graphs to view details for a specific time frame. See "Zooming in on graph details" on page 35 for details.

Zooming in on graph details

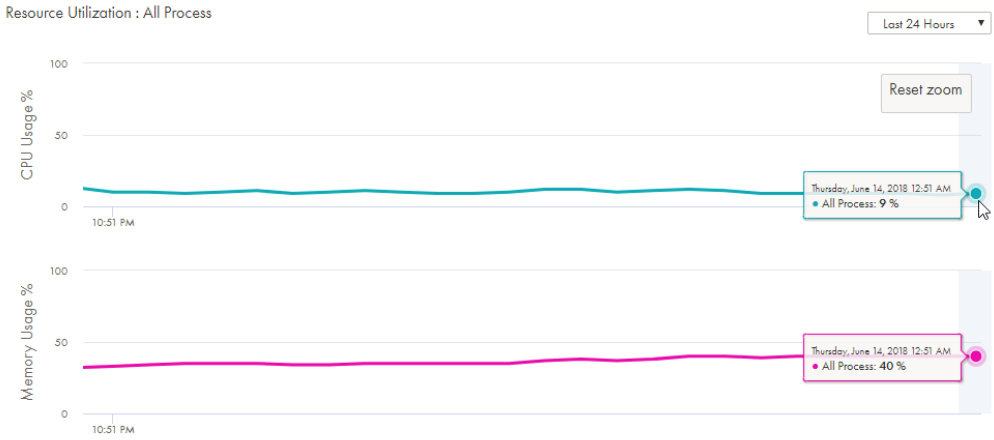
You can zoom in on resource utilization graphs to view details for a specific time frame. You select the start and end times to zoom in on in the graphs.

1. Place the cursor on the time frame start point in a graph.
2. Left-click your mouse.
3. Drag the cursor to the time frame end point in the graph.

   The following image shows the time frame from 10:51 p.m. to 12:51 p.m. selected:



   The resource utilization graphs update to display data only for the specified time frame, as shown in the following image:



4. Click **Reset zoom** to return the graphs to the original state.

CHAPTER 7

Monitoring PowerCenter domains

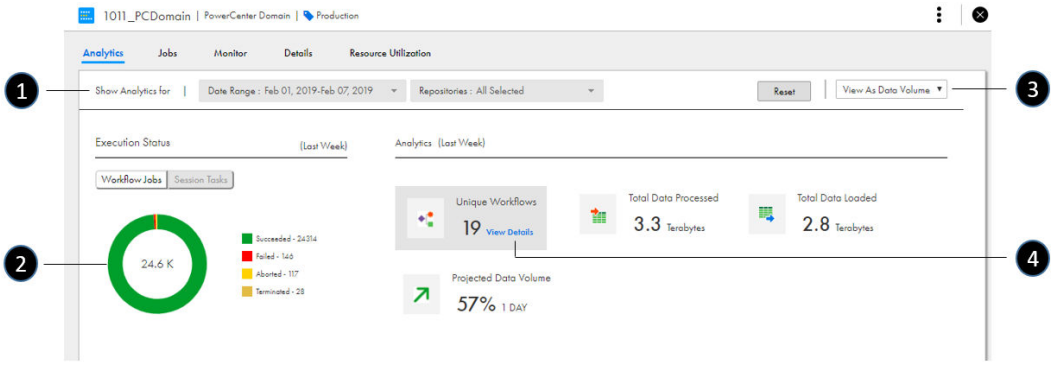
This chapter includes the following topics:

Viewing PowerCenter domain analytics

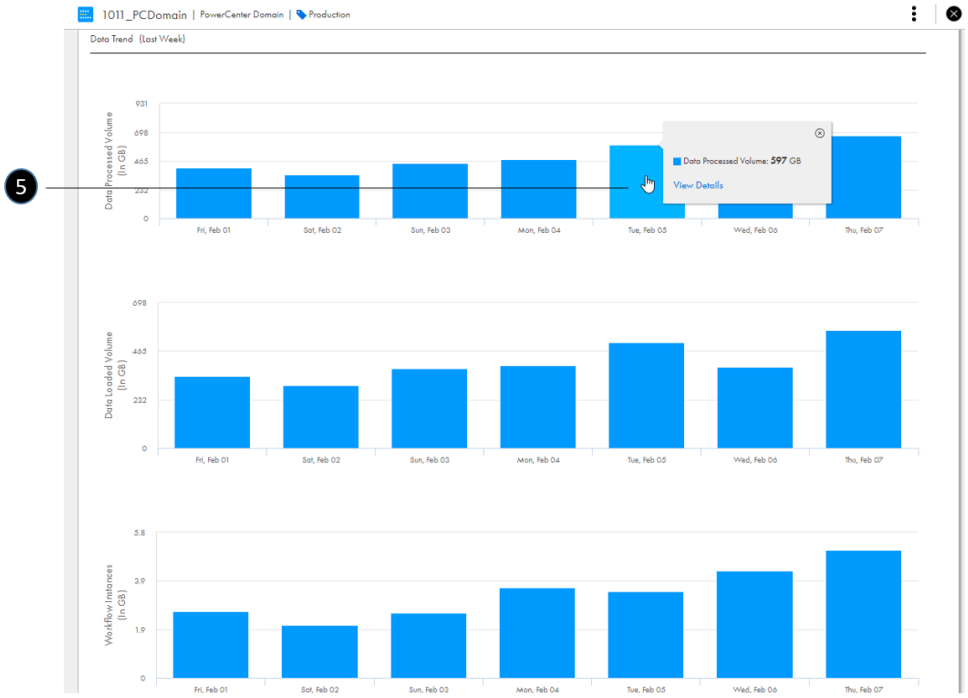
Use the Domain Analytics page to view workflow statistics and data processing trends for a specific domain.

Select a domain, then click the **Analytics** tab to view this page.

The upper portion of the page displays workflow execution and data processing statistics for the selected date range and repositories and folders. You can view data based on data volume processed or on total rows moved. By default, the page displays the data volume processed for workflow jobs run over the last seven days.



The graphs in the lower portion of the page display data processing and workflow data for the same selections:



The following table lists the tasks you can perform from this page:

Task	Description
1	Use filters to select the date range and repositories and folders to view data for. The data displayed in the page is updated based on your filter settings. See "Using PowerCenter repository filters" on page 56 for details. By default, data for the last seven days is shown for all repositories and folders in the domain.
2	Click on **Succeeded** or **Failed** to view job or session details. See "Viewing PowerCenter workflow analytics" on page 46 for details.
3	Choose whether to view data based on data volume processed or on total rows moved.
4	Click **View Details** in the Unique Workflows region to view detailed workflow instance statistics for the selected date range and repositories. See "Viewing PowerCenter workflow analytics" on page 46 for details.
5	Click a bar in the chart to view additional details.

Viewing PowerCenter workflow analytics

Use PowerCenter workflow statistics to assess workflow instance performance, identify failed and long-running workflow runs, and troubleshoot issues.

Use filters to drill down on the data you want to view. The data displayed in each page is based on the combined filters you set.

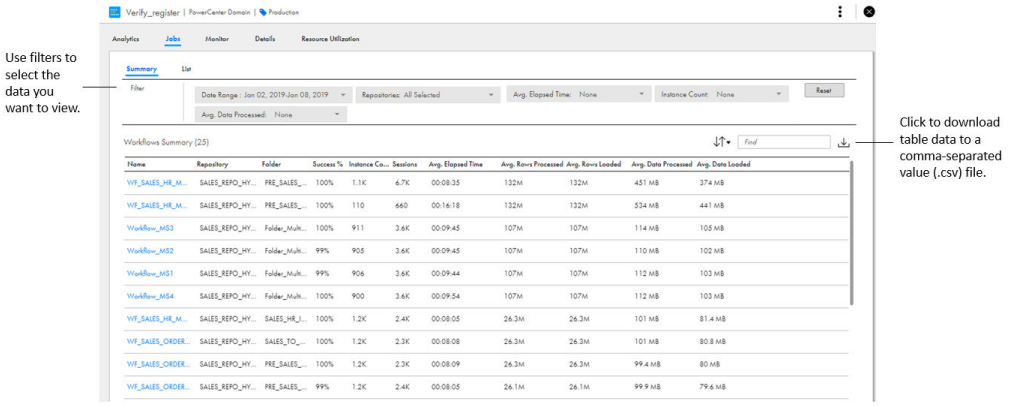
Viewing PowerCenter workflow execution summary data

Use summary data collected on workflow run executions to gain insight into job run performance. You can view the overall performance of workflows for specified time ranges, for specific repositories and folders, by instance count, and by average time elapsed and data processed.

1. Click the **Overview** tab.
2. Click a PowerCenter domain.

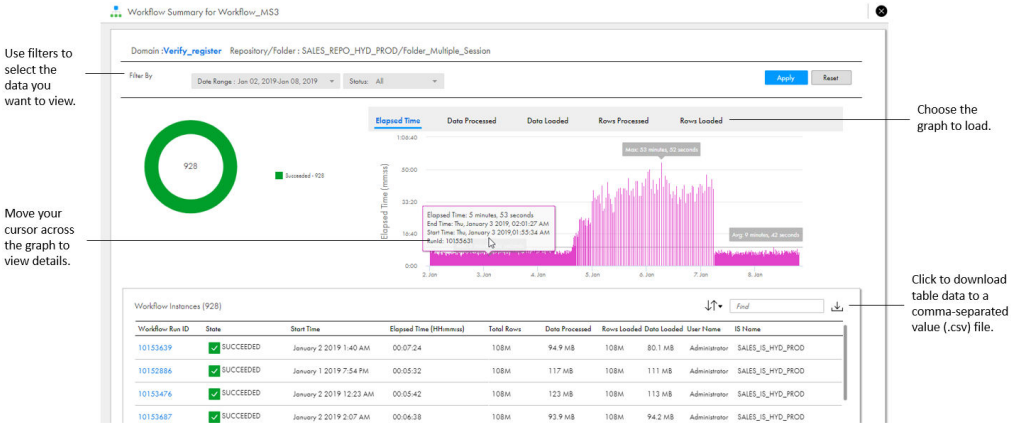
   You might need to first select a location, then select a domain within the location.
3. Click the **Jobs** tab.
4. Click the **Summary** tab.
5. Use filters to select the data to view.

   The table updates according to your filter settings. You can download the table data to a comma-separated value (.csv) file.



6. Click a workflow in the table to view details on workflow instances.

   A page displaying statistics for individual workflow instances loads. Use filters to select the data you want to view.



7. A graph displaying workflow run statistics loads in the page. You can view statistics by elapsed run time, amount of data processed, the amount of data read from source rows read, and the amount of data written to target rows. The charts display data based on the filters you set.

   Move your cursor across the graph to view specific details. You can zoom in on the graph to view details for a specific time frame. See "Zooming in on graph details" on page 35.
8. Click a Workflow Run ID to view task details.

Viewing PowerCenter workflow instance run analytics

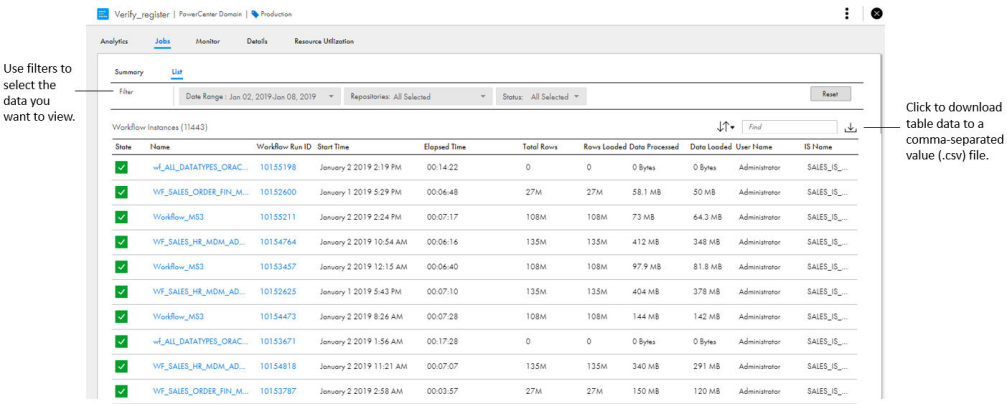
Use data for workflow instances that run in the domain to uncover performance issues. Data collected for each workflow instance includes whether the workflow instance run succeeded or failed, run start and end times, and total data processed.

You can view the overall performance of workflow instances for specified time ranges, for specific repositories and folders, by instance count, or by status. You can download the table data to a comma-separated value (.csv) file.

1. Click the **Overview** tab.
2. Click a PowerCenter domain.

   You might need to first select a location, then select a domain within the location.
3. Click the **Jobs** tab.
4. Click the **List** tab.
5. Use filters to select workflow instances to view data for.

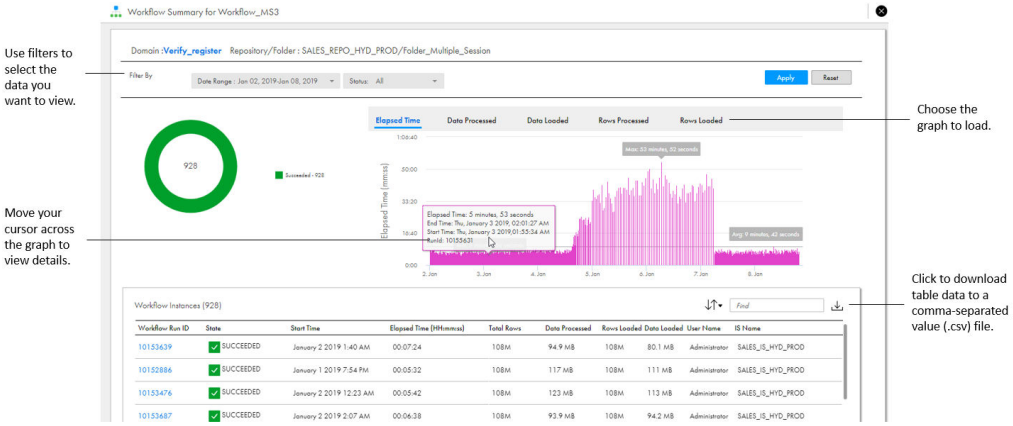
   The table updates according to your filter settings.



6. Click a workflow instance in the table to view details on workflow instances.

   A page displaying statistics for individual workflow instances loads. Use filters to select the data you want to view. You can select additional columns to filter on from the **Add Field** menu.

7. A graph displaying workflow run statistics loads in the page. You can view statistics by elapsed run time, amount of data processed, the amount of data read from source rows read, and the amount of data written to target rows. The charts display data based on the filters you set.



Move your cursor across the graph to view specific details. You can zoom in on the graph to view details for a specific time frame. See "Zooming in on graph details" on page 35.

Viewing anomalous workflow run behavior

Operational Insights leverages the CLAIRE engine, which employs statistical and machine learning approaches to detect data outliers and anomalies, to notify you about abnormal PowerCenter workflow run behavior.

Operational Insights analyzes elapsed run time data for workflows each day, and notifies you when anomalies are detected for specific workflow instances. You can then use the application to identify time periods during which a workflow took more or less time to run than normal, and use this data to determine the root cause.

Note: This feature is currently in preview mode.

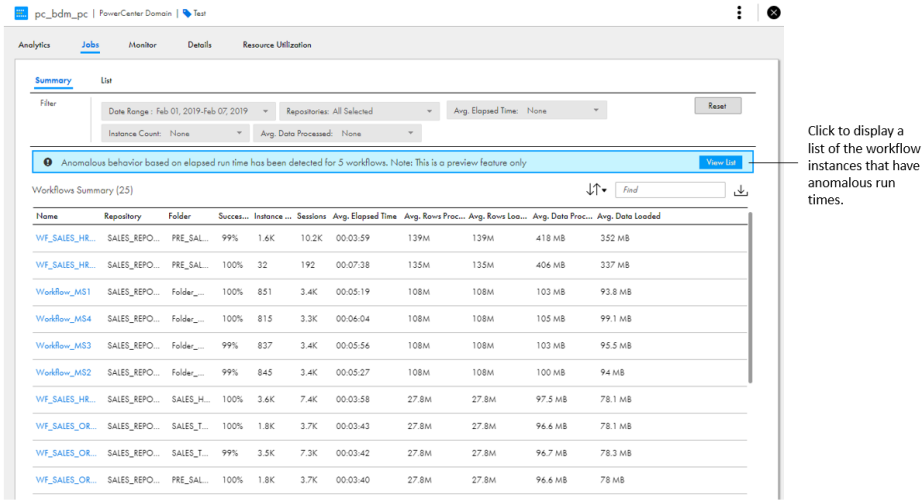
1. Click the **Overview** tab.
2. Click a PowerCenter domain.

   You might need to first select a location, then select a domain within the location.
3. Click the **Jobs** tab.
4. Click the **Summary** tab or the **List** tab.

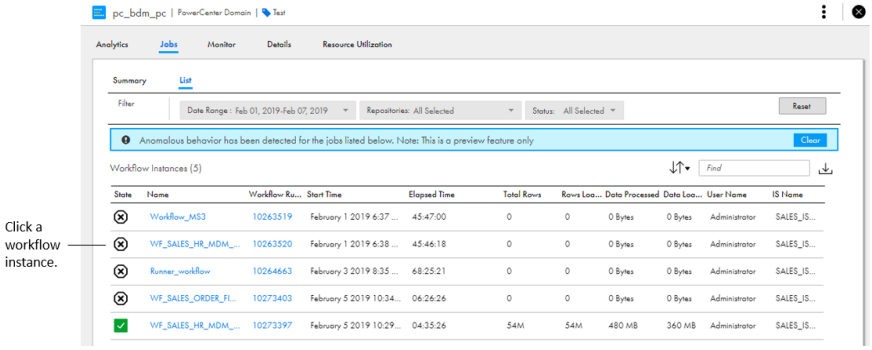
   When you click the **Summary** tab, the application displays a notification noting the number of workflows with workflow instances that have anomalous run times.

   When you click the **List** tab, the application displays a notification noting the number of workflow instances that have anomalous run times.

5. Click **View List** in the notification. The List page opens and displays a list of the workflow instances that have anomalous run times.



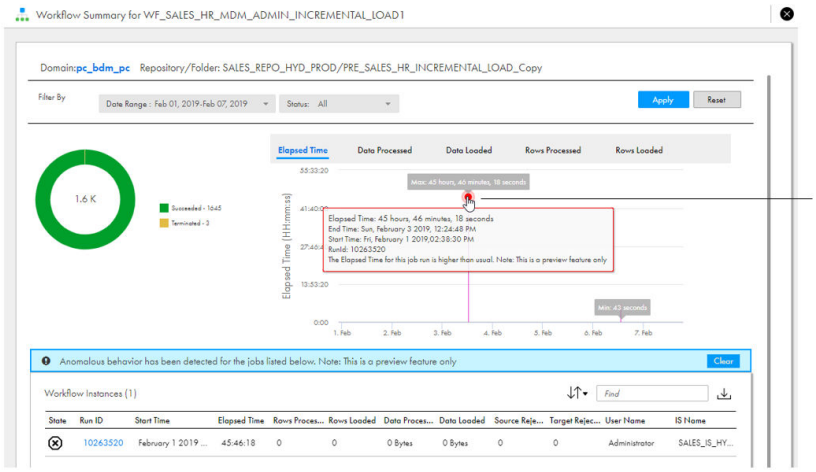
6. Click a workflow instance in the list.



The Workflow Summary page for the workflow opens. The graph contains maximum and minimum elapsed run time data for the workflow instance, and displays a red dot indicating the when the abnormal run occurred.

You can zoom in on the graph to view details for a specific time frame. See "Zooming in on graph details" on page 35.

7. Click the red dot in the graph to view the details for the associated tasks.



Click to view the details for the associated tasks.

Identifying failed workflows

You can quickly identify workflows that failed in the last 24 hours, or in the last seven days. You can also view errors indicating why a specific workflow failed.

1. Click the **Overview** tab.
2. Click a domain.

   You might need to first select a location, then select a domain within the location.
3. Click the **Analytics** tab.
4. Click **View Details** in the Unique Workflows region.

   The Workflow Trends page for the domain appears.
5. Click the number of failed workflows in the Workflow Insights region.

   By default, workflows that failed for the last 24 hours appear in the Workflow Summary table below. To view workflows that failed over the last seven days:

   - Open the Date Range filter.
   - Select **Last 30 Days**.
   - Click **OK**, then click **Apply**.
6. Select a failed workflow in the Workflow Summary table.

   A page showing aggregated data for all workflow run instances for the selected workflow appears. Use the filters to select the statistics you want to view. Click **Apply** when you're finished. The aggregated data shown on the page and in the graph is updated according to your filter settings.
7. Select a workflow run instance.

   A page displaying all of the Command, Email and Session tasks that ran within the workflow run instance appears. Check the Error column for a task to identify the reason the task failed.

Identifying workflows with increasing run times

You can quickly identify workflows that are taking longer to run over the past 24 hours, or over the last seven days.

1. Click the **Overview** tab.
2. Click a domain.

   You might need to first select a location, then select a domain within the location.
3. Click the **Analytics** tab.
4. Click **View Details** in the Unique Workflows region.

   The Workflow Trends page for the domain appears.
5. Click the number of workflows with increasing run times in the Workflow Insights region.

   By default, workflows that are taking longer to run over the last 24 hours appear in the Workflow Summary table below. To view workflows that are taking longer to run over the last seven days:

   - Open the Date Range filter.
   - Select **Last 30 Days**.
   - Click **OK**, then click **Apply**.
6. Click the Avg. Elapsed Time column in the Workflow Summary table to sort workflows by shortest or longest average run times.
7. Select a workflow in the Workflow Summary table.

   A page showing aggregated data for all workflow run instances for the selected workflow appears.
8. Click the Elapsed Time column to sort workflows by shortest or longest run times.

Viewing PowerCenter domain resource usage analytics

Use CPU and memory consumption metrics to trace workflow performance issues.

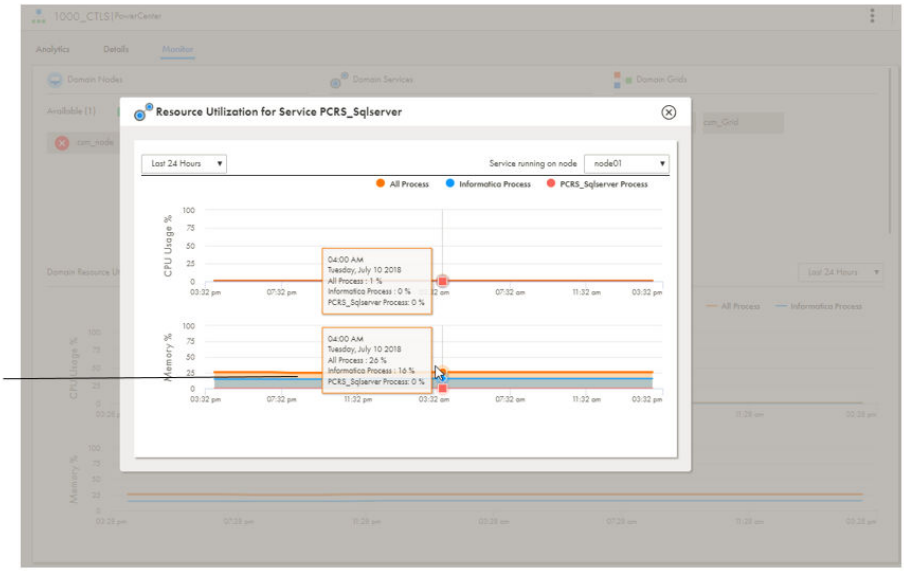
Viewing resource utilization for nodes, services and grids in a PowerCenter domain

Use the Domain Resource Utilization graph to view memory and CPU usage for Informatica nodes, Informatica application services, and PowerCenter grids running on nodes within a domain.

1. Click the **Overview** tab.
2. Click a PowerCenter domain.

   You might need to first select a location, then select a domain within the location.
3. Click the **Monitor** tab.
4. Click an Informatica node, an application service, or a PowerCenter grid.

A graph displaying CPU and memory consumption for the selected node or service loads.



When you select an application, consumption data is shown for all processes running on the node, including the application service. You might want to consider adding more computing resources within a domain if CPU or memory utilization on a node exceeds 75%.

5. Select the time period to view details for in the graph. You can choose to view statistics for the last 24 hours, the last week, or the last month.
6. Move your cursor across the graph to view specific details.

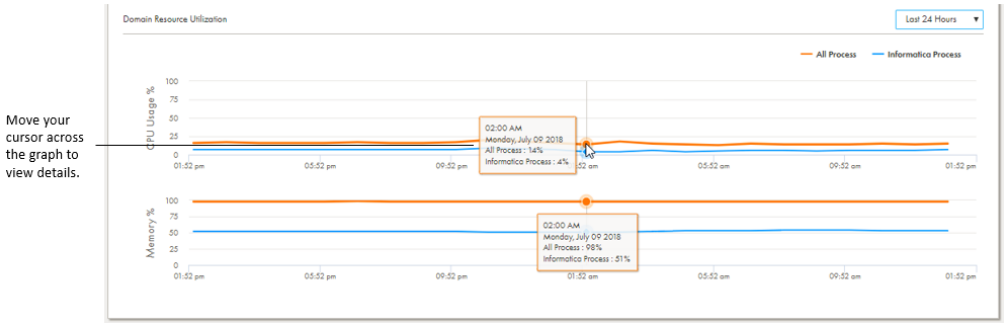
   You can zoom in on resource utilization graphs to view details for a specific time frame. See "Zooming in on graph details" on page 35 for details.

Viewing resource utilization for a PowerCenter domain

Use the Domain Resource Utilization graph to view memory and CPU usage for all processes and Informatica processes running on node within the domain.

1. Click the **Overview** tab.
2. Click a PowerCenter domain.

   You might need to first select a location, then select a domain within the location.
3. Click the **Monitor** tab.
4. Scroll to the Domain Resource Utilization graph at the bottom of the page.



5. Select the time period to view details for. You can choose to view statistics for the last 24 hours, the last week, or the last month.
6. Move your cursor across the graph to view specific details.

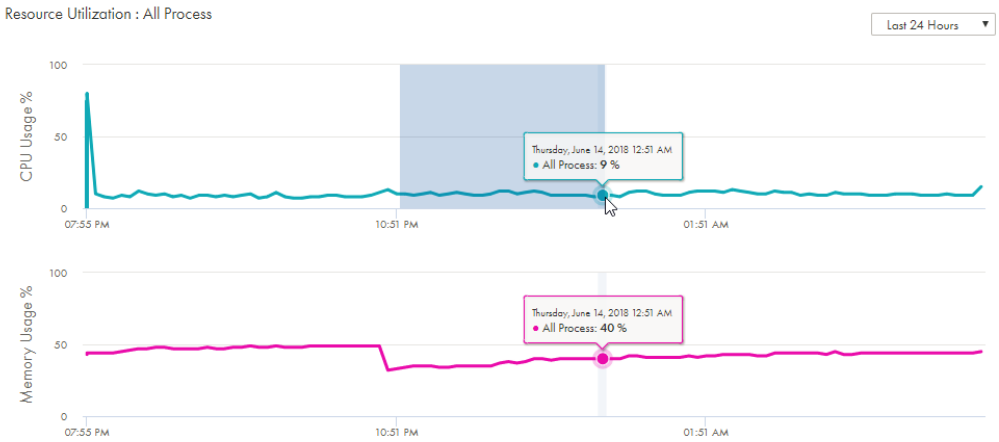
   You can zoom in on resource utilization graphs to view details for a specific time frame. See "Zooming in on graph details" on page 35 for details.

Zooming in on graph details

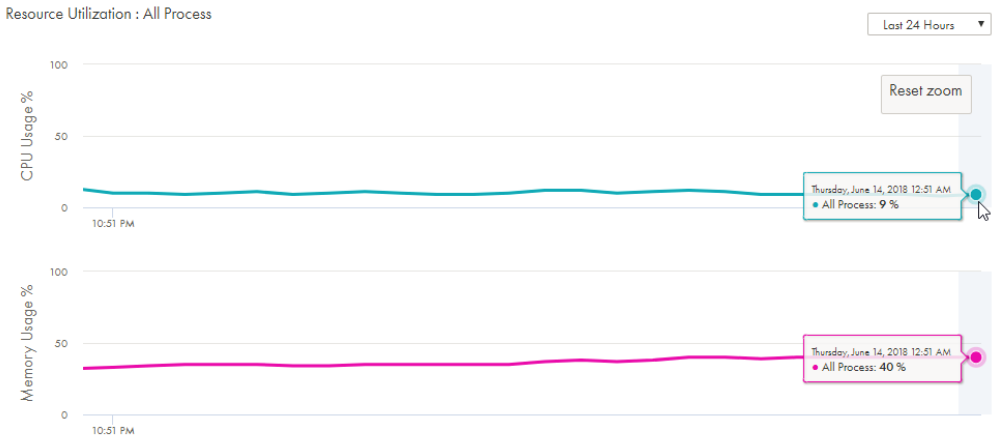
You can zoom in on resource utilization graphs to view details for a specific time frame. You select the start and end times to zoom in on in the graphs.

1. Place the cursor on the time frame start point in a graph.
2. Left-click your mouse.
3. Drag the cursor to the time frame end point in the graph.

   The following image shows the time frame from 10:51 p.m. to 12:51 p.m. selected:



   The resource utilization graphs update to display data only for the specified time frame, as shown in the following image:



4. Click **Reset zoom** to return the graphs to the original state.

Viewing the resource utilization heat map

Use the heat map to quickly identify resource contention issues within a PowerCenter domain, and to analyze bottlenecks caused by too many workflow jobs running within the same time period.

The heat map provides a calendar view showing periods of heaviest and lightest resource consumption. You can view data based on memory consumption or on CPU usage. The heat map displays CPU usage data by default.

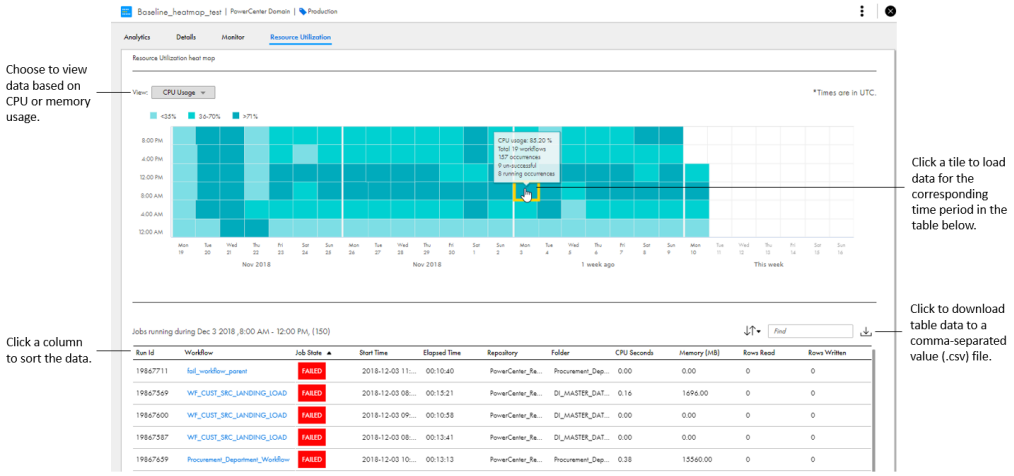
Each tile in the calendar represents a four-hour time period. The darkest colored tiles represent periods of 71% or higher consumption; the lighter colors represent periods with lower consumption. Click a tile to load details for the jobs that ran within the time period.

You can download the table containing the data for a selected tile to a comma-separated value (.CSV) file. If you note that consumption is high for a time period on one day, but lighter for the same time period on a different day, you might want to download the tables for both tiles and compare the data in each to identify possible causes.

Note: To enable the heat map to display data, you must apply an EBF to each PowerCenter node in the domain. See "Enabling the resource usage heat map" on page 56 for details.

1. Click the **Overview** tab.
2. Click a PowerCenter domain.

   You might need to first select a location, then select a domain within the location.
3. Click the **Resource Utilization** tab.



4. Choose to display CPU usage data or memory consumption data.
5. Click the tile for a time period.

   Details for the jobs that ran within the time period load in the table below.

Enabling the resource usage heat map

To enable the heat map to display data, you must apply an EBF to each PowerCenter node in the domain.

The table below specifies the EBF to apply for each PowerCenter release:

PowerCenter Release	EBF to Apply
10.1.0	EBF-12627
10.1.1 Hotfix 2	EBF-12626
10.2 Hotfix 1	EBF-12625

After you apply the EBF to a domain, use Informatica Administrator (the Administrator tool) to add custom properties to the PowerCenter Integration Service. The table below describes the custom properties to add:

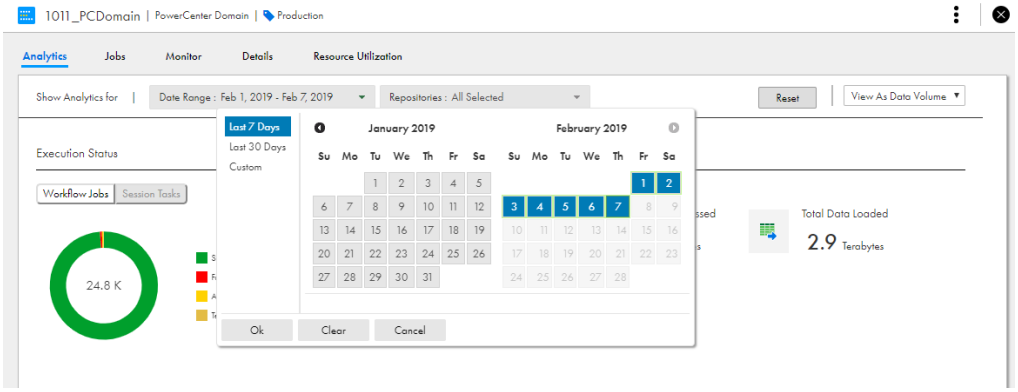
Property	Value	Description
CollectProcessMetricsInterval	2	Interval at which statistics are collected, in seconds. Minimum value is 2 seconds.
NumberOfResourceUtilStatsToPersist	20	Number of workflow instance statistics that are persisted for each workflow.
PersistStatsForRoundRobin	yes	Indicates whether workflow instance statistics are persisted in the round-robin partitioning mode.

Recycle the PowerCenter Integration Service after you add the custom properties.

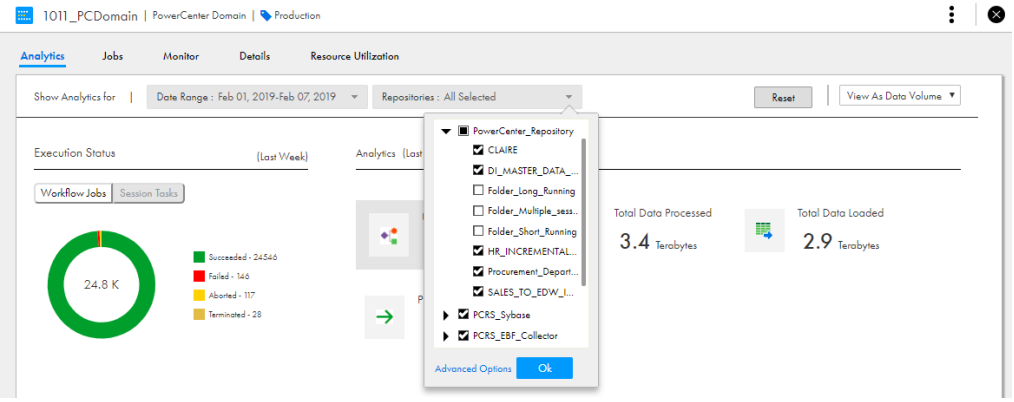
Using PowerCenter repository filters

Use PowerCenter repository filters to display run-time workflow analytics collected during specific time periods. You can also select specific PowerCenter repositories and folders within each repository to display data for. You can use advanced filtering to further refine results.

1. Select the date range you want to show statistics for from the Date Range menu.



2. Select the repositories you want to display data for from the Repositories menu. All repositories are selected by default; clear those you don't want to include.



3. Optionally select the folders within each repository you want to view data for. All folders are selected by default; clear those you don't want to include.

   You can also click **Advanced Options** to select the folders within each repository you want to view data for.

4. Click **OK**.
5. Click **Apply** to reload the page with analytics data for the selected date range and repositories.

CHAPTER 8

Usage scenarios

The extensive analytics provided by Operational Insights provide you with numerous options for troubleshooting issues and improving performance. Use the scenarios provided to help you get started with the service.

Monitoring the enterprise

Operational Insights provides you with a single console for monitoring all of your Informatica assets, including on-premise domains and domains running in the cloud. The rich analytics help you calculate the usage of your Informatica investment, and make decisions about where and when to expand or reallocate Informatica assets.

Monitor the health of all Informatica assets across the enterprise.

View availability and performance statistics for all of your Big Data and PowerCenter domains, nodes, and services.

Configure email alert notifications.

Configure Operational Insights to send email alert notifications to administrator users and groups when an issue occurs within a domain or with a Secure Agent.

You can configure alerts for the following events:

- The domain or Secure Agent is unavailable.
- A collector, service, or node running within the domain is unavailable.
- CPU or memory consumption by a domain node, a domain service, or the Secure Agent host has crossed a configurable threshold.

Review job and workflow execution status.

Identify the best- and worst-performing jobs workflows based on execution time, execution count, and volume of data processed.

Assess how effectively Informatica services are being used across the organization.

View data volume moved across the enterprise and within specific domains over defined time periods.

Predict future demand for Informatica services.

Analyze trends in data volume and node resource usage for specific locations, domains, nodes, clusters, and repositories.

Monitoring tasks

Common monitoring tasks are listed below:

Task	Description	Details
View enterprise health	View the status of domains, nodes and services across the enterprise. Drill down on specific locations and domains and grids within locations.	"Viewing enterprise health" on page 30
Configure alerts	Configure the email alerts to send when an issue occurs within a domain or with a Secure Agent host.	"Configuring alerts" on page 34
View domain-level analytics	Review detailed view run-time workflow statistics and data processing trends for a specific domain. Use filters to specify the date range and PowerCenter repositories and folders you want to view workflow data for.	"Viewing Big Data domain analytics" on page 37 "Viewing PowerCenter domain analytics" on page 45
View job and workflow statistics	View detailed analytics on jobs and workflows processed within the domain to troubleshoot performance issues and identify best and worst performers. For example, if you notice a workflow that typically takes 10 minutes to run recently took 20 minutes, you can investigate the underlying issue.	"Viewing Big Data job analytics" on page 38 "Viewing PowerCenter workflow analytics" on page 46
View node-level resource consumption	Understand trends in CPU usage and memory consumption for assets within a domain. Use this data to address performance issues and predict when you might need to add more capacity.	"Viewing Big Data domain resource usage analytics" on page 41 "Viewing PowerCenter domain resource usage analytics" on page 52

Troubleshooting issues

Use the metadata and metrics collected by Operational Insights to proactively uncover performance issues before they have a significant impact. Once you realize an issue exists, use the data collected to trace the problem back to the source.

Uncover workflow performance issues using run-time analytics.

Review run-time statistics to identify:

- Failed jobs and workflows. Drill down on error codes for failed PowerCenter workflows.
- Jobs and workflows that ran successfully, but processed zero rows of data.
- Jobs and workflows that run too long or too frequently.
- PowerCenter workflows that never run, and should be deleted from the PowerCenter repository to free up space.

Identify fluctuations in data processing times or volume.

Compare historical and average data movement and job execution times to discover aberrations, which could indicate a problem.

Diagnose performance issues caused by node resource issues.

Determine if performance degradation is due to non-Informatica processes running on a node, or because a node is running out of CPU or memory capacity.

Use recommendations to resolve issues.

Use recommendations generated to alert you to errors occurring within a domain. Recommendations contain links to relevant Informatica Knowledge Base articles associated with the reported error code.

Troubleshooting tasks

Common tasks for troubleshooting issues and run-time failures are listed below:

Task	Description	Details
View workflow analytics	Review the status of job and workflow instances to uncover errors or anomalies in data moved or processing times. Use filters to load and compare past run-time statistics with the most recent data for fluctuations in execution times, such as workflows taking longer to process than usual, or in data volume processed.	"Viewing Big Data job analytics" on page 38 "Viewing PowerCenter workflow analytics" on page 46
Identify failed workflows	Review workflows that failed and need to be rescheduled for the last 24 hours or past seven days.	"Identifying failed workflows" on page 51
Identify workflows with increasing run times	Review workflows that are taking longer to run over the last 24 hours or past seven days.	"Identifying workflows with increasing run times" on page 52
View node CPU utilization and memory consumption metrics	Once you've identified a failed or underperforming workflow, check the node it runs on to uncover issues due to resource utilization, such as non-Informatica processes running on the node consuming too much memory.	"Viewing Big Data domain resource usage analytics" on page 41 "Viewing PowerCenter domain resource usage analytics" on page 52
Identify resource contention issues	Identify and analyze PowerCenter workflow bottlenecks caused by resource contention issues. Use the data to identify workflows that are candidates for rescheduling.	"Viewing the resource utilization heat map" on page 55
View error remediation recommendations	Review the recommendations generated each day for errors occurring within a domain. If an article related to the error is available in the Informatica Knowledge Base, click the **Read Article** button to read the article.	"Viewing recommendations" on page 32

Analyzing performance

Use the workflow and resource consumption statistics provided by Operational Insights to detect and proactively address performance issues.

Predict when you will need to add additional capacity to a Big Data domain or PowerCenter grid.

Review average CPU and memory consumption across all nodes in a domain to determine when additional physical or virtual node hosts will be needed. Determine the likely data volume increase based on current processing trends.

Better distribute workload across nodes.

Identify the nodes that are processing the bulk of the workflows within each domain. Determine which nodes have low data volume levels and can take on more workflows, or could be reallocated within the local grid or moved to another domain.

Use recommendations to uncover and resolve issues.

Review the recommendations generated each day to discover errors occurring within a domain. Recommendations contain links to relevant Informatica Knowledge Base articles associated with the reported error code.

Performance tasks

Common tasks for assessing node and workflow performance are listed below:

Task	Description	Details
View domain-level job and workflow statistics	Review processing statistics for the best and worst performing jobs and workflows within the domain. Identify underutilized and overburdened nodes. If you determine that a node's performance is being impacted by workload, determine what projects can be moved from the node, and identify candidates to move it to.	"Viewing Big Data job analytics" on page 38 "Viewing PowerCenter workflow analytics" on page 46
Assess resource consumption	Determine if performance slowdown is due to a node running low on memory or CPU capacity, or to non-Informatica processes running on the node machine. Assess if current capacity is sufficient for expected increase in workload based on trends.	"Viewing Big Data domain resource usage analytics" on page 41 "Viewing PowerCenter domain resource usage analytics" on page 52
Identify resource contention issues	Identify and analyze PowerCenter workflow bottlenecks caused by resource contention issues. Use the data to identify workflows that are candidates for rescheduling.	"Viewing the resource utilization heat map" on page 55
View error remediation recommendations	Review the recommendations generated each day for errors occurring within a domain. If an article related to the error is available in the Informatica Knowledge Base, click the **Read Article** button to access the article.	"Viewing recommendations" on page 32
Configure auto-scaling	If you determine that spikes in processing are exhausting computing resources, configure auto-scaling to dynamically increase capacity.	Chapter 9, "Auto-scaling PowerCenter grids in the cloud" on page 64

Calculating chargeback

If your organization provides Informatica as a shared service, tracking usage by the internal users that fund your team is critical. Use Operational Insights to calculate chargeback costs for departments and business units using measurable workflow statistics.

Calculate service usage by business units.

View statistics on total workflows and sessions processed from the PowerCenter repositories and folders owned by specific business units.

Chargeback tasks

Common tasks for determining chargeback based on Informatica service usage are listed below:

Task	Description	Details
View PowerCenter domain-level workflow statistics	Use the domain analytics filters to view usage metrics based on: - Date range - PowerCenter repositories and folders owned by specific business units	"Viewing PowerCenter workflow analytics" on page 46

Capacity planning

Use predictive analytics to identify capacity shortfalls within a domain before they occur.

Anticipate and avoid computing shortfalls based on CPU and memory consumption across a domain.

You can view consumption statistics across nodes, grids and services. Use this data to ensure your infrastructure has the resources needed to handle spikes in processing.

Predict future processing requirements based on historical data volume growth.

Analyze how extensively Informatica services are being used by different business units to determine future capacity needs.

Dynamically add elastic nodes to a PowerCenter grid running in the cloud.

Configure autoscaling to automatically increase processing power when resource usage thresholds are exceeded.

Capacity planning tasks

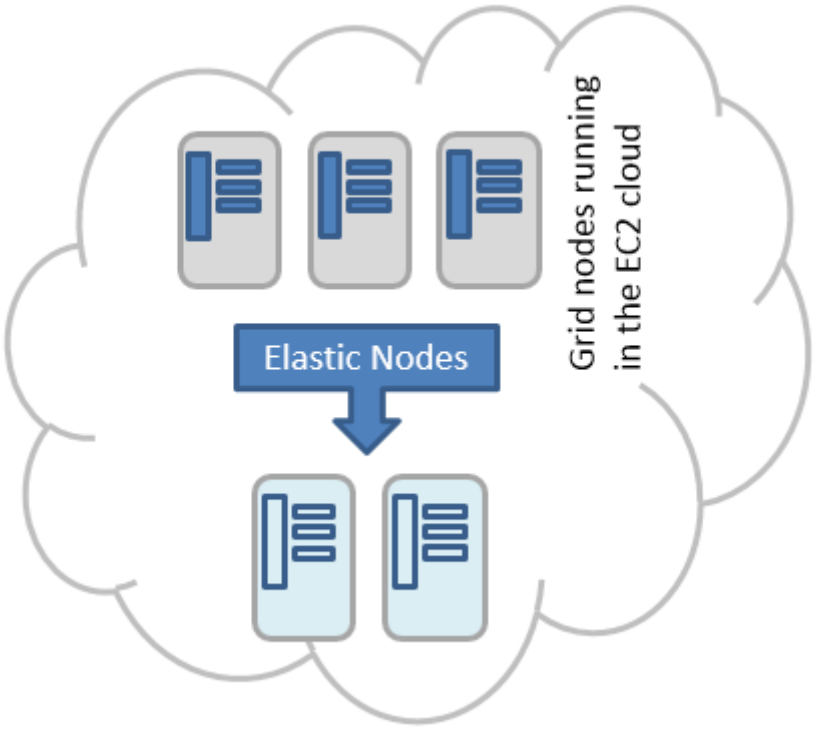
Common capacity planning tasks are listed below:

Task	Description	Details
View data processing trends across the enterprise	View data for the top domains in terms of data processed and workflow jobs to determine if your current Informatica deployments are sufficient to manage current and future workloads.	"Viewing data processing statistics" on page 32
View data processing trends within a domain	Review workflow and task processing statistics to determine whether a domain's current processing capacity is able to manage to manage current and future workloads.	"Viewing Big Data domain analytics" on page 37 "Viewing PowerCenter domain analytics" on page 45
View computing resource statistics	Understand trends in CPU and memory consumption statistics for assets within a domain. Use this data to address performance issues and predict when you might need to add more capacity to prevent resource starvation.	"Viewing Big Data domain resource usage analytics" on page 41 "Viewing PowerCenter domain resource usage analytics" on page 52
Configure auto-scaling	Define conditions for dynamically adding elastic nodes to a PowerCenter grid running in the cloud.	Chapter 9, "Auto-scaling PowerCenter grids in the cloud" on page 64

CHAPTER 9

Auto-scaling PowerCenter grids in the cloud

Use auto-scaling to dynamically add elastic worker nodes to a PowerCenter Integration Service grid running in an Informatica Intelligent Cloud Services domain on Amazon Elastic Compute Cloud (EC2).



You configure auto-scaling by setting scale up conditions based on CPU utilization or memory consumption for nodes within the grid, which is calculated as an average across for all of the processes running on all nodes in the grid every 5 minutes by default.

When a condition is exceeded, Operational Insights executes scripts in the cloud to perform the following steps:

1. Create the virtual machine (VM).
2. Copy the Informatica master key from the domain master gateway node to the cloud storage. The master key is needed to start each node in the cloud.

3. Download the Informatica master key and any custom node startup or shutdown scripts to the node from cloud storage. The truststore file and keystore file are also uploaded if the Informatica domain uses SSL authentication. Any custom scripts used in auto-scaling are also uploaded.
   All of these artifacts are deleted from storage after being downloaded to the node.
4. Provision the worker node on the VM.
5. Start the node and add it to the grid. Operational Insights waits 20 minutes for the node to start and join the grid; if this time is exceeded, the VM is terminated.

A new node is added each time the condition is exceeded. For example, if the CPU utilization threshold is set at 70%, a single node will be added when average CPU utilization for all processes across the domain exceeds this percentage. If the threshold is exceeded again, another node will be added. This process will continue as needed until the maximum number of VMs set in the auto-scaling configuration for a grid is met. Up to 12 VMs can be created.

You also set scale down conditions that specify when elastic nodes are no longer needed. When the condition is met, a node is removed from the grid, and the VM is terminated. Each elastic node is removed from the grid in the reverse order in which it was added.

You configure auto-scaling for each PowerCenter Integration Service grid within a domain. A wizard guides you through the configuration process.

Selecting the cloud vendor configuration

Select the cloud vendor to use when auto-scaling PowerCenter grids within the domain.

1. Select the domain you want to configure auto-scaling for.
2. Click the **Details** tab.
3. Click the **Configure** button under Auto-scaling Configuration section of the page.
4. Select the cloud vendor configuration in the Vendor Settings panel.

Adding or editing a cloud vendor configuration

You can create a new cloud vendor configuration to use when auto-scaling grids. Once a vendor configuration is added, you can select it in the Vendor Settings panel of the wizard.

You can also edit an existing cloud vendor configuration. Note however that editing a vendor configuration may impact other auto-scaling configurations.

1. Click the **Add New Vendor** button on the Vendor Settings panel.
2. Enter values for the following properties.

Property	Description
Vendor Name	Enter a name to assign to the cloud vendor configuration. The name is added to the list of selectable configurations. The name must be unique within Operational Insights.
Vendor Type	The cloud vendor.

Property	Description
IAM Profile	The Identity and Access Management (IAM) instance profile to use to auto-scale grids. The IAM instance profile enables elastic nodes to download the master key and any custom scripts from Amazon S3. The profile must contain an IAM role with privileges to access S3.
Access Key	The access key ID associated with your AWS account. Required if the machine hosting the Secure Agent used by the domain does not run in Amazon EC2. You do not need to provide this value if the domain uses a Secure Agent that runs in Amazon EC2, and you already have an IAM instance profile assigned to the domain's master gateway node.
Secret Key	The secret access key for your AWS root account. Required if the machine hosting the Secure Agent used by the domain does not run in Amazon EC2. You do not need to provide this value if the domain uses a Secure Agent that runs in Amazon EC2, and you already have an IAM instance profile assigned to the domain's master gateway node.

3. Click **Save** to complete the configuration.

Auto-scaling in Amazon Web Services

You can configure auto-scaling in Amazon Web Services to dynamically add processing capacity to PowerCenter Integration Service grids.

License requirements

You must subscribe to the Amazon Elastic Block Store (Amazon EBS) and Amazon Elastic Compute Cloud (Amazon EC2) services to use the auto-scaling feature. If you don't already have an AWS account, create one at http://aws.amazon.com.

Ensure that your Informatica license covers the number of additional cores you intend to add as virtual machines to a grid. Contact Informatica if you need to add additional cores to your license.

Amazon Machine Images for PowerCenter

Auto-scaling in Amazon Web Services requires an Amazon Machine Image (AMI), which provides everything required to launch a virtual machine instance in Amazon EC2 and provision it with a PowerCenter worker node. You specify the AMI to use to auto-scale a grid when you configure auto-scaling in Operational Insights.

The following PowerCenter AMIs are available as "Bring Your Own License" (BYOL) or "Pay As You Go" (PAYG) options on the AWS Marketplace:

- Informatica PowerCenter For Red Hat Linux (BYOL)
- Informatica PowerCenter For Windows (BYOL)
- Informatica PowerCenter For RHEL (PAYG)
- Informatica PowerCenter For Windows (PAYG)

You must have an Informatica license to use the BYOL AMI options.

You can also create custom AMIs to use to enable auto-scaling in your PowerCenter grids. If you plan to create and use your own AMI, ensure that the AWS account you use to configure auto-scaling has access to the AMI.

Preparing to enable auto-scaling in Amazon Web Services

Review the following information to prepare to enable auto-scaling in AWS.

Create an AWS Identity and Access Management (IAM) instance profile.

You must create an AWS Identity and Access Management (IAM) instance profile to enable elastic nodes to download the master key and any custom scripts from Amazon Simple Storage Service (Amazon S3).

An instance profile is a container for an IAM role that you can use to pass role information to an Amazon EC2 instance when the instance starts. When you use the AWS Management Console to create an IAM role for Amazon EC2, the application creates an IAM instance profile and gives it the same name as the role.

The IAM instance profile must contain an IAM role with privileges to access the Amazon S3 bucket containing the master key and custom scripts. You might need to create a policy to grant read-write access to the Amazon S3 bucket, and then attach the policy to the IAM role.

Create access keys for your AWS account.

You must supply the access key ID and the secret access key for your AWS root account if the machine hosting the Secure Agent used by the Informatica domain does not run in Amazon EC2. The access key data is used to access Amazon S3 when auto-scaling elastic nodes in an Informatica grid.

If the domain uses a Secure Agent on a host that runs in Amazon EC2, and you already have an IAM instance profile assigned to the domain's master gateway node, you do not need to supply an access key ID and secret access key.

Create custom elastic node startup and shutdown scripts.

You can create custom scripts to execute before elastic nodes are added to the grid. For example, you might create custom scripts to copy files to the VMs, download third party adapters, or install custom software on the VMs.

You can also create custom scripts to execute before elastic nodes are removed from the grid.

Copy the custom scripts to the machine hosting the Secure Agent used by the domain. You set the path to each script directory when you configure auto-scaling for a grid.

Include the Amazon IPaddress ranges in your list of approved IP addresses.

If your organization uses a firewall, and the Informatica domain and corresponding repository databases are not available in either the us-west-2 or us-east-1 S3 region, you must include the following in the list of approved IP addresses:

- The IP address range for the Amazon S3 region containing the Amazon S3 bucket. Operational insights copies files to the Amazon S3 bucket during auto-scaling.
- The IP address and port for the domain master gateway node host running in EC2. Include the JavaServer Faces (JSF) port if available.
- The IP address and database port for the PowerCenter repository database.
- The IP address and database port for the Model repository database managed by the monitoring Model Repository Service configured for the domain.

You can find the IP ranges for all Amazon S3 regions at the following link:

https://ip-ranges.amazonaws.com/ip-ranges.json

The IP address range for each region is updated several times a week. You might need to update the latest IP ranges for the region containing the Amazon S3 bucket frequently to ensure you have latest address range.

Configuring auto-scaling for a grid

Configure auto-scaling for each PowerCenter grid within the Informatica domain. A configuration panel is displayed for each grid.

1. Select the domain you want to configure auto-scaling for.
2. Click the **Details** tab.
3. Click **Configure Now** under the Auto-scaling Configuration section of the page.
4. Define the scale-up condition that adds elastic nodes to the grid.

   A new VM and node are provisioned each time the condition is exceeded. Valid values for the Percent property are from 10 to 99.
5. Define the scale-down condition that removes elastic nodes from the grid.

   After a node is shut down, the VM is terminated. Each VM is removed in the reverse order in which it was added.
6. Select the cloud vendor to use to auto-scale the grid.
7. Enter values for the following cloud configuration properties:

Property	Description
Amazon Machine Image	Select the Amazon Machine Image (AMI) type, and then enter the AMI ID to use to auto-scale the grid. You can find the AMI ID value for custom AMIs in the Amazon EC2 Management Console. In the console, select **Instances > Images**, and then copy the value under the AMI ID column. The Amazon EC2 Management Console does not display the AMI ID values for AMIs provided by Informatica.
Amazon EC2 Instance Type	The model identifier for the Amazon EC2 instance type. For the list of supported values, see the Pricing Details section for one of the following PowerCenter BYOL options available on the AWS Marketplace: - Informatica PowerCenter For Red Hat Linux (BYOL) - Informatica PowerCenter For Windows (BYOL)
Amazon S3 Region	The Amazon S3 geographic region that contains the Amazon S3 bucket that files are copied to during auto-scaling.
AWS Subnet ID	The range of IP addresses in your Amazon Virtual Private Cloud (VPC) that can be used to isolate different Amazon EC2 resources from each other or from the Internet. Each subnet resides in one Availability Zone. If a gateway node is already running in Amazon EC2, you can use the subnet ID specified for the VM the node runs on. You can find this value in the in the Amazon EC2 Management Console: - Select **Instances > Images**. - Select the VM the node runs on. - Click the **Description** tab. - Copy the value in the Subnet ID row.

Property	Description
Maximum VMs in Grid	The maximum number of VMs that can be provisioned when auto-scaling the grid. Up to 12 VMs can be provisioned.
Startup Script Location	Specify the path and file name of custom startup scripts you copied to the machine hosting the Secure Agent used by the domain. You can specify multiple scripts. Separate each script with a semi-colon(;).
Custom AWS Tags	Tags to use to categorize the auto-scale configuration. Enter tags as key=value pairs, each separated by a comma.
Private Key File	The fully qualified path and file name of the .ppk file for the Amazon EC2 key pair that you specified when you launched the VM instance. You only need to specify this property if the VM runs on a Linux host.
AWS VM Region	The Amazon EC2 geographic region where VMs are provisioned.
Amazon S3 Bucket	The Amazon S3 location to copy temporary files to during auto-scaling.
AWS Security Group	The identifier for the Amazon Web Services security group new VMs are added to.
Amazon EC2 Key Pair Name	The Amazon EC2 key pair name for the VM. You can reuse the key pair name assigned to another VM. In the Amazon EC2 Management Console, select **Network & Security > Key Pairs**, and then copy a value under the Key Pair Name column.
Shutdown Script Location	Specify the path and file name of custom shutdown scripts you copied to the machine hosting the Secure Agent used by the domain. You can specify multiple scripts. Separate each script with a semi-colon(;).
Machine Username	The user name that has permissions to execute the yum command on the Linux machine. You only need to specify this property if the VM runs on a Linux host.

8. Enter values for the following grid properties.

Property	Description
Installation Directory	The directory on VM instances to extract the Informatica node binaries to when using a custom AMI to auto-scale an elastic node. You only need to specify this property if you use a custom AMI to auto-scale nodes in the grid.
Code Page ID	The code page of used by the PowerCenter Integration Service process. An elastic node uses the code page when it extracts, transforms, or loads data. You only need to specify this property if you use a custom AMI to auto-scale nodes in the grid.
Integration Service Name	The name of the PowerCenter Integration Service associated with the grid.

Property	Description
New VM PM Root Directory	The PowerCenter Integration Service root directory on elastic nodes.
Node Port	The HTTP port used by elastic nodes added to the grid. All elastic nodes can use the same port.

9. Click **Save** to save the configuration.

   The **Test Auto-scaling Configuration** region appears.

10. Click **Test Scale Up** to ensure the configuration can successfully add an elastic node to the grid.

    The test creates a new VM and provisions an elastic worker node on it. The test may take up to 20 minutes to complete.
    Review the Secure Agent log files to see whether the test succeeded. The log files are written to the following directory:

    ```
    <Secure Agent installation directory>/apps/OpsInsightsAutoScale/logs
    ```

11. Click **Test Scale Down** to ensure a VM and elastic node can be successfully removed from the grid.

    The test removes the most recently added VM and node from the grid. The node is removed even if it is a production node. You can use this option for forcibly remove a VM and node from the grid if needed.
    Review the Secure Agent log files to see whether the test succeeded.

INDEX