Cloudera Observability

Configuring Cloudera Observability

Date published: 2023-04-31
Date modified: 2023-09-14

CLOUDERA

Legal Notice

Contents

Configuration tasks for CDP Public Cloud

Describes the tasks required to successfully enable Cloudera Observability for a CDP Public Cloud deployment.

Configuring Cloudera Observability for a CDP Public Cloud deployment requires the following tasks:

Enabling telemetry for a Data Hub deployment

Steps for enabling the collection of your workload analytic diagnostic data by Cloudera Telemetry Publisher, which is a service that collects and sends your workload diagnostic information about your job and query processes to Cloudera Observability.

About this task

The following procedure describes how to enable access to your workload diagnostic data by Telemetry Publisher for CDP clusters using a Data Hub service. By default, the collection of diagnostic data for a Data Hub service is disabled.

Cloudera Observability collects diagnostic information about your environment and workloads using either Telemetry Publisher or Databus WXM Client:

- For CDP clusters using a Data Hub service. Telemetry Publisher collects this information and sends it to Cloudera Observability.

  By default, the collection of diagnostic data is disabled. You enable telemetry when you register and configure your Cloudera environment in the Cloudera Management Console by turning on the Enable Workload Analytics option. For more information on how to register your Cloudera environment and the configuration settings, see the *Register an AWS environment from CDP UI* topic in the *Cloudera Docs Management Console Public Cloud* documentation, by clicking the Related Information link below.
- For CDP clusters using Cloudera Data Warehouse (CDW) and/or Cloudera Data Engineering (CDE) services, Databus WXM Client collects this information and sends it to Cloudera Observability:

  - By default, the collection of diagnostic data is enabled for CDW.
  - By default, the collection of diagnostic data is disabled for CDE. You enable telemetry when you enable your CDE service by selecting the Enable Workload Analytics check box. For more information on *Enabling a Cloudera Data Engineering service*, click the Related Information link below.

Before you begin

You must be logged in to the Cloudera Management Console and you must have completed all the prerequisite tasks for registering an environment that are described in the *Cloudera Docs Management Console Public Cloud* documentation. For more information, click the Related Information link below.

Procedure

1. From the Management Console navigation panel, select Environments.
2. In the Environment page, click Register Environment.

3. In the Register Environment wizard, enter your required environment settings using the wizard's instructions on its side panel for the General Information, Data Access and Data Lake Scaling, Region, and Networking and Security pages.

4. In the Logs section on the Storage page, enter your log storage requirements.

5. In the Telemetry section, slide the Enable Workload Analytics toggle switch to the right and then confirm this setting by clicking Enable.

6. Click Register Environment.

Results

Diagnostic information about your jobs, queries, and workloads are enabled for collection by Telemetry Publisher.

Related Information

Cloudera Management Console

Register an AWS environment from CDP UI

Enabling environment telemetry

Enabling a Cloudera Data Engineering service

Cloudera Observability deployment architecture for CDP Public Cloud

Describes the components and architecture of a basic Cloudera Observability environment deployed in CDP Public Cloud.

A Cloudera Observability environment for CDP Public Cloud comprises the following:

- Cloudera Environment, which is a secure and governed cloud service platform. The Cloudera Observability component services run in the Control Plane of the Cloudera Observability framework. Users access the Cloudera Observability web user interface from the web host server in this framework.
- Working Environment, which contains your Workload Clusters in your Workload environments, such as Production, Development, and Staging.
- Workload Cluster, which is one or more CDP clusters managed by Cloudera. Depending on your environment's Cloudera data service, Telemetry Publisher is associated with a cluster, virtual warehouse, or virtual cluster in a data service by either Cloudera Manager for a Data Hub service or by Databus WXM Client for CDE and CDW services.

The below *Cloudera Observability Architecture for Public Cloud* diagram shows the communication between Cloudera Observability and your workload clusters through either Telemetry Publisher or Databus WXM Client. Where, the Cloudera Observability service, including its main component services, run in the Cloudera Control Plane and the area on the right is your Working Environment that contains the clusters and services required to run your workload jobs and queries.

Cloudera Management Console (not shown) manages the clusters and services in each of your working environments.

Telemetry Publisher and Databus WXM Client collect and send diagnostic information about jobs and queries from your Workload Clusters to Cloudera Observability. To ensure that all data transfers are secure between your Workload Clusters and Cloudera Observability, Telemetry Publisher and Databus WXM Client communicate with Cloudera Observability's endpoints using Transport Layer Security (TLS), as follows:

1. When a job or query is completed, the Telemetry Publisher or Databus WXM Client connects to the Cloudera Observability service and asks to upload the workload diagnostic information. Once the request is authorized and verified, Cloudera Observability replies with a signed S3 URL that can then be used to upload the workload diagnostic information by Telemetry Publisher or Databus WXM Client.

2. When the URL is received, Telemetry Publisher and Databus WXM Client perform a secure and direct protocol test using the Cloudera Observability S3 URL, before sending any diagnostic data.

Figure 1: Cloudera Observability Architecture for Public Cloud
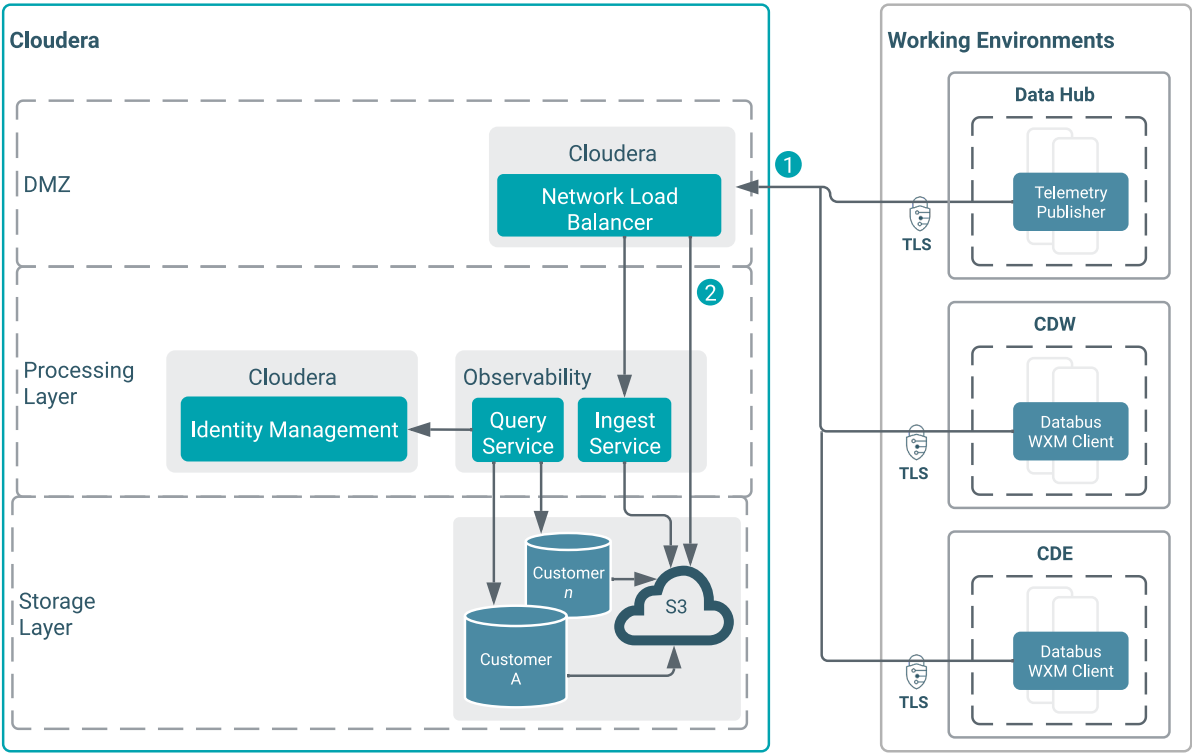
The following diagram shows the services from which Telemetry Publisher collects diagnostic metrics in a Data Hub environment:

Figure 2: Cloudera Data Hub Environment
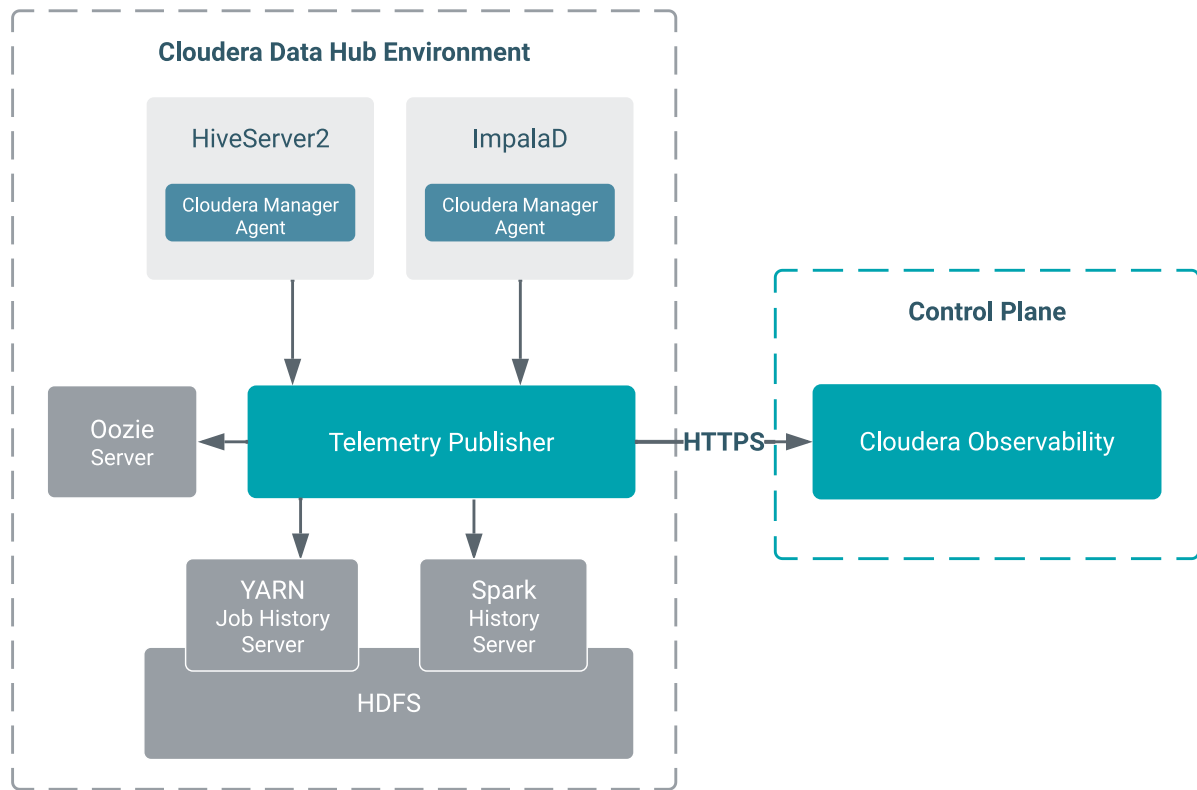
The following diagrams show the services from which Databus WXM Client (formally named Databus Producer) collects diagnostic metrics in a Cloudera Data Warehouse (CDW) and a Cloudera Data Engineering (CDE) working environment. Where, the Databus WXM Client continually communicates and checks for recently completed jobs and queries to see if there is any diagnostic data to transfer, such as task and event logs and job and query history files,

with the Hive DDL History, LLAP History, and Impala History pollers (HiveDDLHistoryPoller, LlapHistoryPoller, and ImpalaHistoryPoller):

- **Figure 3: CDW Environment**
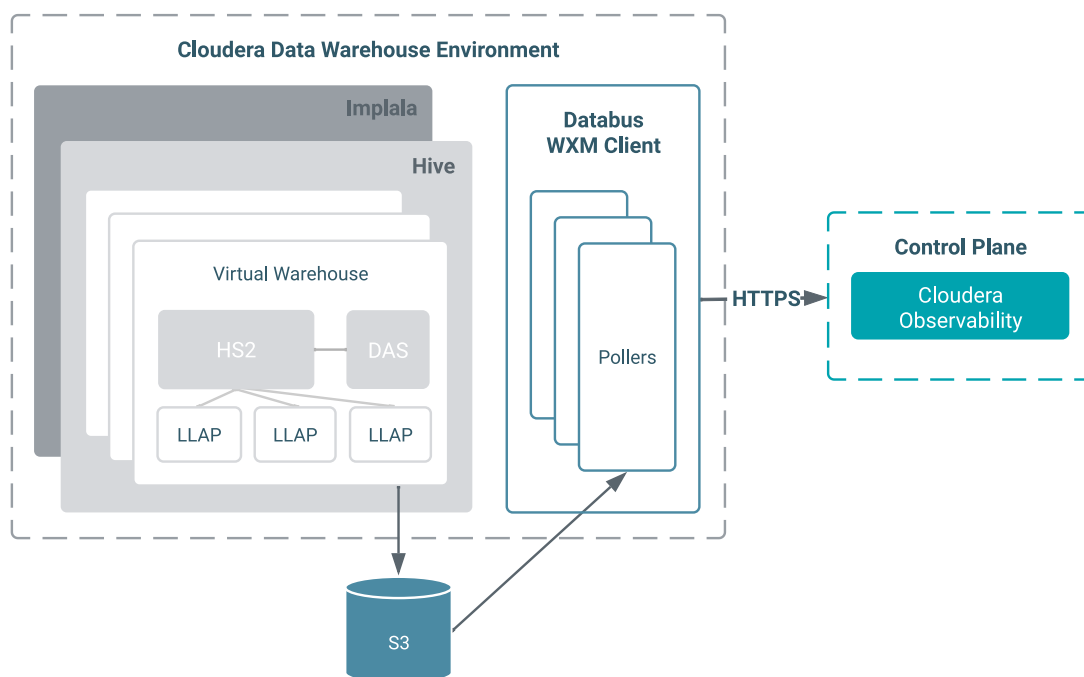
- **Figure 4: CDE Environment**
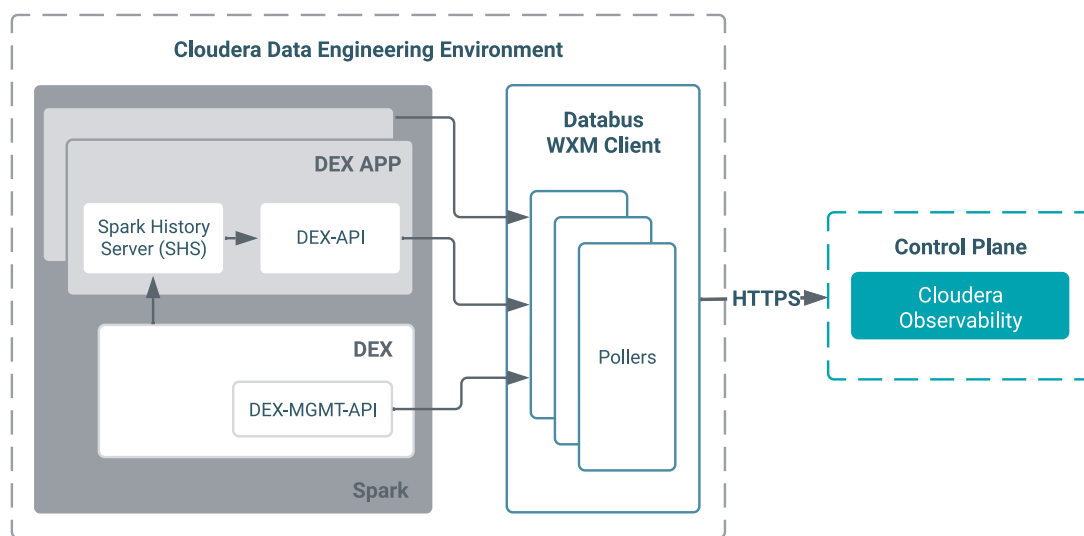
Configuration tasks for CDP Private Cloud

Describes the tasks required to successfully enable Cloudera Observability for a CDP Private Cloud deployment.

Configuring Cloudera Observability for a CDP Private Cloud deployment requires the following tasks:
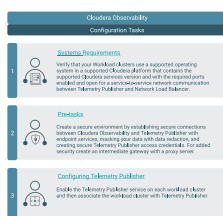
Deployment architecture for CDP Private Cloud

Describes the components and architecture of a basic Cloudera Observability environment deployed in CDP Private Cloud.

A Cloudera Observability environment for CDP Private Cloud comprises the following:

- Cloudera Environment, which is a secure and governed cloud service platform. The Cloudera Observability component services all run in the Control Plane of the Cloudera Observability framework. Users access the Cloudera Observability web user interface from the web host server in this framework.
- Working Environment, which contains your Workload Clusters in your Workload environments, such as Production, Development, and Staging.
- Workload Cluster, which is one or more CDP clusters managed by Cloudera Manager. Telemetry Publisher is associated with a cluster by Cloudera Manager.

The below diagram shows the communication between Cloudera Observability and your workload clusters through Telemetry Publisher. Where, the Cloudera Observability service, including its main component services, run in the Cloudera Control Plane and the area on the right, behind your firewall, is your Working Environment that contains the clusters and services required to run your workload jobs and queries.

Cloudera Manager (not shown) manages one or more clusters in each of your working environments. Telemetry Publisher is enabled and configured for Cloudera Observability from each Cloudera Manager instance in your working environment.

For example, as shown in the diagram below, Customer A's Production environment contains two clusters that are both managed by one instance of Cloudera Manager, whereas the Development environment, which also contains two clusters, is managed by two instances of Cloudera Manager, one for each cluster. In this case, only one Telemetry Publisher service is enabled for the Production environment and two Telemetry Publisher services are enabled for the Development environment.

Telemetry Publisher collects and sends diagnostic information about jobs and queries from your Workload Clusters to Cloudera Observability. To ensure that all data transfer is secure between your Workload Clusters and Cloudera Observability, Telemetry Publisher communicates with Cloudera Observability's endpoints using Transport Layer Security (TLS), as follows:

1. When a job or query is completed, Telemetry Publisher connects to the Cloudera Observability service and asks to upload the workload diagnostic information. Once the request is authorized and verified, Cloudera Observability replies with a signed S3 URL that can then be used to upload the workload diagnostic information by Telemetry Publisher.
2. When the URL is received, Telemetry Publisher performs a secure and direct protocol test using the Cloudera Observability S3 URL, before sending any diagnostic data.
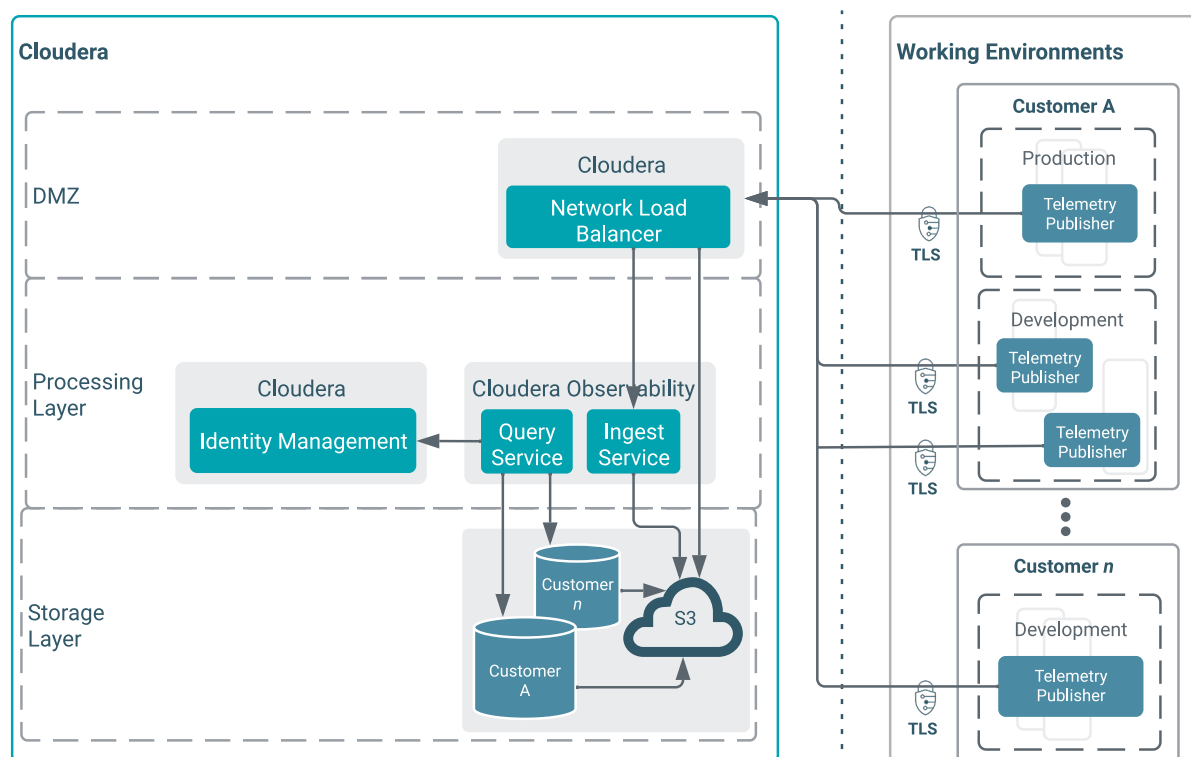
Security considerations

Describes the Cloudera security measures for Cloudera Observability.

Cloudera Observability security involves establishing secure connections, secure access authentication, and protecting customer and user data.

Redaction

Cloudera Manager's Telemetry Publisher collects and sends diagnostic information about job and query processes to Cloudera Observability. As this diagnostic data may contain sensitive information it is important to mask this data on your Workload cluster in Cloudera Manager before Telemetry Publisher sends it to Cloudera Observability. For more information on redaction, see the related links below.

Data in motion

Cloudera Observability and its services run on a secure Cloudera Observability framework. The transfer of diagnostic data to Cloudera Observability from your Workload cluster is completed using the Hypertext Transfer Protocol Secure (HTTPS) and the Transport Layer Security (TLS) protocols and an authentication process. Access to Cloudera Observability requires that each Workload cluster is configured with the Altus access and private keys in Cloudera Manager, which are used by Telemetry Publisher to authenticate the connection and identify the owner of the data. The access key identifies the user making the API call, and the request is signed by the key, using public key encryption. The signature is verified by the service and then the request is processed. You create your Telemetry Publisher access keys in Cloudera Observability, and add, delete, and disable them from Cloudera Manager on your Workload cluster. For more information on creating Telemetry Publisher access keys, see the related link below.

Note: The API connections are always encrypted using the Transport Layer Security (TLS) protocol.

Data at rest

Your diagnostic data goes through several transformation processes before it is stored in the Cloudera Observability S3 bucket. The Cloudera Observability microservices that are granted access to the Cloudera Observability S3 bucket do so through the Amazon Web Services (AWS) Identity Access Management (IAM) service, which securely controls access to AWS resources and is managed by you.

Furthermore, the AWS S3 Server Side encryption is used to encrypt your data at the object level. A dedicated AWS Key Management System (KMS) key is used for the encryption, and only the Cloudera Observability microservices are allowed to access the key.

Data isolation

Cloudera Observability isolates data access at the customer account level, where each customer account has its own dedicated storage database that contains all the data owned by the account, including the account's clusters. This ensures that each account's data is only cross-cluster viewable and not cross-account viewable.

All read access requests to Cloudera Observability use authentication, which identifies the customer account of the user before directing the user's queries to the database associated with the account.

User access

Cloudera Observability supports resource access roles and privilege types that define who is entitled to access your Cloudera Observability Workload clusters and who is entitled to access and administer your Cloudera Observability Workload clusters. The user's identity and Cloudera Observability access rights, such as, the existence of a user account, the correct password, and the correct user role and access credentials, are validated each time a user logs in to the Cloudera Observability UI. For more information, see the related link below.

Note: Cloudera Observability does not support Impersonation.

Related Information
Collecting Cloudera Observability diagnostic metrics
Redacting data
Generating Telemetry Publisher access credentials
Managing user access to workloads

System requirements for your Workload Cluster

Lists the minimum supported system requirements for your Workload Cluster.

Before you start the configuration tasks for Cloudera Observability, you must verify that your Workload clusters contain the minimum supported requirements for software and networks. For more information on the requirements for CDP Private Cloud Base and CDP Private Cloud Data Services, click the Related Information links below.

Related Information
Requirements
CDP Private Cloud Base Requirements and Supported Versions

Supported Cloudera versions

Lists the supported Cloudera platform and software versions for your Workload clusters.

The following table lists the supported Oracle Java Development Kit (JDK):

Table 1: Supported JDK

Product	Versions
Oracle Java Development Kit (JDK)	8 and 11

The following table lists the supported Cloudera platform and software for running your Workload clusters:

Table 2: Supported Cloudera Platforms and Software for your Workload clusters

Cluster	Version	Cloudera Manager
CDP 7.x clusters	Private Cloud Base 7.0.3 and higher	Cloudera Manager version 7.1.1 and higher

Unsupported Versions

The following versions are not supported:

- Cloudera Manager 6.0 and 7.0.3

The following figure shows an example of the Cloudera versions that are supported by Cloudera Observability:
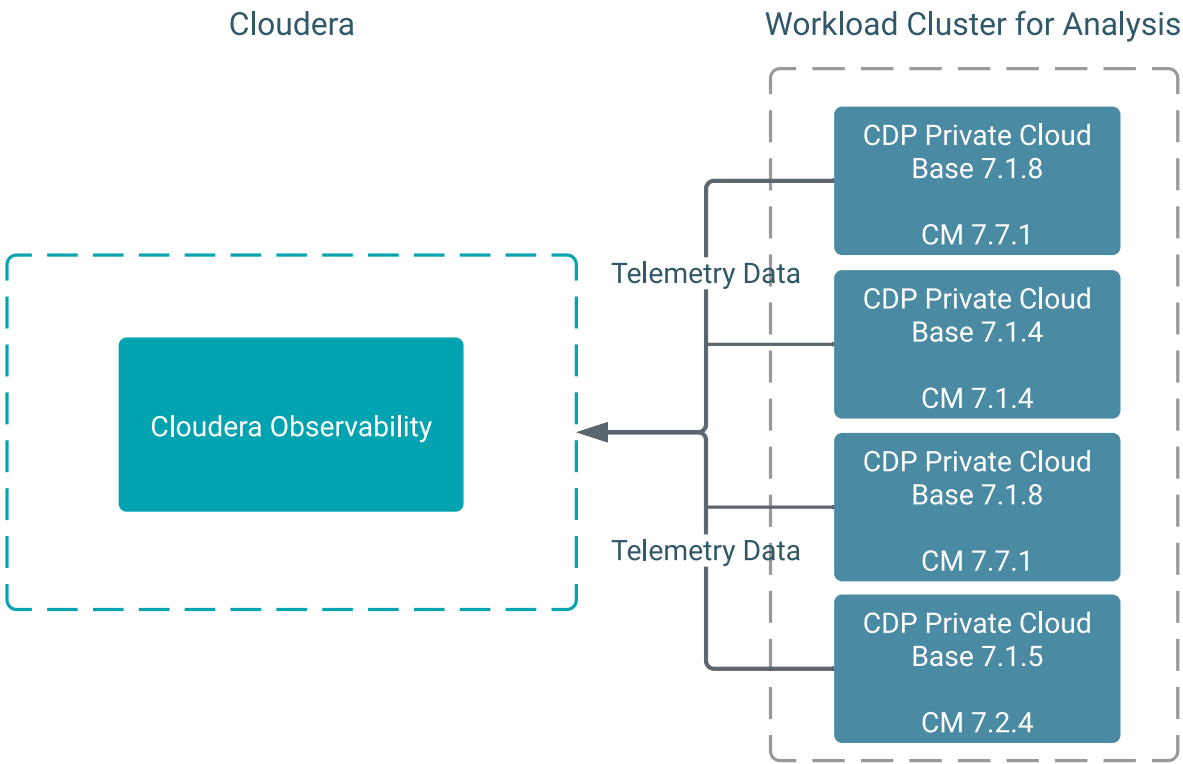
Related Information
Cloudera Product Compatibility

Network port requirements and connectivity verification

Lists the network port numbers and their respective protocols used by Cloudera Observability and dependent services.

Note: To enable communication, you may need to reconfigure or update your firewall.

Communication between Cloudera Observability and your Workload clusters is through Telemetry Publisher, which requires network communication with Network Load Balancer and the Cloudera Observability S3 bucket in the Cloudera Observability framework.

Telemetry Publisher collects and sends diagnostic information about job and query processes from your Workload clusters to Cloudera Observability. It communicates with Cloudera Observability, and its S3 bucket through Network Load Balancer using the Hypertext Transfer Protocol Secure (HTTPS) and the Transport Layer Security (TLS) protocols.

The following table lists the Network Load Balancer and the Cloudera Observability S3 bucket host names and the port numbers that must be enabled for service-to-service network communication between Telemetry Publisher and Cloudera Observability:

Table 3: Network Port Numbers

Port Number	Host Name
443	Host names for a US-based Control Plane cloud region: • dbusapi.us-west-1.sigma.altus.cloudera.com • cloudera-dbus-prod.s3.amazonaws.com
443	Host names for a EU-based Control Plane cloud region: • dbusapi.eu-1.cdp.cloudera.com • mow-prod-eu-central-1-sigmadbus-dbus.s3.eu-central-1.amazonaws.com
443	Host names for a AP-based Control Plane cloud region: • dbusapi.ap-1.cdp.cloudera.com • mow-prod-ap-southeast-2-sigmadbus-dbus.s3.ap-southeast-2.amazonaws.com

Network connectivity between Telemetry Publisher and Cloudera Observability verification

Cloudera recommends verifying access from Telemetry Publisher to Network Load Balancer and the Cloudera Observability S3 bucket on port 443, by running the following commands for your cloud region:

• For a US-based Control Plane cloud region run:

```
curl -v https://dbusapi.us-west-1.sigma.altus.cloudera.com:443
```

```
curl -v https://cloudera-dbus-prod.s3.amazonaws.com:443
```

• For a EU-based Control Plane cloud region run:

```
curl -v https://dbusapi.eu-1.cdp.cloudera.com:443
```

```
curl -v https://mow-prod-eu-central-1-sigmadbus-dbus.s3.eu-central-1.ama
zonaws.com:443
```

• For a AP-based Control Plane cloud region run:

```
curl -v https://dbusapi.ap-1.cdp.cloudera.com:443
```
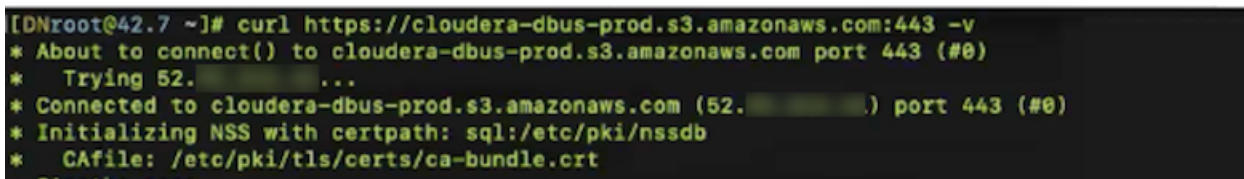
```
curl -v https://mow-prod-ap-southeast-2-sigmadbus-dbus.s3.ap-southeast-2
.amazonaws.com:443
```

Where, the -v option outputs the results in the terminal, which enables you to verify that the network port is open. For example, the following image shows that a successful connection was established after running the curl command for the US-based Control Plane cloud region with the cloudera-dbus host name.

Public key and CA certificate verification

Note: The public certificates offered with the Network Load Balancer and the Cloudera Observability S3 bucket web URL hosts can be verified with your Certified Authority (CA) truststore certificates. For maximum security, Cloudera recommends that you use your organization's private key infrastructure (PKI) and manage your own PKI and truststore certificates and certificate keys.

Cloudera recommends that you verify that the public keys and CA certificates are present in the Telemetry Publisher truststore by running the following command for your cloud region:

- For a US-based Control Plane cloud region run:

```
openssl s_client -connect dbusapi.us-west-1.sigma.altus.cloudera.com:443 -
showcerts
```

- For a EU-based Control Plane cloud region run:

```
openssl s_client -connect dbusapi.eu-1.cdp.cloudera.com:443 -showcerts
```

- For a AP-based Control Plane cloud region run:

```
openssl s_client -connect dbusapi.ap-1.cdp.cloudera.com:443 -showcerts
```

You can import a certificate from another truststore file to your truststore file with the keytool utility, by running the following command:

```
keytool -importkeystore -srckeystore cacerts -srcstorepass changeit -destkey
store truststore -deststorepass truststore_password
```

Where,

- *truststore*, is the destination truststore file name.
- *truststore_password*, is the password that opens the destination truststore file.

Pre-tasks

Describes how to generate Telemetry Publisher access credentials, how to enable endpoint services between Telemetry Publisher and Cloudera Observability, and how to enable or disable data redaction, which provide secure data transfers and conceal your sensitive data. For added security, optional configuration steps are also included for enabling an intermediary gateway using a proxy server.

Before configuring Telemetry Publisher and enabling its service on your cluster you must complete the Telemetry Publisher pre-requisite tasks.

Firewall configuration for Cloudera Observability

Connecting Telemetry Publisher to Cloudera Observability through endpoint services creates a secure connection between your CDP Data Hub cluster and the Cloudera Observability cloud service.

The Cloudera Telemetry Publisher service collects metrics from various components in a CDP Data Hub cluster and securely sends these metrics by way of the Hypertext Transfer Protocol Secure (HTTPS) protocol and the Transport Layer Security (TLS) encryption over the internet to Cloudera Observability.

Enabling secure communication from a CDP Data Hub cluster to the Cloudera Observability service that runs on an Amazon Web Services (AWS) cloud platform, requires that Telemetry Publisher connects to Cloudera Observability through the following endpoint services:

Depending on your cloud region, pair an Endpoint #1 (EC2 service) with an Endpoint #2 (S3 service) as follows:

- For a US-based Control Plane cloud region, pair Endpoint #1 (EC2 service):

  ```
  https://dbusapi.us-west-1.sigma.altus.cloudera.com
  ```

  with Endpoint #2 (S3 service):

  ```
  https://cloudera-dbus-prod.s3.amazonaws.com
  ```

- For a EU-based Control Plane cloud region, pair Endpoint #1 (EC2 service):

  ```
  https://dbusapi.eu-1.cdp.cloudera.com
  ```

  with Endpoint #2 (S3 service):

  ```
  https://mow-prod-eu-central-1-sigmadbus-dbus.s3.eu-central-1.amazonaws.com
  ```

- For a AP-based Control Plane cloud region pair Endpoint #1 (EC2 service):

  ```
  https://dbusapi.ap-1.cdp.cloudera.com
  ```

  with Endpoint #2 (S3 service):

  ```
  https://mow-prod-ap-southeast-2-sigmadbus-dbus.s3.ap-southeast-2.amazona
  ws.com
  ```

Where, these endpoints map to a dynamic IP address in AWS. For more information on the IP address ranges, see the Amazon documentation.

You can also configure a HTTP proxy between Telemetry Publisher and Cloudera Observability. In this configuration, the proxy acts as a HTTP tunnel for the encrypted TLS communication between Telemetry Publisher and Cloudera Observability.

Redacting data

Telemetry Publisher collects diagnostic data from logs, job configurations, and SQL queries, and then sends this data to Cloudera Observability. As this diagnostic information may contain sensitive information it is important to mask this data before Telemetry Publisher sends it to Cloudera Observability.

Data redaction works separately from Cloudera data encryption techniques. Data encryption alone does not preclude administrators with full access to the cluster from viewing sensitive user data. Redaction ensures that cluster administrators, data analysts, and others cannot see personally identifiable information (PII) or other sensitive data that is not within their job domain. At the same time, it does not prevent users with appropriate permissions from accessing data to which they have privileges.

In the event that redaction was disabled, such as during testing, Cloudera recommends that before you configure Telemetry Publisher you verify that redaction has not been disabled.

Redacting log and query data
By default, redaction for log and SQL query data is enabled for Telemetry Publisher.

Note: Only the sensitive data in the actual file is redacted. Metadata, such as the file's name, the file's owner, and information about the data in the file is not redacted.

Redacting Spark data

By default, redaction is enabled in the YARN service for Spark SQL data.

The YARN service redacts Apache Spark SQL sensitive data from event and executor logs.

Important: To ensure that Telemetry Publisher only sends redacted data to Cloudera Observability do not change the spark.redaction.regex configuration property.

Redacting MapReduce data

Telemetry Publisher reads the job configuration file from HDFS. You can enable data redaction for your MapReduce jobs pulled from HDFS by Telemetry Publisher by adding your MapReduce job configuration properties in the Cloudera Manager YARN configuration settings.

About this task

In YARN you add MapReduce job configurations that enable data redaction when MapReduce data is pulled from HDFS.

Procedure

1. In a supported web browser on your Workload cluster, log in to Cloudera Manager.
2. In Cloudera Manager, select Clusters, YARN, and then click the Configuration tab.
3. Search for the Redacted MapReduce Job Properties property.

   Note: By default, several MapReduce job configuration properties are set for you by the YARN service. Do not change these settings.

4. Add additional MapReduce job configurations by clicking the plus sign (+), which is located after the last configured property, and entering the default gateway group.
5. Click Save Changes.
6. Restart the YARN service.

Disabling redaction for testing

Steps for disabling the Log and Query redaction property in Telemetry Publisher for testing tasks.

About this task

Describes how to disable the Log and Query Redaction property, which by default is enabled for Telemetry Publisher.

Important: To protect sensitive data from being accessed by unauthorized users, Cloudera recommends that log and query redaction is enabled for both HDFS and the Telemetry Publisher service.

The Log and Query Redaction property works with the Log and Query Redaction property in HDFS. Both redaction properties must be disabled for Telemetry Publisher to start.

Note: The Log and Query Redaction configuration property is available in Cloudera Manager version 5.16 and higher.

Procedure

1. In a supported web browser on a Workload cluster, log in to Cloudera Manager with administrative privileges.
2. In Cloudera Manager, select Clusters, HDFS, and then click the Configuration tab.
3. In the Search field, enter redact, which locates the Log and Query redaction properties for HDFS.
4. Deselect the Enable Log and Query Redaction property.
5. Click Save Changes.

6. In the Cloudera Manager Home page, select Clusters, locate and select Cloudera Management Service, and then select the Configuration tab.

7. From the Filters panel in the SCOPE section, select Telemetry Publisher.

8. In the Search field, enter redact, which displays the Log and Query Redaction property.

9. Deselect the Log and Query Redaction property for the Telemetry Publisher Default Group.

10. Click Save Changes.

11. Restart both the HDFS and the Telemetry Publisher services, which disables the log and query redaction feature.

Adding a proxy server

Steps for configuring a proxy server, which adds extra security by enabling an intermediary gateway for sending your workload data to Cloudera Observability.

About this task

Describes how to add a proxy server as an intermediary gateway.

Note: You cannot upload data from Amazon Web Services (AWS) using a proxy server.

You can configure the Telemetry Publisher service to send data by way of a proxy server for database and metric data uploads. By default, this configuration property is disabled.

Telemetry Publisher uses the TLS and HTTPS protocols to send telemetry information to Cloudera Observability, which ensures that the data is encrypted. The proxy you use must support the HTTP CONNECT method to be able to pass through the encrypted messages. For more information, see the associated HTTP CONNECT Request for Comments (RFC) document.

Note: Telemetry Publisher support for proxy servers is only available in Cloudera Manager version 5.16.2 and higher.

Procedure

1. In a supported web browser on a Workload cluster, log in to Cloudera Manager with administrator privileges.

2. In Cloudera Manager, select Clusters, locate and select Cloudera Management Service, and then select the Configuration tab.

3. From the Filters panel in the SCOPE section, select Telemetry Publisher.

4. In the Search field, enter proxy, which displays the proxy configuration properties.

5. In the Proxy Support for Telemetry Publisher property, select the Telemetry Publisher Default Group check box and do the following:

   a. In the Proxy Server field, enter the proxy server name.
   b. In the Proxy Port field, enter the port number for the proxy server.
   c. In the Proxy User field, enter the proxy server user name, which is used for access authentication.
   d. In the Proxy Password field, enter the password for the proxy server user name.

   Note: If these properties do not appear, search for the Java Configuration Options for Telemetry Publisher property and in its entry field, enter the following:

   ```
   -Djdk.http.auth.tunneling.disabledSchemes=""
   ```

6. Click Save Changes, and then restart the Telemetry Publisher service.

Generating Telemetry Publisher access credentials

Steps for generating access credentials that enable communication between your Workload clusters and Cloudera Observability.

About this task

Describes how to create the access and private keys for Cloudera Manager's Telemetry Publisher, which collects and sends your workload information to Cloudera Observability. You will be required to supply these values when you enable the Telemetry Publisher service on your Workload Cluster.

Note: Only Cloudera Observability administrators (ObservabilityClusterAdmin formerly WXMClusterAdmin) can create Telemetry Publisher access credentials.

Procedure

1. In a supported browser, log in to the Cloudera Data Platform (CDP) as the user with administrative privileges.

   The CDP Public Cloud web interface landing page opens.

2. From the Your Enterprise Data Cloud landing page, select the Management Console tile.

   The Management Console home page opens.

3. From the Management Console's navigation panel, scroll down, click your user name, and then select Profile.

   Your Profile page opens.

4. Click Generate Access Key.

   The Generate Access Key dialog box opens.

5. In the Generate Access Key dialog box, click Generate an old format access key.

   The Telemetry Publisher Access Key ID and Private Key credentials are created and their respective fields populated with the credential text.

6. Do one of the following:

   - Manually save the access credentials:

     a. Record the Access Key ID credential text and store it somewhere safe. You will be required to supply this value when you enable the Telemetry Publisher service on your Workload Cluster.

     b. In a new text file, copy and paste the Private Key credential text exactly as provided without trailing spaces and then name and save the file somewhere safe. You will be required to supply this file when you enable the Telemetry Publisher service on your Workload Cluster.

   - Download the credentials file.

     a. Click Download Credentials File.

     b. Go to your Downloads directory.

     c. Open the file in a text editor and record the Access Key ID credential text and store it somewhere safe.

     d. Remove the text containing the Access Key ID, starting from [default] and ending with cdp_private_key=.

        The credentials file should now only contain the Private Key credentials text, starting from -----BEGIN P RIVATE KEY----- and ending with -----END PRIVATE KEY-----.

     e. Save the file somewhere safe. You will be required to supply this file when you enable the Telemetry Publisher service on your Workload Cluster.

Related Information

Enabling the Telemetry Publisher service

Cloudera Observability access roles

Configuring Telemetry Publisher

Tasks for enabling Cloudera Telemetry Publisher, which collects and transmits diagnostic information about job and query processes to Cloudera Observability.

Cloudera Telemetry Publisher is a role in the Cloudera Manager Management Service that collects and sends your workload information to Cloudera Observability. For example, when new clusters are added with Cloudera Manager, Telemetry Publisher automatically sends the new cluster information to Cloudera Observability.

Note: Cloudera recommends that you assign a dedicated disk for the Telemetry Publisher Service role on your Workload cluster. This prevents any issues when sending data to Cloudera Observability from affecting operations other than those performed by Telemetry Publisher.

Depending on the number and size of the jobs run on the cluster, the minimum supported disk drive size is 500GB. This size includes enough disk space for Telemetry Publisher to store data locally when it is unable to send data to Cloudera Observability due to connectivity or other issues.

Enabling the telemetry network communication for Cloudera Observability

Learn how to enable the network communication between Telemetry Publisher and Cloudera Observability.

About this task

Describes how to configure the host URL and access credentials for Telemetry Publisher.

Before you begin

Verify that you have the following values before enabling the Telemetry Publisher service, as you will be required to supply their values during this task.

- The Telemetry Publisher access credentials, which are required to register the Telemetry Publisher account.
- The Databus endpoint EC2 service URL for your cloud region:
    - For a US-based Control Plane cloud region use:

      ```
      https://dbusapi.us-west-1.sigma.altus.cloudera.com
      ```

    - For a EU-based Control Plane cloud region use:

      ```
      https://dbusapi.eu-1.cdp.cloudera.com
      ```

    - For a AP-based Control Plane cloud region use:

      ```
      https://dbusapi.ap-1.cdp.cloudera.com
      ```

Procedure

1. In a supported web browser on a Workload cluster, log in to Cloudera Manager.
2. In Cloudera Manager, select Clusters, locate and select Cloudera Management Service, and then select the Configuration tab.
3. Search for the Telemetry Publisher Advanced Configuration Snippet (Safety Valve) for telemetrypublisher.conf property and in its text field, enter one of the following according to your cloud region using the Databus endpoint EC2 service URL that you recorded as a prerequisite for these steps:
    - For a US-based Control Plane cloud region enter:

      ```
      telemetry.upload.job.logs=true
      telemetry.altus.url=https://dbusapi.us-west-1.sigma.altus.cloudera.com
      ```

    - For a EU-based Control Plane cloud region enter:

      ```
      telemetry.upload.job.logs=true
      telemetry.altus.url=https://dbusapi.eu-1.cdp.cloudera.com
      ```

    - For a AP-based Control Plane cloud region enter:

      ```
      telemetry.upload.job.logs=true
      telemetry.altus.url=https://dbusapi.ap-1.cdp.cloudera.com
      ```

4. Click Save Changes.

5. Back in the Cloudera Manager Home page, from the navigation panel, select Administration and then External Accounts.

> **Tip:** You can refresh Cloudera Manager and come back to the Cloudera Manager Home page by clicking on the Cloudera Manager icon at the top of the navigation panel.

6. In the External Accounts page, click the Altus Credentials tab, which displays your resource access account certificates.

7. Add a new Telemetry Publisher access account certificate by clicking Add Access Key Authentication and then in the **Add Access Key Authentication** dialog box do the following:

    a) In the Name field, enter an identifiable name for the Telemetry Publisher access key account.
    b) In the Access Key ID field, enter the Telemetry Publisher access key text exactly as provided without trailing spaces.
    c) From the Private Key list, select Choose File and then browse and select your Telemetry Publisher private key file.

    > **Note:** The Cloudera Observability Telemetry Publisher credentials are not related to Altus, but act as a pay-wall mechanism to use Cloudera Observability.

    d) Click Add, which saves the credentials as an Altus account certificate using the account name you provided and adds it on the Altus Credentials External Accounts page.

8. Back in the Cloudera Manager Home page, from the navigation panel, select Administration and then Settings.

9. Under the Filter CATEGORY section, select Altus, which populates the Settings page with your Telemetry Publisher access key accounts.

10. In the filtered result, select the Telemetry Publisher Altus account credential that you require for this Workload Cluster. In this case, the name you provided in step 7a.

11. Click Save Changes.

(Optional) Renaming the Workload cluster

Describes how to rename the Workload cluster with a human-readable name in Cloudera Manager.

About this task

Cloudera Observability identifies the cluster from a random string of 32 characters, such as 44a6e75e-8630-4773-9 ea9-6272478e84c2, which is difficult to identify and manage. Cloudera recommends completing the following task to rename your Workload cluster before you add and start the Telemetry Publisher role instance.

Procedure

1. In a supported web browser on a Workload cluster, log in to Cloudera Manager.
2. In Cloudera Manager, select Clusters, and then select the Workload cluster that requires a human-readable name.
3. From the Actions menu, select Rename Cluster.
4. In the Name field of the Rename Cluster dialog box, enter a new name that is easily identifiable by you.
5. Click Rename Cluster.

What to do next
Add and start a role instance of the Telemetry Publisher service on the Cloudera Manager Server node.

Adding and starting an instance of Telemetry Publisher

Describes how to associate a Workload cluster with Telemetry Publisher by designating a host cluster with the Telemetry Publisher service role and starting the Telemetry Publisher service for Cloudera Observability.

About this task

After configuring and adding the Telemetry Publisher credentials, the Telemetry Publisher service must be associated with your Working cluster by adding the Telemetry Publisher service role to a designated host and starting the Telemetry Publisher role instance.

Before you begin

Verify that you have configured and added the Telemetry Publisher credentials in Cloudera Manager.

Procedure

1. In Cloudera Manager, select Clusters and then locate and select Cloudera Management Service.
2. Add the Telemetry Publisher role on the Cloudera Manager Server nodes by doing the following:
   a) From the Actions menu, select Add Role Instances.

   The Add Role Instances for Cloudera Management Service wizard opens.
   b) Click inside the Telemetry Publisher field.

   The **Hosts Selected** dialog box opens.
   c) Select the check box of the host for the Telemetry Publisher and click OK.

   The wizard populates the Telemetry Publisher field with the selected host name.
   d) Click Continue.
   e) In the Review Changes page, review your selection and click Continue.
3. Start Telemetry Publisher by selecting the Instances tab, selecting the Telemetry Publisher check box, and then from the Actions for Selected menu, select Start.
4. In the Start message, confirm starting the Telemetry Publisher Service on your cluster by clicking Start.
5. Monitor the progress until the *Successfully started service* message appears and then click Close.

HDFS file access requirements

Describes how to access files from HDFS with Telemetry Publisher when your access keys are stored in the Cloudera Key Trustee Server.

By default, when keys are stored in the Key Trustee Server, the HDFS user for Telemetry Publisher (hdfs) does not have permission to access files.

To enable access to your files in HDFS, the Telemetry Publisher user must belong to the user groups that authenticate user access for the Job History Server and the Spark History Server. For example, if the hadoop user group authenticates access for the Job History Server and the spark user group authenticates access for the Spark History Server, then the Telemetry Publisher user must belong to the hadoop group and the spark group to download files from HDFS.

Enabling the Auto Actions feature in Telemetry Publisher

Steps for enabling the Cloudera Observability Auto Actions feature in the Telemetry Publisher service.

About this task

Describes how to access and enable the Telemetry Publisher Auto Actions property settings.

 Note: This task must be performed by a user who has either cluster or full administrator privileges.

Procedure

1. Verify that you have enabled the Telemetry Publisher service.
2. In a supported web browser, log in to Cloudera Manager as a user with full system administrator privileges.

3. From the Cloudera Manager navigation panel, click Clusters and then scroll down, locate, and select Cloudera Management Service.

4. In the Status Summary section on the Cloudera Management Service page, click Telemetry Publisher.

5. In the Telemetry Publisher page, select the Configuration tab.

6. In the Alert message that appears, click Continue Editing Role Instance.

7. In the Search field, enter Telemetry Publisher Advanced Configuration Snippet (Safety Valve) for telemetrypublisher.conf.

> **Tip:** Entering the full property name in the Search field is not mandatory. For example, in this case you can enter *safety* to locate the Telemetry Publisher Advanced Configuration Snippet (Safety Valve) for telemetrypublisher.conf configuration property.

8. In the Telemetry Publisher Default Group text box, enter the Telemetry Publisher Auto Actions configuration property settings that you require.

   For example, to enable Auto Actions for Spark applications, set the following:

   ```
   autoactions.yarn.app.collector.enabled=true
   autoactions.collection.spark.enabled=true
   ```

   For the full list of available Auto Actions property settings for Telemetry Publisher, see *Telemetry Publisher configuration settings for Auto Actions*.

9. Click Save Changes.

10. Restart the Telemetry Publisher service on your cluster by doing the following:

   a) Go back to the Cloudera Manager Home page.

   > **Tip:** Clicking the CLOUDERA Manager icon in the upper-left corner takes you back to the Cloudera Manager Home page.

   b) From the Cloudera Manager Navigation panel, click Clusters and then scroll down, locate, and select Cloudera Management Service.

   c) In the Status Summary section on the Cloudera Management Service page, click Telemetry Publisher.

   d) From the Actions menu, select Restart this Telemetry Publisher.

   e) In the Restart message, confirm restarting the Telemetry Publisher Service on your cluster by clicking Restart.

   f) Monitor the restart progress until the Successfully restarted role message appears and then click Close.

Related Information
Triggering action alerts across jobs and queries

Telemetry Publisher configuration settings for Auto Actions

Lists the Cloudera Observability Telemetry Publisher configuration property settings. Set in the Telemetry Publisher Safety Valve section in Cloudera Manager they enable Telemetry Publisher to collect data that is required by the Auto Actions feature.

Table 4: Telemetry Publisher configuration settings for Auto Actions

Property Name	Default Value	Description	Example
autoactions.yarn.app.collector.enabled	FALSE	• When set to true, the Telemetry Publisher collector is enabled and tests for workloads related to YARN. • When set to false, no YARN related collection configurations are considered.	autoactions.yarn.app.collector.enabled=true
autoactions.impala.collector.enabled	FALSE	• When set to true, the Telemetry Publisher collector is enabled and tests for workloads related to Impala. • When set to false, no Impala related collection configurations are considered.	autoactions.impala.collector.enabled=true

Property Name	Default Value	Description	Example
autoactions.collection.yarn.enabled	FALSE	• When set to true, the Telemetry Publisher collector evaluates all the YARN application Auto Actions. This is the starting point for an Auto Actions evaluation. • When set to false, no YARN application Auto Actions are evaluated on the cluster.	autoactions.collection.yarn.enabled=true
autoactions.collection.mr.enabled	FALSE	• When set to true, the Telemetry Publisher collector evaluates all the MapReduce Auto Actions. This is the starting point for an Auto Actions evaluation. • When set to false, no MapReduce Auto Actions are evaluated on the cluster.	autoactions.collection.mr.enabled=true
autoactions.collection.spark.enabled	FALSE	• When set to true, the Telemetry Publisher collector evaluates all the Spark application Auto Actions. This is the starting point for an Auto Actions evaluation. • When set to false, no Spark application Auto Actions are evaluated on the cluster.	autoactions.collection.spark.enabled=true
autoactions.collection.hive.enabled	FALSE	• When set to true, the Telemetry Publisher collector evaluates all the Hive query Auto Actions. This is the starting point for an Auto Actions evaluation. • When set to false, no Hive query Auto Actions are evaluated on the cluster.	autoactions.collection.hive.enabled=true
autoactions.definition.cache.refresh.minutes	5	This setting controls the number of minutes an Auto Action definition is stored in cache on the cluster. Your Auto Action definitions can be stored in cache on the cluster. As an Auto Action definition rarely changes, setting this value increases the delivery speed of requests by reducing the number of calls to Cloudera Observability.	autoactions.definition.cache.refresh.minutes=5

Logging in to Cloudera Observability

Learn how to access the Cloudera Observability web user interface to start viewing your diagnostic data for analysis.

About this task

Describes how to access Cloudera Observability and begin working with the Main and the Environment navigation panels.

Note: There can be a delay from job completion to when the job is available in Cloudera Observability, where large jobs can take up to 10 minutes to display.

Before you begin

Do the following:

• For CDP Private Cloud:

  • Verify that Telemetry Publisher is enabled for Cloudera Observability on your Workload clusters and that they are associated with Telemetry Publisher.
  • If applicable, verify that your environment's data services are using Cloudera Data Engineering (CDE) version 1.19 or above and/or Cloudera Data Warehouse (CDW) version 1.6.3 or above. Starting with these versions

the collected diagnostic data is categorized and displayed within their Data Service category in the Cloudera Observability web UI.

- For CDP Public Cloud:

  Verify that Enable Workload Analytics was turned on when you registered your Data Hub environment and when you installed your Data Engineering service.

  ⚠ **Important:** When you are not working in Cloudera Observability, Cloudera recommends that you explicitly log out by selecting your user name in the main navigation panel and clicking Log Out.

Procedure

1. In a supported web browser log in to the Cloudera Observability web UI by doing the following:
   a) In a supported browser, log into the Cloudera Data Platform.

      The CDP Public Cloud web interface landing page opens.
   b) From the Your Enterprise Data Cloud landing page, select the Observability tile.

   The Cloudera Observability web UI Landing page opens to the main navigation panel.

   📝 **Note:** For a list of supported web browsers, click the Related Information link below.

2. From the main navigation panel, select **Financial Governance**, which opens the **Chargeback** page. When configured, by you, this page displays the total costs and the hourly CPU and memory usage for all of your cost centers, including the unutilized resource usage costs from the **Uncategorised** section. For more information about the Financial Governance feature and how to configure your cost centers and assign them to your resources, click the Related Information link below.

3. From the main navigation panel, select Analytics.

   The Cloudera Observability **Environments** page opens.

4. Select an environment required for analysis.

   💡 **Tip:** To filter and display only those environment platforms or services of interest, from the Environments list, select the environment's Type.

   The **Environment** navigation panel opens, which hierarchically lists the environment and its services hosted on the selected environment.

5. Depending on the environment selected, verify that the **Cluster Summary** page is displayed for the environment's cluster required for analysis.

   To display the **Cluster Summary** page for a Data Lake, Database Catalog, Data Engineering, and Data Hub environment type, do one of the following:

   - From the Environment panel, expand the service's category and depending on the service, locate and select the Data Hub's cluster, the Data Engineering's Virtual Cluster, or the Data Warehouse's Virtual Warehouse that is required for analysis.
   - In the Data Services table, drill-down through the service links to locate and select the Data Hub's cluster, the Data Engineering's Virtual Cluster, or the Data Warehouse's Virtual Warehouse that is required for analysis.

   The **Cluster Summary** page, which is displayed as the title in your browser tab, displays performance trends and metrics about the processed jobs and queries and enables you to view historical trends for analysis when you select a predefined or custom time period from the Time-Range filter list.

6. From the cluster's ENGINES, the Data Engineering's Virtual Cluster, or the Data Warehouse's Virtual Warehouse, select a workload engine of interest.

   When an engine is selected, the name of the engine is displayed in the browser tab and the page's chart widgets display information about the workload jobs run by the selected engine, such as which jobs or queries have failed or are slow, their processing time, missed SLAs (thresholds), user and pool metrics, and outlier issues.

7. In the workload engine's page, review its chart widgets and then select a chart widget, such as Suboptimal. Select a link or bar and drill down further to view more information, such as health checks, execution details, baselines, and trends.

> **Tip:** Breadcrumbs are displayed at the top of each page, which displays the name of your current location and its preceding page levels. You can move between these levels by clicking on a breadcrumb location.

Related Information

Supported browsers

Analyzing your environment costs with Cloudera Observability

About the Cloudera Observability user interface hierarchy