Workload XM 2.1.3

Installing Workload XM

Date published: 2020-12-04
Date modified: 2021-07-22

CLOUDERA

https://docs.cloudera.com/

Legal Notice

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

On-Premises Installation Overview

A brief overview of the tasks required to successfully install Workload XM on a dedicated cluster within your environment.

Installing Workload XM requires the following tasks that are performed by you:

- Creating a CDP cluster for Workload XM that contains a minimum of 5 nodes and that is managed by Cloudera Manager.

  Note: Workload XM must be installed in a dedicated cluster, separate from your development, test, or production workload clusters. This configuration minimizes the impact on the cluster and prevents the need to upgrade your workload clusters to meet the needs of Workload XM.
- Verifying that your environment's system has the required supported software and hardware and the required network services and devices for installing Workload XM.
- Performing the pre-installation tasks.
- Downloading the Workload XM installation files from the Cloudera Downloads website to the host server on your Workload XM cluster.
- Activating the Workload XM parcel files, enabling secure communication and data encryption between components, Workload XM on-premises, and your Workload clusters, and setting the required Workload XM component configurations.
- Adding the Workload XM service in Cloudera Manager.
- Verifying the cluster network topology of your Workload XM environment.

Tip: The pre-installation, installation, and deployment tasks collect a series of parameter and property values. These values are used during the Workload XM Manager installation and deployment to configure and setup Workload XM specifically for your system. Cloudera recommends recording these values before starting a task.

Follow these guidelines to ensure a successful Workload XM installation :

- Decide on the type of Workload XM environment that best suites your business requirements.
- Read the system requirements. This ensures that your Workload XM on-premises cluster has the required base hardware and software.
- Read the installation pre-requisites and installation steps. This ensures that you understand the tasks required and how they are completed.
- Understand what software services are required and what account information is needed when configuring dependent services. Third-party software services, such as LDAP and network and firewall security, must be configured by you. For example, your SSL key pair file locations and private key are required during installation.
- Record all the required configuration values, such as host names, port numbers, user names and passwords.
- During the pre-installation tasks, record any new configuration values as you create them. You will be required to enter these configuration values later when you install and deploy Workload XM.
- After installing Workload XM, verify that the software stack installed successfully.

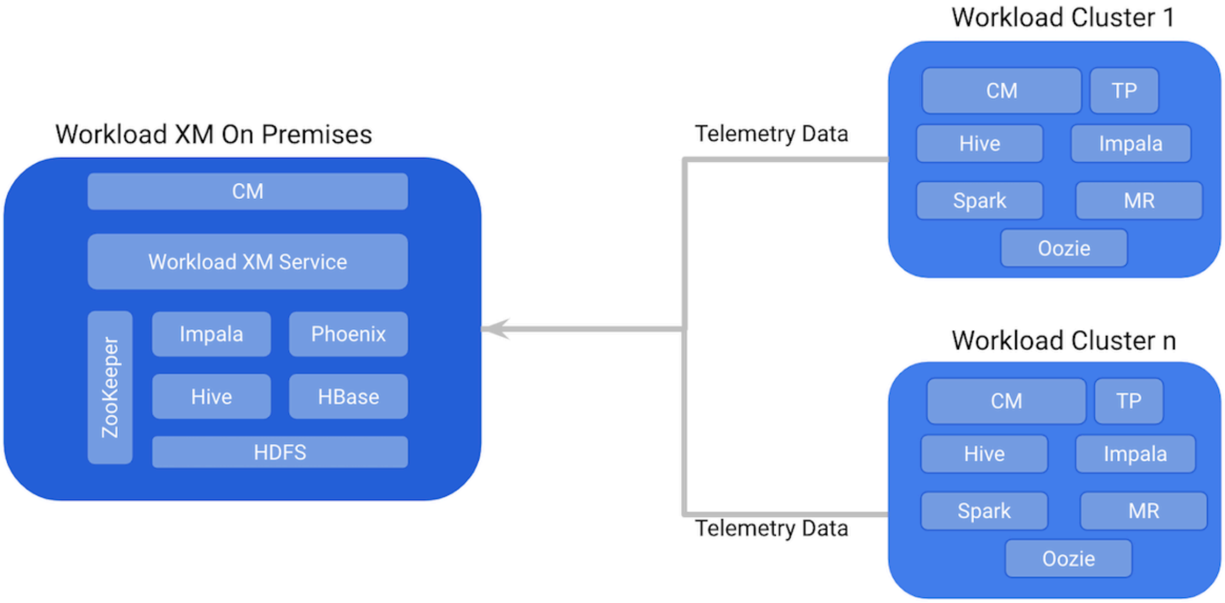
Architecture

Describes the components and architecture of a basic Workload XM environment.

Workload XM on-premises consists of two or more clusters:

- Workload XM on-premises cluster, which is a CDP cluster that is managed by Cloudera Manager. Workload XM and all its main component services are installed and run in this cluster. Users access the Workload XM web user interface from the web host server in this cluster.
- Workload Cluster, which is a CDH, CDP, or HDP cluster managed by Cloudera Manager. This cluster is associated with Telemetry Publisher in Cloudera Manager and runs your workload processes.

The below diagram shows the communication between Workload XM on-premises and your workload clusters through Telemetry Publisher. Where, the Workload XM service is installed on the left cluster, including the Workload XM main component services, and the Workload clusters on the right contain the services required to run your workload processes. Telemetry data collected by Telemetry Publisher is passed from these clusters to the Workload XM on-premises cluster.



System Requirements

Lists the minimum supported system requirements for your Workload XM on-premises cluster.

Before you install Workload XM on-premises, you must verify that your environment contains the minimum supported requirements for software, hardware, and networks.

Hardware Requirements

Lists the minimum supported hardware requirements for your Workload XM on-premises cluster.

In addition to the minimum supported hardware requirements for the services that you have installed on your Workload cluster nodes, you must verify that the dedicated cluster for Workload XM on-premises also contains the recommended minimum hardware requirements.

The recommended minimum hardware requirements for the Workload XM on-premises cluster are:

- A computer cluster of 5 nodes that hosts Workload XM.
- Where, each computer node in the Workload XM cluster must contain a minimum of:
  - 16 CPU cores
  - 64 GB of RAM
  - 12 TB of disk space

Note: To prevent issues from impacting operations other than those performed by Telemetry Publisher, such as sending data to Workload Manager, Cloudera highly recommends that the host server on which you assigned the Telemetry Publisher Service role is allocated its own dedicated disk.

Supported File Systems

Lists the supported file systems for your Workload XM on-premises cluster.

The following files systems are supported:

- Hadoop Distributed File System (HDFS)
- Amazon Simple Storage Service (S3)
- Azure Data Lake Storage (ADLS)

Supported Operating Systems

Lists the supported operating systems for your Workload XM on-premises cluster.

You must verify that the dedicated cluster for Workload XM on-premises runs on one of the supported Linux operating systems listed in the following table:

Table 1: Supported Linux Operating Systems

Product	Linux Version
CentOS Enterprise Linux	7.6, 7.5, 7.4
Red Hat Enterprise Linux	7.6, 7.5, 7.4

Note: All the nodes within the Workload XM cluster must run the same version of the Workload XM supported Linux operating system.

Supported Cloudera Versions

Lists the supported Cloudera platform and software versions for your Workload XM on-premises cluster and Workload clusters.

The following table lists the supported Cloudera platform and software for running Workload XM on-premises:

Table 2: Supported Cloudera Platform and Software for your Workload XM On-premises Cluster

Product	Version
CDP Private Cloud Base	7.0.3 or later
Cloudera Manager	7.0.3 or later

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

Table 3: Supported Cloudera Platforms and Software for your Workload Clusters

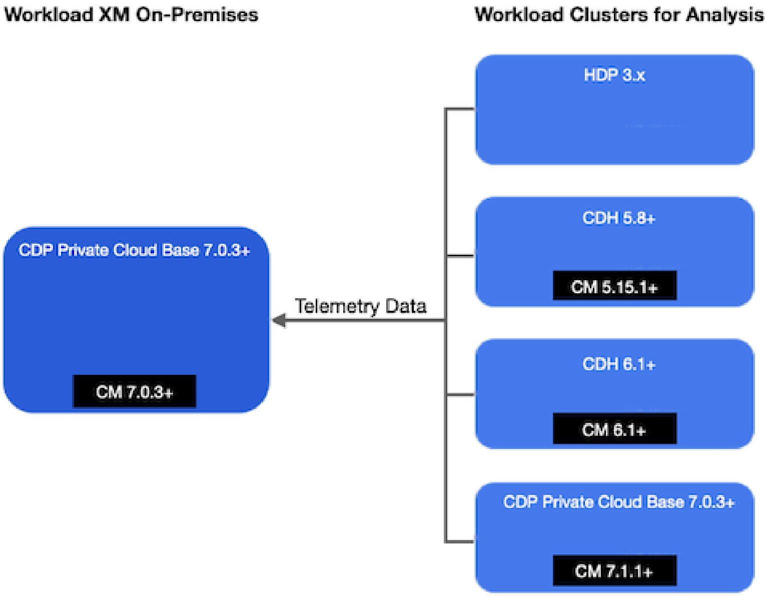
Cluster	Version	Cloudera Manager
HDP 3.x cluster		
CDH 5.x cluster	CDH version 5.8 and later	Cloudera Manager version 5.15.1 and later
CDH 6.x clusters	Cloudera Manager version 6.1 and later	Cloudera Manager version 6.1 and later
CDP 7.x clusters	Private Cloud Base 7.0.3 or later clusters	Cloudera Manager version 7.1.1 or later

Unsupported Versions

The following versions are not supported:

- CDH 6.0
- Cloudera Manager 6.0 and 7.0.3

The following figure shows an example of the Cloudera versions that are supported by Workload XM:



Network Port Requirements

Lists the network port numbers and their respective protocols used by Workload XM and dependent services.

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

Protocols are defined as follows:

- **UI Port** (ui.port), which serves the Workload XM user interface (UI) and communicates using HTTPS when TLS/SSL is enabled, otherwise it communicates using HTTP.
- **API Port** (api.port), which listens for REST calls to API-based servers. It communicates using HTTPS when TLS/SSL is enabled, otherwise it communications using HTTP.
- **Metrics Port** (webservice.port), which exposes an interface to the metrics that the Workload XM roles collect.
- **GRPC Port** (grpc.port), which listens for gRPC requests against the backend servers. This protocol is used for inter-role communication.

The following table lists the port numbers that must be enabled for Workload XM communication.

Table 4: Workload XM port numbers

Service	Web Port Number	gRCP Port Number
API Server	12011, 12012	
Databus API Server	12021, 12022	
Analytic Database Server	12031	12032
Baseline Server	12041	12042
Databus Server	12051	12052
Entities Server	12061	12062
Pipeline Server	12071	12072

Service	Web Port Number	gRCP Port Number
Admin API Server	12111	12112

The table below contains the complete list of services, protocols, and their default port number values.

Table 5: Default network port numbers

Service	UI Port (ui.port)	API Port (api.port)	Metrics Port (webservice.port)	gRPC Port (grpc.port)
Console Server	12001			
API Server		12012	12011	
Databus API Server		12022	12021	
Analytic Database Server			12031	12032
Baseline Server			12041	12042
Databus Server			12051	12051
Entities Server			12061	12062
Pipelines Server			12071	12072
SDX Server			12081	12082
Admin API Server			12111	12112

The following services are exposed service-wide:

- The Phoenix Query Server Port (phoenix.queryserver.port), which is the port for the Phoenix Query Server used by Workload XM.
- The Impala Daemon Port (impala.daemon.port), which is the port for the Impala Daemon used by Workload XM.

Installation Prerequisites

The tasks that must be completed before you install Workload XM on-premises.

Configure the Java Heap Requirements

Setting the supported Java heap size for the Zookeeper, HBase, HDFS, and Phoenix services, ensures the long-term success of a Workload XM deployment.

About this task

Describes how to set the Java heap size in Cloudera Manager for the Workload XM services listed in the following table.

Table 6: Java Heap Size Settings

Service	Size Setting Name	Minimum Value
ZooKeeper	Java Heap Size of ZooKeeper Server in Bytes	4 GB
HBase	HBase RegionServer(RegionServer Default Group)	16 GB
HDFS	Java Heap Size of NameNode	4 GB
Phoenix	Phoenix Query Server Max Heapsize	4 GB

Procedure

1. In a supported web browser on the Workload XM on-premises cluster, log in to Cloudera Manager.
2. In Cloudera Manager, select **Clusters** and then select the *Service* name. For example, Zookeeper.
3. In the *Service* name page, click the **Configuration** tab and then search for the *Size Setting Name*. For example, in the Search field, enter **java heap**, which locates the **Java Heap Size of ZooKeeper Server in Bytes** setting for the Zookeeper product.
4. Change the setting to the minimum supported value or higher for the service. For example, change the **Java Heap Size of ZooKeeper Server in Bytes** setting, to 4 GiB.
5. Click **Save**.
6. Repeat theses steps for each service using the above **Java Heap Size Settings** table.

Configure Performance Improvements

Improve the performance of Workload XM by configuring the ZooKeeper and HBase property settings.

About this task

Describes how to set the ZooKeeper and HBase property values listed in the following table.:

Table 7: Performance Improvement Settings

Service	Property	Value
ZooKeeper	maxClientCnxns	300
HBase	hbase.regionserver.handler.count	40
	hbase.hstore.blockingStoreFiles	100
HBase	hbase.ipc.server.max.callqueue.size	2147483648 bytes (2GiB)

Procedure

1. In a supported web browser on the Workload XM on-premises cluster, log in to Cloudera Manager.
2. In Cloudera Manager, select **Clusters** and then the *Service* name. For example, ZooKeeper.
3. In the *Service* name page, click the **Configuration** tab, and then search for the *Property* name. For example, in the Search field, enter **maxClientCnxn**, which locates the **Maximum Client Connections** setting for ZooKeeper.
4. Change the setting to the value for the service as listed in the **Performance Improvement Settings** table. For example, change the **Maximum Client Connections** setting, to 300.
5. Repeat steps 2 to 4 for the hbase.regionserver.handler.count and the hbase.hstore.blockingStoreFiles properties listed for the HBase service.
6. For the hbase.ipc.server.max.callqueue.size property do the following:
   a. In the HBase **Configuration** tab, search for **safety valve**.
   b. In the **HBase Service Advanced Configuration Snippet (Safety Value) for hbase-site.xml** section, click **Add Another**.
   c. Add the **hbase.ipc.server.max.callqueue.size** setting and set the value to 2147483648 bytes (2GB).
7. Restart HBase, ZooKeeper, and any other dependent services.

Confirm Installation of Main Components

Lists the Workload XM main component services. These services must be installed before installing Workload XM.

You must verify that the following services are installed on the cluster in which Workload XM is to be installed.

- In a supported web browser on the Workload XM on-premises cluster, log in to Cloudera Manager.

- In Cloudera Manager, select **Clusters**, and then confirm that your Workload XM on-premises cluster has the following required services:
- HBase
- HDFS
- Hive
- (Optional) Hue.

  Note: Though an optional service, Cloudera recommends installing Hue on the cluster in which Workload XM is to be installed as it provides troubleshooting and extraction services.

- Impala
- Phoenix
- ZooKeeper

Note: To optimize performance, Cloudera highly recommends that you do not install any other services on the Workload XM cluster.

Downloading the Installation Files

Lists the Cloudera download archive URLs for accessing the Workload XM installation files.

Important: The Workload XM installation parcel and files must be downloaded to a computer that is on the same network as the Workload XM on-premises cluster.

To download the Workload XM installation components:

1. Verify that you have an active Workload XM subscription agreement, license key, and access authentication credentials for the Cloudera Workload XM download archive repository. For information on how to obtain these contact your Cloudera sales representative.

   Note: Your Workload XM download credentials are not the same as the access credentials you use for the Cloudera support portal.

2. In a web browser on a computer that is on the same network as the Workload XM on-premises cluster, enter the repository URL location for each component, as listed in table Workload XM 2.1.3.
3. In the Sign In dialog box, enter the user name and password access authentication credentials that you received.

   Tip: To directly download the components, you can use your access authentication credentials for the Cloudera Workload XM download archive repository as part of the URL.

   For example, `https://username:password@archive.cloudera.com/p/wxm/2.1.3/WXM-2.1.3.2.1.3-b9-7082632-el7.parcel`

4. Verify that you have downloaded the following:

   - Parcel: WXM-*version_build*-el7.parcel
   - SHA: WXM-*version_build*-el7.parcel.sha
   - CSD: WXM-*version_build*.jar

Table 8: Workload XM 2.1.3

Repository Location	Type	Description
https://archive.cloudera.com/p/wxm/2.1.3/manifest.json	manifest.json	For Advanced users who are installing the Workload XM installation parcels in a local parcel repository.
https://archive.cloudera.com/p/wxm/2.1.3/WXM-2.1.3.2.1.3-b9-7082632-el7.parcel	parcel	The Workload XM installation parcel for a RedHat RHEL 7 or CentOS operating system.

Repository Location	Type	Description
https://archive.cloudera.com/p/wxm/2.1.3/ WXM-2.1.3.2.1.3-b9-7082632-el7.parcel.sha	sha file	The Workload XM shell archive parcel for a RedHat RHEL 7 or CentOS operating system.
https://archive.cloudera.com/p/wxm/2.1.3/ WXM-2.1.3.2.1.3-b9-7082632.jar https://archive.cloudera.com/p/wxm/2.1.3/ WXM-2.1.3.2.1.3-b9-7082632.jar.sha	CSD	Custom service descriptor jar files that enable you to install Workload XM on-premises.

Deploying the Installation Files

Steps for copying the Workload XM installation files from the computer where the files were downloaded to the Cloudera Manager Server parcel directories on the Workload XM on-premises cluster.

About this task

Describes how to deploy the downloaded Workload XM installation files to the on-premises cluster on which you plan to install Workload XM.

Procedure

1. Verify that you have the domain name of the Cloudera Manager Server host on the Workload XM on-premises cluster.
2. In a terminal on the computer where the installation files were downloaded, log in to the Cloudera Manager Server host and verify that you can establish a secure shell (SSH) and a secure copy protocol (SCP) connection between the computer where the installations files were downloaded and the Cloudera Manager Server host.
3. Go the directory where the Workload XM installation files were downloaded.
4. As the root user, SSH to the Cloudera Manager Server host.

   For example,

   ```
   sh root@domainname
   ```

5. In the directory where the WXM installation files were downloaded, do the following:

   a) Using the SCP protocol, copy the Workload XM parcel files to the `/opt/cloudera/parcel-repo` directory of the Cloudera Manager Server by entering the following command:

   ```
   scp WXM-version_build-el7.parcel root@cm_mgr_server_host:/opt/cloudera/p
   arcel-repo/
   scp WXM-version_build-el7.parcel.sha root@cm_mgr_server_host:/opt/cloude
   ra/parcel-repo/
   ```

   b) Copy the WXM CSD file to the `/opt/cloudera/csd` directory of the Cloudera Manager Server by entering the following command:

   ```
   scp WXM-version_build.jar root@cm_mgr_server_host:/opt/cloudera/csd/
   ```

6. In Cloudera Manager Server go to the `/opt/cloudera/parcel-repo` and the `/opt/cloudera/csd` directories and set the ownership of the copied files and change the read, write, and execute permissions to 644 by entering the following commands:

   ```
   chown cloudera-scm:cloudera-scm /opt/cloudera/parcel-repo/WXM-*;
   chmod 644 /opt/cloudera/parcel-repo/WXM-*;

   chown cloudera-scm:cloudera-scm /opt/cloudera/csd/WXM-*;
   chmod 644 /opt/cloudera/csd/WXM-*;
   ```

7. Restart the Cloudera Manager Server by entering the following command:

   ```
   service cloudera-scm-server restart
   ```
8. In a supported web browser, log in to Cloudera Manager on the Workload XM on-premises cluster.
9. In Cloudera Manager, select **Clusters**, and locate and select **Cloudera Management Service**.

   The Cloudera Management Service page opens.
10. From the **Actions** menu, click **Restart**.
11. In the Restart message, confirm restarting the management roles by clicking **Restart**.

Activating the Workload XM Parcel

Distribute the Workload XM installation files on all the nodes in the Workload XM on-premises cluster.

About this task

Describes how to activate the Workload XM installation parcel for an on-premises installation.

Procedure

1. In a supported web browser, log in to Cloudera Manager on the Workload XM on-premises cluster.
2. In Cloudera Manager, select **Hosts** and then **Parcels**.
3. In the Parcels page, verify that Cluster 1 is the Workload XM on-premises cluster.
4. From the **Parcel Name** section, locate and select **WXM** and then click **Distribute**.
5. When the **Distributed** indicator appears, click **Activate**.
6. In the Activate WXM confirmation message, click **OK**.

Results

The indicators for the WXM parcel are displayed as **Distributed** and **Activated**.

Securing the Workload XM Service Data

Describes how to enable secure connections and access authenticity when transferring data between components of Workload XM and your data.

Workload XM on-premises stores your workload data in HDFS and HBase, where the HDFS data is created in the root path and the directories have `wxm:impala` ownership. Configuring Kerberos and TLS/SSL ensures access authenticity and protects connections to your data.

Configuring Kerberos

Workload XM must be able to create Phoenix tables in data storage. If you are installing Workload XM on-premises in a Kerberized environment, it must be able to securely create these tables.

About this task

Describes how to add the wxm user as a HBase superuser, which securely enables Workload XM to create and store Phoenix tables.

Procedure

1. In a supported web browser on the Workload XM on-premises cluster, log in to Cloudera Manager.

2. In Cloudera Manager, select **Hosts** and then **Parcels**.
3. In the Parcels page, select the HDFS service.
4. In the HDFS-1 page, click the **Configuration** tab.
5. In the Search field, enter **hbase superusers**, which displays the HBase Superusers property.
6. In the HBASE-1 (Service-Wide) field, enter **wxm**, which adds the wxm user as a HBase superuser.

   Tip: If the HBASE-1 (Service-Wide) field is not visible, click the plus icon.

7. Click **Save Changes**.

Configuring TLS

Enable secure connections for data transfers and user access with either the Transport Layer Security (TLS) protocol or the Secure Socket Layer (SSL) protocol, which ensures access authenticity and securely protects your data.

Note: Cloudera recommends that you configure your cluster to use auto-TLS, which eases the process of configuring TLS/SSL.

TLS/SSL is supported between the following services:

- The supported web browser and the Workload XM UI.
- Telemetry Publisher and the Workload XM API.
- The Workload XM UI and the Workload XM API.
- The Workload XM Servers and Impala.

Configure the TLS properties based on the edge connection that you want to encrypt.

The following tables list the property settings for enabling TLS/SSL encrypted communication between the Workload XM system components:

- The supported web browser connected to the Workload XM UI.
- The Console Server and other REST Clients connected to the Admin API Server, the API Server, and the Databus API Server.
- The Pipeline Server, the Analytic Database Server, the Entities Server, the Databus Server, and a SDX Server connected to Impala Server.

Table 9: TLS/SSL parameters for a secure connection between your browser and the Workload XM UI

Component	Property	Value
Console Server	TLS/SSL Server Private Key File (PEM)	ssl.privatekey.path
Console Server	TLS/SSL Server Certificate File (PEM)	ssl.cert.path
Console Server	TLS/SSL Private Key Password	ssl.privatekey.password
Console Server	Enable TLS/SSL	ssl.enabled

Table 10: TLS/SSL parameters for a secure connection between the Console Server and other REST clients and the Admin API Server, the API Server, and the Databus API Server

Component	Property	Value
Console Server	TLS/SSL Certificate Trust Store File	ssl.cacert.path
Admin API Server	TLS/SSL Certificate Trust Store File	ssl.trustStore.path

Component	Property	Value
Admin API Server API Server Databus API Server	Enable TLS/SSL	ssl.enabled
Admin API Server API Server Databus API Server	TLS/SSL Server JKS Keystore File Location	ssl.keyStore.path
Admin API Server API Server Databus API Server	TLS/SSL Server JKS Keystore File Password	ssl.keyStore.password
Admin API Server API Server Databus API Server	TLS/SSL Server JKS Keystore Key Password	ssl.keyManager.password

Table 11: TLS/SSL parameters for a secure connection between the Pipeline Server and several other servers to the Impala Server

Component	Property	Value
Pipelines Server Analytic Database Server Entities Server Databus Server SDX Server	TLS/SSL Client Trust Store File	ssl.trustStore.path
Pipelines Server Analytic Database Server Entities Server Databus Server SDX Server	TLS/SSL Client Trust Store Password	ssl.trustStore.password

Enabling Phoenix Operations in HBase

Sets the Phoenix service operations in HBase for your Workload XM environment. The properties are added in Cloudera Manager using safety valves, which safely enable the changes to the HBase service.

About this task

Enables the Phoenix service for the Workload XM environment by safely adding Phoenix properties in the HBase service.

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

Procedure

1. Verify that the **ZooKeeper maxClientCnxns** property was set to 300.
2. In a supported web browser on the Workload XM on-premises cluster, log in to Cloudera Manager.
3. In Cloudera Manager, select **Clusters**, **HBase**, and then click the **Configuration** tab.

4. In the Configuration page, search for the **HBase Service Advanced Configuration Snippet (Safety Valve) for hbase-site.xml** property.

   Tip: Entering the full property name in the Search field is not always required. For example, in this case you can enter *snippet* to locate the **HBase Service Advanced Configuration Snippet (Safety Valve) for hbase-site.xml** property.

5. Above the **Name** field of the **HBase Service Advanced Configuration Snippet (Safety Valve) for hbase-site.xml** property, click **View as XML**.

6. In the XML field, add the following either before or after the existing XML:

   Tip: Dragging the bottom right corner downwards increases the size of the field.

```
<property>
   <name>hbase.regionserver.wal.codec</name>
<value>org.apache.hadoop.hbase.regionserver.wal.IndexedWALEditCodec</val
ue>
   <description>Set hbase.regionserver.wal.codec to enable custom Write
Ahead Log ("WAL") edits to be written</description>
</property>
   <property>
<name>hbase.region.server.rpc.scheduler.factory.class</name>
<value>org.apache.hadoop.hbase.ipc.PhoenixRpcSchedulerFactory</value>
   <description>Factory to create the Phoenix RPC Scheduler that uses s
eparate queues for index and metadata updates</description>
</property>
<property>
   <name>hbase.rpc.controllerfactory.class</name>
<value>org.apache.hadoop.hbase.ipc.controller.ServerRpcControllerFactory</
value>
   <description>Factory to create the Phoenix RPC Scheduler that uses sepa
rate queues for index and metadata updates</description>
</property>
<property>
  <name>phoenix.functions.allowUserDefinedFunctions</name>
   <value>true</value>
   <description>enable UDF functions</description>
</property>
<property>
   <name>phoenix.queryserver.serialization</name>
   <value>JSON</value>
   <description>serialization format between client and query server</d
escription>
</property>
<property><name>hbase.server.keyvalue.maxsize</name>
   <value>52428800</value>
   <description>limits max file size for blobs</description>
</property>
<property>
   <name>phoenix.schema.isNamespaceMappingEnabled</name><value>true</val
ue>
</property>
<property>
   <name>hbase.ipc.server.max.callqueue.size</name>
<value>2147483648</value>
</property>
```

7. Search for the **Write-Ahead Log (WAL) Codec Class** property and verify that the property is set to the following value:

```
org.apache.hadoop.hbase.regionserver.wal.IndexedWALEditCodec
```

8. Do the following:

   a. Search for the **Maximum Size of HBase Client KeyValue** property and set the value to 50 Mib.
   b. Search for the **HBase RegionServer Handler Count** property and set the value to 40.
   c. Search for the **HStore Blocking Store Files** property and set the value to 100.

9. (Optional) If you are installing Workload XM on a Kerberized environment, search for the **HBase Superusers** property and verify that the **wxm** user is added.
10. Click, **Save Changes**.
11. Back in the Cloudera Manager Home page, select **Clusters**, **Phoenix**, and then click the **Configuration** tab.

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

12. In the Configuration page, search for the **Query Server Advanced Configuration Snippet (Safety Valve) for phoenix-site.xml** property by entering *snippet*.
13. Click View as XML and then in the XML field, add the following either before or after the existing XML:

```
<property>
   <name>phoenix.queryserver.serialization</name>
   <value>JSON</value>
   <description>serialization format between client and query server</description>
</property>
<property>
   <name>phoenix.schema.isNamespaceMappingEnabled</name>
   <value>true</value>
</property>
```

    Note: This step ensures that the configuration setting for the `phoenix.schema.isNamespaceMappingEnabled` property is consistent on both the client and the server.

14. Click, **Save Changes**.

15. Apply your changes and restart the HBase and Phoenix services, by doing the following:

a) Back in the Cloudera Manager Home page, select the **Status** tab and then from the **Actions** menu, denoted by the vertical ellipses icon, select **Deploy Client Configuration**.

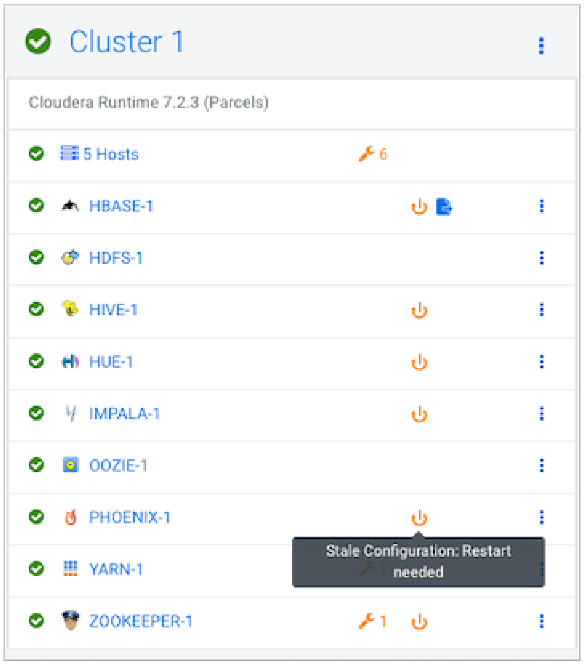
b) In the Deploy Client Configuration message, confirm deployment by clicking **Deploy Client Configuration**.

c) Monitor the progress of the client's configuration deployment until you see the *successfully deployed* message.

d) Click **Close**.

e) Back in the Cloudera Manager Home page, select the **Status** tab.

Notice that the **Stale Configuration: Restart needed** indicator is displayed for both the HBase and Phoenix services.



Stale Configuration: Restart needed

f) Restart the HBase and Phoenix services by doing one of the following:

- Click the HBase or Phoenix vertical ellipses icon and then select **Restart**.
- Click the HBase or Phoenix Stale Configuration indicator.

g) In the Stale Configurations page, click **Restart Stale Services**.

h) In the Restart Stale Services page, select the **Re-deploy client configuration** check box and click **Restart Now**.

i) Monitor the restart progress until the *All requested services successfully restarted* message appears and then click **Finish**.

Adding the Phoenix Query Server Role

Assign the Phoenix Query Server role to all the hardware devices in the Workload XM environment.

About this task

Describes how to assign the Phoenix Query Server role to all your hosts.

Procedure

1. In a supported web browser on the Workload XM on-premises cluster, log in to Cloudera Manager.
2. In Cloudera Manager, select **Clusters**, **Phoenix**, and then from the **Actions** menu, select **Add Role Instances**.
3. In the Add Role Instances to PHOENIX page, click inside the **Query Server x *n*** field, which opens the Hosts Selected page.

4. Add the Query Server role to all hosts by doing the following:

    a. Select the check box by the side of each host, which adds a Query Server role icon in the **Added Roles** column for each selected host.
    b. Click **OK**, which takes you back to the **Add Role Instances to PHOENIX** page where the **Query Server x *n*** field is now populated with the selected host names.
    c. Click **Continue**.
    d. In the Review Changes page, verify the changes and click **Continue**.
    e. Click **Finish**.
5. Back in the Cloudera Manager Home page, select **Clusters**, **Phoenix**, and then click the **Instances** tab.
6. Select the check box by the side of each host.
7. From the **Actions for Selected** list, select **Restart**.
8. In the Restart message, confirm restarting the hosts by clicking **Restart**.
9. Monitor the progress until the *Successfully restarted service* message appears for each restarted host and then click **Close**.

Deploying Workload XM

Steps for successfully configuring and installing Workload XM on-premises on all nodes in the Workload XM cluster.

About this task

Describes how to install Workload XM on-premises on all nodes in the Workload XM cluster.

Before you begin

The following tasks must be completed before deploying Workload XM.

- Verify that you successfully downloaded, distributed, and activated the WXM parcel.
- Verify that you assigned the Phoenix Query Server role to all the hosts in the Workload XM environment, by selecting the **Status** tab and then **Roles** in the Cloudera Home page, and then confirming that each host displays the QS icon.
- Verify that you have the following TLS/SSL key pair values, which you will be required to supply during the deployment task:
    - The location of your TLS/SSL private key file.
    - The location of your TLS/SSL certificate file.
    - the password of your TLS/SSL private key.
- Verify that you recorded a Phoenix and Impala Daemon host name, which you will be required to supply during the deployment task, by doing the following:
    1. From the Cloudera Manager's Home page, select **Phoenix** and then **Instances**. Record the host name of one of the Query Server hosts.
    2. From the Cloudera Manager's Home page, select **Impala** and then **Instances**. Record the host name of one of the Query Server hosts.

    Tip: Click the CLOUDERA Manager icon to go back to Cloudera Manager's Home page.

Procedure

1. Verify that you are in a supported web browser on the Workload XM on-premises cluster and have logged in to Cloudera Manager.
2. From the Cloudera Manager's Home page, select the **Status** tab.

3. From the cluster Actions menu, denoted by the vertical ellipses icon, select **Add Service**.
4. From the **Service Type** column in the Service to Cluster page, locate **Workload XM**.
5. Select **Workload XM** and click **Continue**.

   The Add Workload XM Service to Cluster 1 page opens.
6. In the Assign Rolls panel, click **Continue**.
7. In the **Review Changes** panel do the following:

   a. If not visible, display the **Phoenix Query Server Host** and the **Impala Daemon Host** value entry fields by clicking the plus (+) icon under Workload XM (Service-Wide).
   b. In the **Phoenix Query Server Host** field, enter the host name of the Phoenix Query Server that you recorded.
   c. In the **Impala Daemon Host** field, enter the host name of the Impala Daemon that you recorded.
   d. In the **Console Service TLS/SSL Server Private Key File** field, enter the location of your TLS private key file.
   e. In the **Console Service TLS/SSL Server Certificate File** field, enter the location of your TLS certificate file.
   f. In the **Console Service TLS/SSL Private Key password** field, enter the password of your TLS private key.
   g. For the **Console Service TLS/SSL Server CA Certificate** field, do nothing by leaving this field blank.
   h. (Optional) Scroll through the rest of the properties and make your changes.
   i. When satisfied with your changes, click **Continue**.

   The Workload XM service is deployed and configured on the Workload XM cluster and its progress is displayed.
8. When completed, as denoted by the **Status** field displaying **Finished**, click **Continue**.
9. In the Summary panel, click **Finish**.
10. In the Cloudera Manager Home page, select the **Status** tab and locate the **Cloudera Management Service** section.
11. From the Actions menu, denoted by the vertical ellipses icon, select **Restart**.
12. In the Restart message, confirm restarting the Cloudera Management Service by clicking **Restart**.
13. Monitor the restart progress until the *Successfully restarted service* message appears and then click **Close**.

Results

On the Cloudera Manager Home page, the Workload XM service appears in the list of services.

Laying Out Components

Horizontally scaling improves performance by enabling multiple devices to share the processing and memory workload. Cloudera recommends that you leverage the Workload XM on-premises cluster resources by installing its components as described.

About this task

Describes how to display your current layout and how to layout the Workload XM services for optimum performance.

The following table lists the components and the layout for a five node cluster. Where,

- One node must include *all* the Workload XM component role types.
- The Databus API Server, Databus Server, Analytic Database Server, Baseline Server, Entities Server, SDX Server, and Pipelines Server role types can scale out to multiple nodes. As listed in the Node 2, 3, and 4 columns.
- Due to inter service dependencies, the following role types are grouped. Where, if one of the components is on a host then all the other components in that group must be on the host, which is enforced by Cloudera Manager:

  - Databus API Server, Databus Server.
  - Analytic Database Server, Baseline Server, Entities Server, SDX Server, and Pipelines Server.
  - Admin API Server, API Server, Console Server.

  For example, if you add a new Databus API Server, you must also add a Databus Server to that node.

- Configure multiple Phoenix Query Server hosts, which reduces bottlenecks. Where, the number of Phoenix Query Server hosts should be proportional to the number of Workload XM roles.

  For example, if you have roles on 5 nodes, at least 5 Query Servers are recommended for Phoenix. Workload XM internally balances loads on those hosts.

Important: Only one host must be configured for Impala.

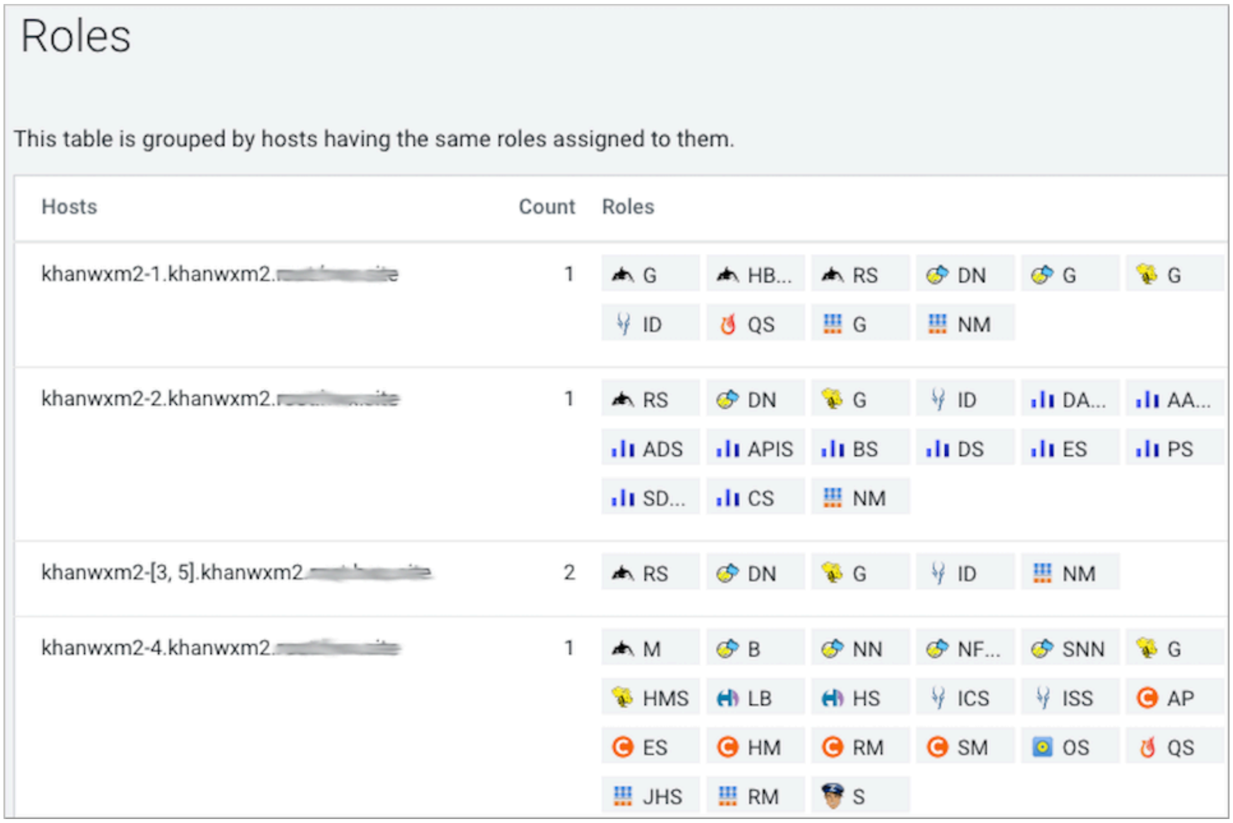
Table 12: Component layout for a five node Workload XM cluster

Service	**Node 1** **(All master components of all services)**	**Node 2, 3, 4** **(Worker nodes + ZooKeeper + WXM processing components)**	**Node 5** **(Worker nodes + WXM processing components + WXM UI)**
Cloudera Management	• `Alert Publisher` • `Event Server` • `Host Monitor` • `Reports Manager` • `Service Monitor`		
HBase	• `Gateway` • `Master` • `Thrift Server (optional)`	• `Gateway` • `RegionServer`	• `Gateway` • `RegionServer`
HDFS	• `Balancer` • `Gateway` • `NameNode` • `NFS Gateway (optional)` • `SecondaryNameNode`	• `DataNode` • `Gateway`	• `DataNode` • `Gateway`
Hive	• `Gateway` • `Metastore Server` • `HiveServer`	• `Gateway`	• `Gateway`
Hue (Optional)	• `Load Balancer` • `Hue Server`		
Impala	• `Catalog Server` • `StateStore`	• `Impala Daemon`	• `Impala Daemon`
Phoenix	• `Query Server`	• `Query Server`	• `Query Server`
Workload XM		• `Analytic Database Server` • `Baseline Server` • `Databus API Server` • `Databus Server` • `Entities Server` • `Pipelines Server` • `SDX Server`	• `Admin API Server` • `Analytic Database Server` • `API Server` • `Baseline Server` • `Console Server` • `Databus API Server` • `Databus Server` • `Entities Server` • `Pipelines Server` • `SDX Server`
ZooKeeper		• `Server`	

Procedure

1. In a supported web browser on the Workload XM on-premises cluster, log in to Cloudera Manager.
2. In Cloudera Manager, select **Hosts** and then **Roles**.

   The roles assigned to each node appear as shown in the below example:



Roles

This table is grouped by hosts having the same roles assigned to them.

Hosts	Count	Roles
khanwxm2-1.khanwxm2.[redacted]	1	G HB... RS DN G G ID QS G NM
khanwxm2-2.khanwxm2.[redacted]	1	RS DN G ID DA... AA... ADS APIS BS DS ES PS SD... CS NM
khanwxm2-[3, 5].khanwxm2[redacted]	2	RS DN G ID NM
khanwxm2-4.khanwxm2.[redacted]	1	M B NN NF... SNN G HMS LB HS ICS ISS AP ES HM RM SM OS QS JHS RM S

3. Compare your existing layout with the layout described in the **Component layout for a five node Workload XM cluster** table above.
4. (Optional) To leverage resources, spread the Workload XM (WXM) roles throughout the cluster. For more information on how to assign roles, see the Cloudera Manager documentation.

Troubleshooting Installation Issues

Provides guidelines when encountering issues with a Workload XM installation. If problems still exist, contact Cloudera support.

Failure Creating a Phoenix Schema

Mapping a Phoenix schema to a HBase namespace enables multitenancy. Before running a Phoenix job you must verify that namespace mapping is enabled in the HBase safety valve. Once enabled, tables that are created with the Phoenix schema are mapped to the HBase namespace.

The following example shows a stack track error report that was generated after running a Phoenix job. It shows that the phoenix schema namespace mapping property is not enabled:

```
Role Log
 at org.apache.phoenix.shaded.org.eclipse.jetty.util.thread.strategy.Execut
eProduceCon
```

```
sume.executeProduceConsume(ExecuteProduceConsume.java:303)
 at org.apache.phoenix.shaded.org.eclipse.jetty.util.thread.strategy.Exec
uteProduceConsume.produceConsume(ExecuteProduceConsume.java:148)
 at org.apache.phoenix.shaded.org.eclipse.jetty.util.thread.strategy.Execut
eProduceConsume.run(ExecuteProduceConsume.java:136)
 at org.apache.phoenix.shaded.org.eclipse.jetty.util.thread.QueuedThreadPool
.runJob(QueuedThreadPool.java:671)
 at org.apache.phoenix.shaded.org.eclipse.jetty.util.thread.QueuedThreadPool
$2.run(QueuedThreadPool.java:589)
 at java.lang.Thread.run(Thread.java:748)
Caused by: java.sql.SQLException: ERROR 725 (43M08): Cannot create schema
 because config phoenix.schema.isNamespaceMappingEnabled for enabling name s
pace mapping isn't enabled. schemaName=SIGMA_DB
```

Solution:

Add the following property in the HBase safety valve:

```
<property><name>phoenix.schema.isNamespaceMappingEnabled</name><value>true</
value></property>
```

After adding the property, redeploy the client configurations and restart HBase and the dependent services.

Incorrect Installation Parcel Placement

Adding the Workload XM installation parcels in the wrong directory on the Cloudera Management Server host causes distribution and activation issues. The Workload XM installation parcel files must reside in the `/opt/cloudera/parcel-repo` directory.

About this task

Issues arise when the Workload XM installation parcels are incorrectly placed in the wrong directory and the Cloudera SCM server is restarted. This task discusses the type of error messaged generated, where to locate the parcel error messages, and what to do when you receive this type of message.

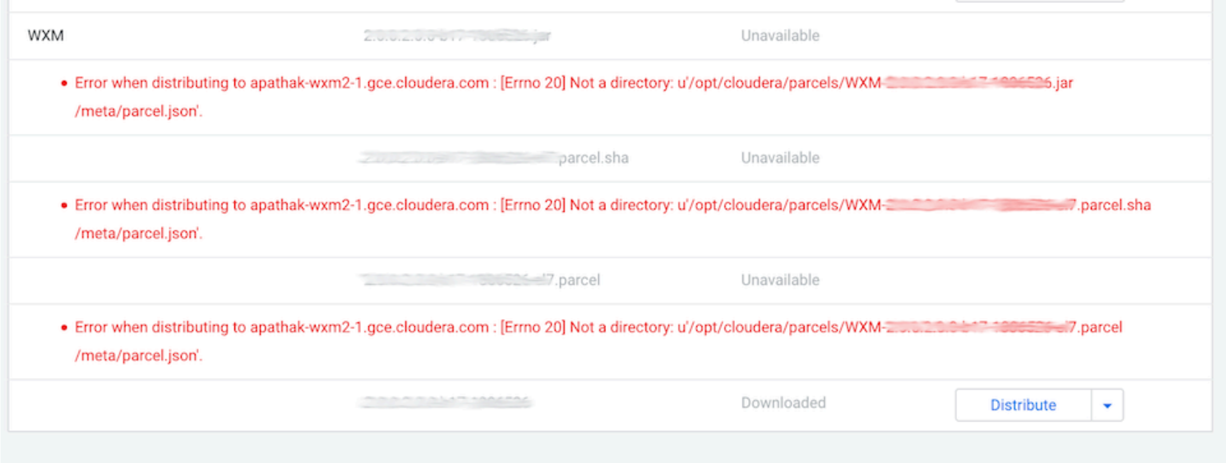
If your installation of Workload XM fails and you receive a message that reports a "getpwnam()" error, do the following:

Procedure

1. Verify that the Workload XM installation parcels are residing in the /opt/cloudera/parcel-repo directory of the Cloudera Management Server.

2. Verify whether the parcel is correctly distributed and activated, by going to the Parcels page in Cloudera Manager.

   An example of the errors displayed on the Parcel page during the Workload XM parcel distribution process is shown below:



WXM [redacted] Unavailable

- Error when distributing to apathak-wxm2-1.gce.cloudera.com : [Errno 20] Not a directory: u'/opt/cloudera/parcels/WXM-[redacted].jar /meta/parcel.json'.

[redacted]parcel.sha Unavailable

- Error when distributing to apathak-wxm2-1.gce.cloudera.com : [Errno 20] Not a directory: u'/opt/cloudera/parcels/WXM-[redacted]7.parcel.sha /meta/parcel.json'.

[redacted]7.parcel Unavailable

- Error when distributing to apathak-wxm2-1.gce.cloudera.com : [Errno 20] Not a directory: u'/opt/cloudera/parcels/WXM-[redacted]7.parcel /meta/parcel.json'.

[redacted] Downloaded

3. In a terminal, move the displaced parcel from the wrong directory to the `/opt/cloudera/parcel-repo` directory.
4. Restart the Cloudera SCM server and Cloudera SCM agent by using the following commands:

```
service cloudera-scm-server restart
```

```
service cloudera-scm-agent restart
```

Granting User Access

Workload XM supports two authentication methods for granting user access; Local authentication and the Lightweight Directory Access Protocol (LDAP). You configure user access to Workload XM in one of these supported authentication methods.

Granting Local Authentication

Granting user access using local authentication.

About this task

Describes how to locate the Workload XM local authentication directory and user authentication files in Cloudera Manager, and how to add a user, or remove or list existing users using the Console Server executable tool.

Procedure

1. In a supported web browser on the Workload XM on-premises cluster, log in to Cloudera Manager.
2. In Cloudera Manager, select **Clusters**, **Workload XM**, and then the **Configuration** tab.

3. Search for the following properties:

   - **User Authorization File Directory (user-file.dir)**, which is the local directory for storing the user authorization file required by the Console Server. By default, `/etc/wxm/conf`.
   - **User Authorization File Name (user-file.name)**, which is the name of the user authorization file required by the Console Server. By default, `user-file.json`.

     Note: If this file does not exist, it is created during the service startup and is then stored in the directory set by the `user-file.dir` parameter.

4. In a terminal, SSH to the cluster node that has the Workload XM Console role.
5. On the Workload XM host, go to the following directory by entering the following command:

   ```
   ${PARCELS_ROOT}/WXM/lib/thunderhead-sigma-console
   ```

6. According to your task, enter one of the following commands:

   - To add a user, enter the following command and then follow the prompts to create the user's user name and password:

     ```
     ./onprem-linux user add --user-file user-file.dir user-file.name
     ```

   - To remove a user, enter the following command:

     ```
     ./onprem-linux user remove --user-file user-file.dir user-file.name
     ```

   - To list existing users, enter the following command:

     ```

     ./onprem-linux user list --user-file user-file.dir user-file.name
     ```

7. (Optional) To access the help for other commands, enter the following command:

   ```
   ./onprem-linux -h
   ```

   Note: You cannot change a user's user name or password, instead you must first remove the user and then recreate the user with their new credentials. Also, if you attempt to edit a nonexistent user file, a prompt appears asking if you would like to create the file.

LDAP Authentication Properties

Granting user access using the Lightweight Directory Access Protocol (LDAP) .

Workload XM supports LDAP authentication through the following properties:

- Enable LDAP (ldap.enabled)
- LDAP URL (ldap.url)
- LDAP Bind User Distinguished Name (ldap.bind_dn)
- LDAP Bind Password (ldap.bind_password)
- LDAP Search Base (ldap.search_base)
- LDAP Search Filter Property (ldap.search_filter_property)
- LDAP Server CA Certificate (ldap.ca_cert)

Configuring Telemetry Publisher

Tasks for enabling Cloudera Telemetry Publisher, which collects and sends diagnostic information about job and query processes to Workload XM.

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

Note: Cloudera highly 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 Workload XM from affecting operations other than those performed by Telemetry Publisher.

Telemetry Publisher Pre-tasks

Describes how to enable endpoint services between Telemetry Publisher and Workload XM and data redaction, which provide secure data transfers and masks sensitive data.

Before configuring Telemetry Publisher you must complete the Telemetry Publisher pre-requisite tasks.

Configuring your Firewall

Connecting Telemetry Publisher to Workload XM through endpoint services creates a secure connection between your on-premises CDH cluster and the Workload XM cloud service.

The Cloudera Telemetry Publisher service collects metrics from various components in a CDH 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 Workload XM.

Enabling secure communication from an on-premises CDH cluster to a Workload XM cloud service that runs on an Amazon Web Services (AWS) cloud platform, requires that Telemetry Publisher connects to Workload XM through the following endpoint services:

- Endpoint #1 (EC2 service):

  ```
  https://dbusapi.us-west-1.sigma.altus.cloudera.com
  ```
- Endpoint #2 (S3 service):

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

Where, these endpoints map to a dynamic IP address in AWS us-west-2. For more information on the IP address ranges, see the Amazon documentation.

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

Redacting Data

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

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.

Note: To ensure that Telemetry Publisher only sends redacted data to Workload XM 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 configurations in the YARN **Redacted MapReduce Job Properties** property.

About this task

Steps for adding MapReduce job configurations in YARN that enable data redaction when MapReduce data is pulled from HDFS.

Procedure

1. In a supported web browser on the Workload XM on-premises 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 Jobs 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.

Enabling the Telemetry Publisher Service

Activating the Telemetry Publisher service for Workload XM on-premises.

About this task

Describes how to enable the Telemetry Publisher service for Workload XM on-premises.

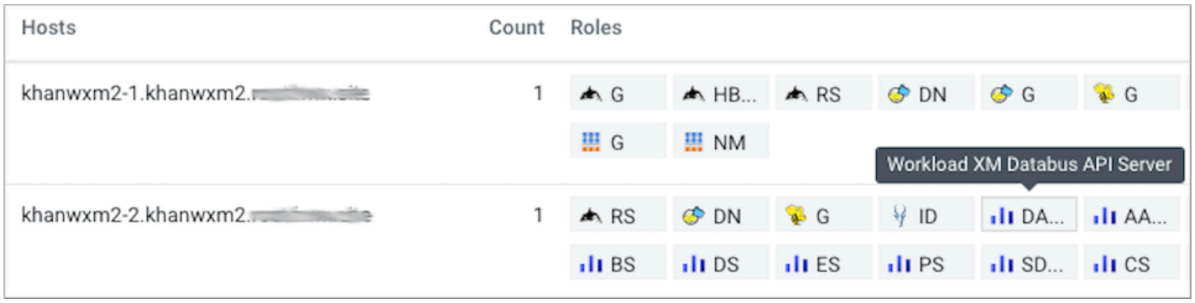
Before you begin

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 Workload XM license key text.
- The name of the node that contains the **Workload XM Databus API Server** role, by doing the following:

  1. In Cloudera Manager, select **Hosts** and then **Roles**.
  2. Search for the **Workload XM Databus API Server** role and record its host name. For example:

     Figure 1: Roles on the Workload XM On-Premises Cluster



Hosts	Count	Roles
khanwxm2-1.khanwxm2.[redacted]	1	G HB... RS DN G G G NM
khanwxm2-2.khanwxm2.[redacted]	1	RS DN G ID DA... AA... BS DS ES PS SD... CS

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 the following using the Workload XM Database API Server host name that you recorded as a prerequisite for these steps:

   ```
   telemetry.upload.job.logs=true
   telemetry.altus.url=http|
   https://Databus_API_Server_hostname:Databus_API_Server_port_number
   ```

   Where,

   - If you have enabled TLS/SSL for the Databus API Server (ssl.enabled), enter **https**.
   - If you have not enabled TLS/SSL for the Databus API Server, enter **http**.
   - By default, the Databus API Server port number is 12022.
4. Click **Save**.
5. Back in the Cloudera Manager Home page, select **Administration** and then **External Accounts**.
6. Click the **Altus Credentials** tab and then click **Add Access Key Authentication**.
7. In the **Add Access Key Authentication** dialog box, do the following:

   a) In the **Name** field, enter an identifiable name for the access key.

   b) In the **Altus Access Key ID** field, enter your Workload XM license key text exactly as provided but without quotation marks and without trailing spaces.

   c) Click **Choose File** and then browse and select your WXM license private key file.

   > **Note:** The Workload XM license is not related to Altus, but acts as a pay-wall mechanism to use Workload XM.

   d) Click **Add**. Your credentials are displayed in the Altus Credentials tab.
8. Back in the Cloudera Manager Home page, select **Clusters** and then locate and select **Cloudera Management Service**.
9. From the **Actions** menu, select **Restart**.
10. In the Restart message, confirm restarting the Cloudera Management Service by clicking **Restart**.
11. Monitor the restart progress until the *Successfully restarted service* message appears and then click **Close**.

Enabling Key Trustee Keys

Accessing 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.

Associating a Workload Cluster with Telemetry Publisher

Steps for connecting your Workload cluster with the Telemetry Publisher service.

About this task

Describes how to associate a Workload Cluster with Telemetry Publisher by designating a host cluster with the Telemetry Publisher service role.

Note: If you are using Java 7, additional steps are required for adding the Telemetry Publisher service role.

Before you begin

The following pre-tasks must be completed before associating a Workload cluster with Telemetry Publisher.

- Verify that you have the JCE Policy installed before enabling the role in the Cloudera Manager Service.

  Note: If you are using JDK version 1.8.0_160 or earlier, verify that you have installed the JCE policy file as described in the Cloudera Manager documentation.

- Rename the Workload cluster with a human-readable name in Cloudera Manager.

  Workload XM identifies the cluster from a random string of 32 characters, such as `44a6e75e-8630-4773-9ea9-6272478e84c2`, which is difficult to identify and manage. Cloudera recommends completing the following task to rename your Workload cluster.

  To rename a workload cluster:

  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**.

Procedure

1. In a supported web browser on a Workload cluster, log in to Cloudera Manager.
2. In Cloudera Manager, select **Clusters** and then locate and select **Cloudera Management Service**.
3. From the **Actions** menu, select **Add Role Instances**.

   The Add Role Instances to Cloudera Management Service opens.
4. Do one of the following:
   - If a Telemetry Publisher role already exists, do nothing. Cloudera Manager does not let you add another.
   - If a Telemetry Publisher role does not exist, continue.
5. Select a host for the Telemetry Publisher by doing one of the following:
   - If you are using Java 8, click **Select a host** and in the Hosts Selected page, select the check box by the side of the required host and click **OK**.
   - If you are using Java 7, configure Telemetry Publisher as follows:

     a. Go back to Cloudera Management Service and click the **Configuration** tab.

     b. Under **Scope**, select **Telemetry Publisher**.

     c. In the **Search** field, enter **java configuration**, which displays the **Java Configuration Options for Telemetry Publisher** filter.

     d. In Telemetry Publisher Default Group field, add the following property:

     ```
     -Dhttps.protocols=TLSv1.2 -Dhttps.cipherSuites=TLS_RSA_WITH_AES_256_
     CBC_SHA256
     ```

     e. Click **Save Changes**.
6. Back in the Cloudera Manager Home page, select **Clusters**, **Hive**, and then **Instances**.
7. Select the **Role Type** check box, which selects all the Hive Roles.

8. From the **Actions for Selected** list, select **Restart**.
9. In the Restart message, confirm restarting the HIVE roles by clicking **Restart**.
10. Monitor the restart progress until the *Successfully restarted service* message appears and then click **Close**

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 Workload XM.

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/HTTPS protocol to send telemetry information to Workload XM, which ensures that the data is encrypted. The proxy you use must support the HTTP CONNECT method in order to be able to pass through the encrypted messages. For more information, see the associated RFC.

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

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. 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.

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 strongly 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 later.

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 check box.
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 check box 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.

Logging in to the Workload XM Web User Interface

Steps for accessing the Workload XM web user interface for the first time.

About this task

Describes how to log in to Workload XM.

Before you begin

Verify that the following is completed:

- Workload XM is installed.
- Telemetry Publisher is enabled for Workload XM and your Workload clusters are associated with the service.
- Cloudera Manager is connected to Workload XM.

Procedure

1. In a supported web browser on the Workload XM on-premises cluster, log in to Cloudera Manager.
2. In Cloudera Manager, select **Clusters**, **Workload XM**, and then **Workload XM UI**.

   Which opens the Workload XM login page.
3. Enter your user name and password.

   Tip: By default, for testing purposes, you can enter **admin** for both the user name and password. Cloudera recommends changing this default setting as soon as possible.

4. From the Workload XM navigation panel, select **Clusters**.

   The Clusters page opens displaying the Workload clusters in your environment.

   Note: If no cluster names appear or a specific Workload cluster is not displaying, verify that you have associated the Telemetry Publisher service with the Workload cluster.

5. To start to view a Workload cluster's workload metrics, select the name of a Workload cluster.

   Note: At this time, the following sections of the Workload XM web UI are disabled or are not present:

   - Feedback link
   - Impala query potential SQL issues
   - Cluster Email Reports
   - HDFS Tables Scanned

Enabling File Size Reports

Steps for enabling file size reports from HDFS. These reports help you to identify where your data is stored inefficiently, such as in small files or partitions, which can cause performance issues.

About this task

Describes how to enable file size reports in Cloudera Manager.

Important: At this time the Workload XM File Size Report feature is only supported on CDH Workload clusters, version 6.3 to version 7.0, with Cloudera Navigator enabled. CDP Workload clusters are not supported.

Procedure

1. In a supported web browser on the Workload XM on-premises cluster, log in to Cloudera Manager.
2. In Cloudera Manager, select **Clusters** and then locate and select **Cloudera Management Service**.
3. Click the **Configuration** tab and in the **Search** field enter **small files**, which displays the small files properties.
4. Select the **Small Files Reporting: Enable Data Collection** check box.
5. From the **Small Files Reporting: HDFS Service for Data Staging** property list, select the HDFS service option that you require.
6. In the **Small File Reporting: HDFS Staging Location** field, enter the path and directory name that you require as the staging area for file size analysis.
7. Click **Save Changes**.

Upgrading Workload XM

Tasks for upgrading Workload XM on-premises.

Before upgrading to a new version of Workload XM on-premises, do the following:

1. Schedule the upgrade and inform your users of the Workload XM service interruption.
2. (Optional) Upgrade to the latest LINUX patch release.
3. Verify that your Workload XM cluster is running a supported version of Cloudera Manager and the Cloudera Data Platform (CDP) and that your Workload clusters are running a supported version of Cloudera Manager and Cloudera Platform.

4. Record the Cloudera Manager host name of the server that contains an installation of Workload Manager, as this value is required during upgrading.

   Tip: Log in to Cloudera Manager and record the name of the host running the Cloudera Manager UI.

5. Stop the Workload XM Service and its services.

   Important: Before stopping Workload XM and its services you must verify that there are no remaining messages in the ZooKeeper queue by doing the following:

   a. In a terminal log in to a host running ZooperKeeper and list the current queues by running the following command:

   ```
   /opt/cloudera/parcels/CDH/lib/zookeeper/bin/zkCli.sh -server [zk
   _server]:2181 ls /wxm/onprem/zkqueue
   ```

   The following terminal output is an example of a current queue output:

   ```
   HiveAudit, HiveHistoryProtobuf, HiveOnMrTable, ImpalaQueryProfile,
    LlapHistoryProtobuf, MrJhist, MrTaskLog, OozieWorkflow, Pse, SdxD
   etails, SparkEventLog, SparkTaskLog, TezHistoryProtobuf, YarnApp
   , YarnAppMetrics, sigmaadb-broadcast, upload-processing-update-q
   ueue
   ```

   b. Verify that all the queues are empty by running the following:

   ```
   for q_n in HiveAudit HiveHistoryProtobuf HiveOnMrTable ImpalaQue
   ryProfile LlapHistoryProtobuf MrJhist MrTaskLog OozieWorkflow Pse
    SdxDetails SparkEventLog SparkTaskLog TezHistoryProtobuf YarnApp
    YarnAppMetrics sigmaadb-broadcast upload-processing-update-queue
     do
       echo $q_n: $(/opt/cloudera/parcels/CDH/lib/zookeeper/bin/zk
   Cli.sh -server [zk_server]:2181 stat -w /wxm/onprem/zkqueue/${q_n}
    | grep "numChildren")
     done
   ```

   c. Verify that all the count values are **0**.

   d. If there are messages in the queue, do the following:

   1. Stop the DBUS services, which stops any new incoming data.
   2. Allow Workload XM to finish processing any existing messages.
   3. Stop any remaining components.

6. Download the latest Workload XM installation parcel and its checksum parcels from the Cloudera Downloads website.
7. Download the latest Workload XM installation CSD (.jar) file from the Cloudera Downloads website.

Upgrading version 2.1.0 to 2.1.3 of Workload XM

Steps for upgrading version 2.1.0 to 2.1.3 of Workload XM.

About this task

Describes how to upgrade from version 2.1.0 to 2.1.3.

Before you begin

These steps assume that you have:

- Scheduled the upgrade and informed your users of the Workload XM service interruption.

- Verified that your Workload XM cluster is running a supported version of Cloudera Manager and the Cloudera Data Platform (CDP) and that your Workload clusters are running a supported version of Cloudera Manager and Cloudera Platform.
- Recorded the Cloudera Manager host name of the server that contains an installation of Workload Manager, as this value is required during upgrading.

  Tip: Log in to Cloudera Manager and record the name of the host running the Cloudera Manager UI.

- Downloaded the 2.1.3 version of the Workload XM installation parcel and checksum from the Cloudera Downloads website.
- Downloaded the 2.1.3 version of the Workload XM installation CSD (.jar) file from the Cloudera Downloads website.
- Verified that there are no remaining messages in the ZooKeeper queue, by doing the following:

  1. In a terminal log in to a host running ZooperKeeper and list the current queues by running the following command:

     ```
     /opt/cloudera/parcels/CDH/lib/zookeeper/bin/zkCli.sh -server [zk_server]
     :2181 ls /wxm/onprem/zkqueue
     ```

     The following terminal output is an example of a current queue output:

     ```
     HiveAudit, HiveHistoryProtobuf, HiveOnMrTable, ImpalaQueryProfile, LlapH
     istoryProtobuf, MrJhist, MrTaskLog, OozieWorkflow, Pse, SdxDetails, Spar
     kEventLog, SparkTaskLog, TezHistoryProtobuf, YarnApp, YarnAppMetrics, si
     gmaadb-broadcast, upload-processing-update-queue
     ```

  2. Verify that all the queues are empty by running the following:

     ```
     for q_n in HiveAudit HiveHistoryProtobuf HiveOnMrTable ImpalaQueryProfil
     e LlapHistoryProtobuf MrJhist MrTaskLog OozieWorkflow Pse SdxDetails Spa
     rkEventLog SparkTaskLog TezHistoryProtobuf YarnApp YarnAppMetrics sigmaa
     db-broadcast upload-processing-update-queue
       do
         echo $q_n: $(/opt/cloudera/parcels/CDH/lib/zookeeper/bin/zkCli.sh -
     server [zk_server]:2181 stat -w /wxm/onprem/zkqueue/${q_n} | grep "numCh
     ildren")
       done
     ```

  3. Verify that all the count values are **0**.
  4. If there are messages in the queue, do the following:

     a. Stop the DBUS services, which stops any new incoming data.
     b. Allow Workload XM to finish processing any existing messages.
     c. Stop any remaining components.

- Stopped the Workload XM service and its services.

Procedure

1. Verify that the Workload XM Service and its services are stopped.
2. Do the following:

   a) In a terminal, SSH into the Cloudera Manager host server.
   b) Copy the downloaded parcel (.parcel) and its checksum (.sha) file to the `/opt/cloudera/parcel-repo` directory of the Cloudera Manager Server on the Workload XM on-premises cluster.
   c) Copy the downloaded CSD (.jar) file to the `opt/cloudera/csd` directory of the Cloudera Manager Server on the Workload XM on-premises cluster.

3. In the `/opt/cloudera/parcel-repo` directory, set the ownership of the parcel and sha files to **cloudera-scm:cloudera-scm**.

4. In the `/opt/cloudera/csd` directory, set the ownership of the CSD .jar file to **cloudera-scm:cloudera-scm**.
5. In a supported web browser on the Workload XM on-premises cluster, log in to Cloudera Manager.
6. In Cloudera Manager, select **Clusters**, **Parcels**, and then **Distribute**.

   In the list of parcels, the latest version appears with a gray Distribute label.
7. Activate the Workload XM installation files by selecting **Activate Only**. Do not restart.
8. From the Cloudera Manager Host, restart the Cloudera Manager Server by entering the following:

   ```
   service cloudera-scm-server restart
   ```
9. Navigate to the Workload XM service in the Cloudera Manager UI and then from the **Actions** menu do the following:

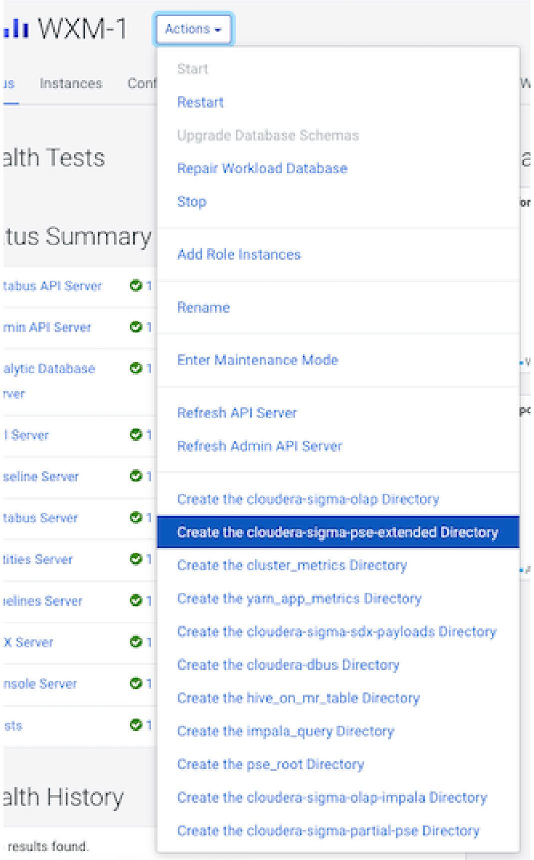
   Note: Verify if a Cloudera Sigma PSE extended directory has been created. If yes, you can skip the action for this component. This component is required by Workload XM and its services, redoing the action does not harm the upgrade but skipping the action when the component has not been created could cause issues.

   a) Run the command to create the Cloudera Sigma PSE extended directory by selecting **Create the cloudera-sigma-pse-extended Directory**.

   b) Run the command to upgrade the database schemas by selecting **Upgrade Database Schemas**.

   c) Restart the Workload XM service, by selecting **Restart**.

   As shown in the following image:


WXM-1
Actions
us
Instances
Conf
alth Tests
tus Summary
tabus API Server
min API Server
alytic Database
rver
l Server
seline Server
tabus Server
tities Server
elines Server
X Server
nsole Server
sts
alth History
results found.
Start
Restart
Upgrade Database Schemas
Repair Workload Database
Stop
Add Role Instances
Rename
Enter Maintenance Mode
Refresh API Server
Refresh Admin API Server
Create the cloudera-sigma-olap Directory
Create the cloudera-sigma-pse-extended Directory
Create the cluster_metrics Directory
Create the yarn_app_metrics Directory
Create the cloudera-sigma-sdx-payloads Directory
Create the cloudera-dbus Directory
Create the hive_on_mr_table Directory
Create the impala_query Directory
Create the pse_root Directory
Create the cloudera-sigma-olap-impala Directory
Create the cloudera-sigma-partial-pse Directory

10. Once all the roles have started, from the **Actions** menu, select **Repair Workload Database**.

Upgrading version 2.0.0 to 2.1.3 of Workload XM

Steps for upgrading version 2.0.0 to 2.1.3 of Workload XM.

About this task

Describes how to upgrade from version 2.0.0 to 2.1.3.

Before you begin

These steps assume that you have:

- Scheduled the upgrade and informed your users of the Workload XM service interruption.
- Verified that your Workload XM cluster is running a supported version of Cloudera Manager and the Cloudera Data Platform (CDP) and that your Workload clusters are running a supported version of Cloudera Manager and Cloudera Platform.
- Recorded the Cloudera Manager host name of the server that contains an installation of Workload Manager, as this value is required during upgrading.

  Tip: Log in to Cloudera Manager and record the name of the host running the Cloudera Manager UI.

- Downloaded the 2.1.3 version of the Workload XM installation parcel and checksum from the Cloudera Downloads website.
- Downloaded the 2.1.3 version of the Workload XM installation CSD (.jar) file from the Cloudera Downloads website.
- Verified that there are no remaining messages in the ZooKeeper queue, by doing the following:

  1. In a terminal log in to a host running ZooperKeeper and list the current queues by running the following command:

     ```
     /opt/cloudera/parcels/CDH/lib/zookeeper/bin/zkCli.sh -server [zk_server]
     :2181 ls /wxm/onprem/zkqueue
     ```

     The following terminal output is an example of a current queue output:

     ```
     HiveAudit, HiveHistoryProtobuf, HiveOnMrTable, ImpalaQueryProfile, LlapH
     istoryProtobuf, MrJhist, MrTaskLog, OozieWorkflow, Pse, SdxDetails, Spar
     kEventLog, SparkTaskLog, TezHistoryProtobuf, YarnApp, YarnAppMetrics, si
     gmaadb-broadcast, upload-processing-update-queue
     ```
  2. Verify that all the queues are empty by running the following:

     ```
     for q_n in HiveAudit HiveHistoryProtobuf HiveOnMrTable ImpalaQueryProfil
     e LlapHistoryProtobuf MrJhist MrTaskLog OozieWorkflow Pse SdxDetails Spa
     rkEventLog SparkTaskLog TezHistoryProtobuf YarnApp YarnAppMetrics sigmaa
     db-broadcast upload-processing-update-queue
       do
         echo $q_n: $(/opt/cloudera/parcels/CDH/lib/zookeeper/bin/zkCli.sh -
     server [zk_server]:2181 stat -w /wxm/onprem/zkqueue/${q_n} | grep "numCh
     ildren")
       done
     ```
  3. Verify that all the count values are **0**.
  4. If there are messages in the queue, do the following:

     a. Stop the DBUS services, which stops any new incoming data.
     b. Allow Workload XM to finish processing any existing messages.
     c. Stop any remaining components.
- Stopped the Workload XM service and its services.

Procedure

1. Verify that the Workload XM Service and its services are stopped.

2. Do the following:
   a) In a terminal, SSH into the Cloudera Manager host server
   b) Copy the downloaded parcel (.parcel) and its checksum (.sha) file to the `/opt/cloudera/parcel-repo` directory of the Cloudera Manager Server on the Workload XM on-premises cluster.
   c) Copy the downloaded CSD (.jar) file to the `opt/cloudera/csd` directory of the Cloudera Manager Server on the Workload XM on-premises cluster.
3. In the `/opt/cloudera/parcel-repo` directory, set the ownership of the parcel and sha files to **cloudera-scm:cloudera-scm**.
4. In the `/opt/cloudera/csd` directory, set the ownership of the CSD .jar file to **cloudera-scm:cloudera-scm**.
5. In a supported web browser on the Workload XM on-premises cluster, log in to Cloudera Manager.
6. In Cloudera Manager, select **Clusters**, **Parcels**, and then **Distribute**.

   In the list of parcels, the latest version appears with a gray Distribute label.
7. Activate the Workload XM installation files by selecting **Activate Only**. Do not restart.
8. From the Cloudera Manager Host, restart the Cloudera Manager Server by entering the following:

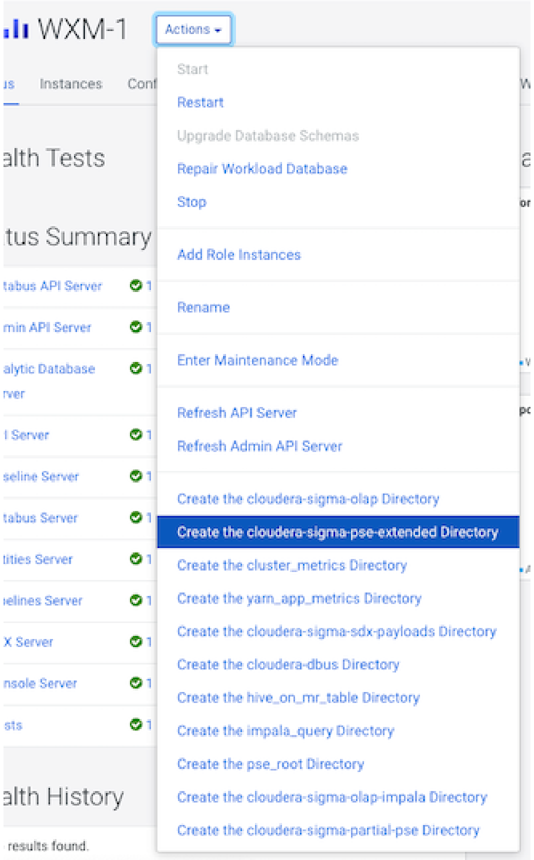
```
service cloudera-scm-server restart
```

9. Navigate to the Workload XM service in the Cloudera Manager UI and then from the **Actions** menu do the following:

   Note: Verify if a Cloudera Sigma PSE extended directory, a Hive MapReduce table directory, or a Shared Data Experience (SDX) role have been created. If yes, you can skip the action for one or all of the aforementioned components. These components are required by Workload XM and its services, redoing the action does not harm the upgrade but skipping the action when a component has not been created could cause issues.

   a) Run the command to create the Cloudera Sigma PSE extended directory by selecting **Create the cloudera-sigma-pse-extended Directory**.
   b) Run the command to create a Hive MapReduce table directory by selecting **Create the hive_on_mr_table Directory**.
   c) Add a Shared Data Experience (SDX) role by selecting **Add Role Instances**. In the **Assign Roles** panel, click **Select hosts** in the SDX server field and then in the Host Selected dialog box, select the check box next to the hostname you require for the Workload XM SDX Server role and click **OK**.
   d) Run the command to upgrade the database schemas by selecting **Upgrade Database Schemas**.
   e) Restart the Workload XM service, by selecting **Restart**.

   As shown in the following image:


WXM-1
Actions
us
Instances
Conf
alth Tests
tus Summary
tabus API Server
min API Server
alytic Database rver
I Server
seline Server
tabus Server
tities Server
pelines Server
X Server
nsole Server
sts
alth History
results found.
Start
Restart
Upgrade Database Schemas
Repair Workload Database
Stop
Add Role Instances
Rename
Enter Maintenance Mode
Refresh API Server
Refresh Admin API Server
Create the cloudera-sigma-olap Directory
Create the cloudera-sigma-pse-extended Directory
Create the cluster_metrics Directory
Create the yarn_app_metrics Directory
Create the cloudera-sigma-sdx-payloads Directory
Create the cloudera-dbus Directory
Create the hive_on_mr_table Directory
Create the impala_query Directory
Create the pse_root Directory
Create the cloudera-sigma-olap-impala Directory
Create the cloudera-sigma-partial-pse Directory

10. Once all the roles have started, from the **Actions** menu, select **Repair Workload Database**.