



Configuring Hadoop Cluster Profile Templates

This chapter contains the following sections:

- [Hadoop Cluster Profile Templates, page 1](#)
- [Creating a Hadoop Cluster Profile Template, page 2](#)
- [Cloning a Hadoop Cluster Profile Template, page 16](#)

Hadoop Cluster Profile Templates

The Hadoop cluster profile template specifies the number of nodes in the cluster and takes care of provisioning and configuring the Hadoop cluster services. Numerous Apache Software Foundation projects develop and comprise the services required to deploy, integrate, and work with Hadoop. Some Hadoop distributions support only a subset of these services, or may have their own distribution-specific services.

Each of the following has been developed to deliver an explicit function:

- HDFS
- CLDB
- YARN
- ZooKeeper
- HBase
- Hive
- Oozie
- Hue
- Spark
- Key-Value Store Indexer
- Solr
- Sqoop
- Impala

- Flume
- PIG
- MAHOUT
- Falcon
- Tez
- Storm
- Ganglia

**Note**

You are not allowed to deselect some of the services because to create a Hadoop cluster, it is necessary to have these services enabled.

Creating a Hadoop Cluster Profile Template

- Step 1** On the menu bar, choose **Solutions > Big Data Containers**.
- Step 2** Click the **Hadoop Cluster Profile Templates** tab.
- Step 3** Click **Add (+)**.
- Step 4** On the **Hadoop Cluster Profile Template** page of the **Create Hadoop Cluster Profile Template** wizard, complete the following fields:

Name	Description
Template Name field	A unique name for the template.
Template Description field	A short description for the template.
Enter Node Count field	The number of nodes in the cluster. The default is four nodes.
Distribution drop-down list	Choose the type of Hadoop distribution.
Hadoop Distribution Version	Choose the Hadoop Distribution Version.

- Step 5** Click **Next**.

What to Do Next

Create a Services Selection policy.

Creating a Services Selection Policy

The cluster policy contains the Hadoop cluster services that you want to enable in the Hadoop cluster.



Note The **Service Selection Page** displays the Hadoop cluster services depending on the Hadoop distribution that you select before on the **Hadoop Cluster Profile Template** page.

Step 1

On the **Services Selection Policy** page of the **Create Hadoop Cluster Profile Template** wizard, check the check box for the optional Hadoop cluster services that you want to enable in your cluster. Some Hadoop cluster services are required for the distribution and cannot be disabled. The available Hadoop cluster services include the following:

- **HDFS**—A file system that spans all nodes in a Hadoop cluster for data storage. This service replicates data across multiple nodes to avoid data loss.
- **CLDB**—A MapR specific management service that keeps track of container locations and the root of volumes.
- **YARN**— A resource-management platform responsible for managing compute resources in clusters and using them for scheduling users' applications.
- **ZooKeeper**—An infrastructure for cross-node synchronization that can be used by applications to ensure that tasks across the cluster are serialized or synchronized.
- **HBase**—A high speed read and write column-oriented database.
- **Hive**—The query engine framework for Hadoop that facilitates easy data summarization, ad-hoc queries, and the analysis of large datasets stored in HDFS and HBase. With SQL-like semantics, Hive makes it easy for RDBMS users to transition into querying unstructured data in Hadoop.
- **Oozie**—A workflow environment for coordinating complex data processing operations.
- **Hue**—An interface that aggregates the most common Hadoop components to improve user experience and to enable users to avoid the underlying complexity and the command line interface.
- **Spark**—An open-source data analytics cluster computing framework.
- **Key-Value Store Indexer**—A method for indexing data across the cluster.
- **SOLR**—A method for searching data across the cluster.
- **Sqoop**—A client-server tool that transfers bulk data between Hadoop and structured datastores, such as relational databases.
- **Impala**—A massively parallel processing (MPP) SQL query engine that runs natively in Apache Hadoop.
- **Flume**—A distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of streaming data into the Hadoop Distributed File System (HDFS).
- **PIG**—A platform for analyzing large data sets that consists of a high-level language for expressing data analysis programs. It allows developers to write complex MapReduce transformations using a simple scripting language.
- **MAHOUT**—A library of scalable machine learning algorithms that provide the data science tools for clustering, classification, and collaborative filtering to automatically find meaningful patterns.

- **Falcon**—Simplifies the development and management of data processing and provides out-of-the-box data management services.
- **Tez**—An extensible framework for building high performance batch and interactive data processing applications, coordinated by YARN in Apache Hadoop.
- **Storm**—A distributed real-time computation system for processing large volumes of high-velocity data..
- **Ganglia**—A scalable distributed system monitoring tool for high-performance computing systems such as clusters and grids.

Step 2 Click **Next**.

What to Do Next

Configure the Rack Assignment policy.

Configuring the Rack Assignment Policy

Step 1 On the **Rack Assignment Policy** page of the **Create Hadoop Cluster Profile Template** wizard, do one of the following:

- To create one or more custom Hadoop node configure policies, click **Add (+)** and continue with Step 2.
- To modify the default Hadoop node configure policy, choose the policy in the table and click **Edit**. For information about the fields in the **Edit Hadoop Node Configure Policy Entry** dialog box, see Step 2.

Step 2 In the **Add Entry to Hadoop Node Configure Policy** dialog box, do the following:

- In the **Rack Name** field, enter the name of the rack server.
- In the **DataNodes** field, click **Select** and check the checkbox for each node that you want to configure on that server.

Note Some Hadoop cluster services require a minimum number of nodes. For example, Zookeeper requires a minimum of 3 nodes.
- Click **Submit**.

Step 3 Click **Next**.

What to Do Next

Configure the HDFS policy.

Configuring the HDFS Policy

Step 1 On the **HDFS Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the HDFS policy configuration and click **Edit**.

If you do not see a node you need for HDFS on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2

In the **Edit HDFS Policy Entry** dialog box, review and, if required, change the following fields:

Name	Description
DataNode drop-down list	Choose Yes if you want the node to act as the DataNode for HDFS. Otherwise, choose No . The data nodes store and retrieve data on request by the name node or by the client.
Primary NameNode drop-down list	Choose Yes if you want the node to act as the primary name node for HDFS. Otherwise, choose No . All operations of the HDFS cluster are maintained by the primary name node. There can be only one primary name node for the HDFS Cluster.
Secondary NameNode drop-down list	Choose Yes if you want the node to act as a secondary name node for HDFS. Otherwise, choose No . The secondary name node is not a direct replacement for the primary name node. The main role of a secondary name node is to periodically merge the FSImage and edit log, to prevent the edit log from becoming too large. A secondary name node runs on a separate physical system because it requires a lot of memory to merge two files. It keeps a copy of the merged file in its local file system so that it is available for use if the primary name node fails.
Balancer drop-down list	Choose Yes if you want the node to act as a balancer for HDFS. Otherwise, choose No .
HTTPFS drop-down list	Choose Yes if you want the node to act as HTTPFS for HDFS. Otherwise, choose No . Note This service provides HTTP access to HDFS.
Fail Over Controller drop-down list	Choose Yes if you want the node to act as Fail Over Controller for HDFS. Otherwise, choose No .
Gateway drop-down list	Choose Yes if you want the node to act as Gateway for HDFS. Otherwise, choose No .
Journal Node drop-down list	Choose Yes if you want the node to act as Journal node for HDFS. Otherwise, choose No .

- Step 3** Click **Submit**.
- Step 4** Repeat Steps 1 and 2 to configure the other nodes for HDFS.
- Step 5** Click **Next**.

What to Do Next

Configure the CLDB policy.

Configuring the CLDB Policy

- Step 1** On the **CLDB Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the CLDB policy configuration and click **Edit**.
If you do not see a node you need for CLDB on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.
- Step 2** In the **Edit CLDB Policy Entry** dialog box, choose **Yes** if you want the node to act as a CLDB agent.
- Step 3** Click **Submit**.
- Step 4** Repeat Steps 1 and 2 to configure the other nodes for CLDB.
- Step 5** Click **Next**.

What to Do Next

Configure the YARN policy.

Configuring the YARN Policy

- Step 1** On the **YARN Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the YARN policy configuration and click **Edit**.
If you do not see a node you need for the YARN policy on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.
- Step 2** In the **Edit YARN Policy Entry** dialog box, review and, if required, change the following fields:

Name	Description
Resource Manager drop-down list	Choose Yes if you want the node to act as a Resource Manager. Otherwise, choose No . The Resource Manager is the ultimate authority that allocates resources among all the applications in the system.

Name	Description
Node Manager drop-down list	Choose Yes if you want the node to act as a task Node Manager. Otherwise, choose No . The Node Manager is responsible for launching the applications' containers, monitoring their resource usage (CPU, memory, disk, network), and reporting to the Resource Manager.
Gateway drop-down list	Choose Yes if you want the node to act as a Gateway. Otherwise, choose No .
JobHistory drop-down list	Choose Yes if you want the node to preserve the JobHistory. Otherwise, choose No .

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for Yarn.

Step 5 Click **Next**.

What to Do Next

Configure the ZooKeeper policy.

Configuring the ZooKeeper Policy



Note You must configure a minimum of three nodes for ZooKeeper.

Step 1 On the **ZooKeeper Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the ZooKeeper policy configuration and click **Edit**. If you do not see a node you need for ZooKeeper on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit ZooKeeper Policy Entry** dialog box, choose **Yes** to make the node to act as a ZooKeeper.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for ZooKeeper.

Step 5 Click **Next**.

What to Do Next

Configure the HBase policy.

Configuring the HBase Policy

Step 1 On the **HBase Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the HBase policy configuration and click **Edit**.
If you do not see a node you need for HBase on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit HBase Policy Entry** dialog box, review and, if required, change the following fields:

Name	Description
HBase Master drop-down list	Choose Yes if you want the node to act as the HBase master. Otherwise, choose No .
Region Server drop-down list	Choose Yes if you want the node to act as a region server. Otherwise, choose No .
HBase Thrift Server drop-down list	Choose Yes if you want the node to host HBase Thrift. Otherwise, choose No .

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for HBase.

Step 5 Click **Next**.

What to Do Next

Configure the Hive policy.

Configuring the Hive Policy

Step 1 On the **Hive Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Hive policy configuration and click **Edit**.
If you do not see a node you need for Hive on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit Hive Policy Entry** dialog box, review and, if required, change the following fields:

Name	Description
HiveServer2 drop-down list	Choose Yes if you want the node to host HiveServer2. Otherwise, choose No .

Name	Description
Hive Metastore Server drop-down list	Choose Yes if you want the node to act as a Hive metastore. Otherwise, choose No .
WebHCat drop-down list	Choose Yes if you want the node to act as a WebHCat. Otherwise, choose No . WebHCat is the REST API for HCatalog, a table and storage management layer for Hadoop.
Gateway drop-down list	Choose Yes if you want the node to act as a Gateway for Hive. Otherwise, choose No .

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for Hive.

Step 5 Click **Next**.

What to Do Next

Configure the Oozie policy.

Configuring the Oozie Policy

Step 1 On the **Oozie Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Oozie policy configuration and click **Edit**.
If you do not see a node you need for Oozie on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit Oozie Policy Entry** dialog box, click **Yes** to make the node to act as an Oozie server.

Step 3 Repeat Steps 1 and 2 to configure the other nodes for Oozie.

Step 4 Click **Next**.

What to Do Next

Configure the Hue policy.

Configuring the Hue Policy

-
- Step 1** On the **Hue Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Hue policy configuration and click **Edit**.
If you do not see a node you need for Hue on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.
- Step 2** In the **Edit Hue Policy Entry** dialog box, do the following:
- From the **Hue Server** drop-down list, choose **Yes** if you want the node to act as a Hue server.
 - From the **BeesWax Server** drop down-list, choose **Yes** if you want the node to act as a BeesWax server.
 - From the **Kt Renewer** drop down-list, choose **Yes** if you want the node to act as a Kt Renewer.
 - Click **Submit**.
- Step 3** Repeat Steps 1 and 2 to configure the other nodes for Hue.
- Step 4** Click **Next**.
-

What to Do Next

Configure the Spark policy.

Configuring the Spark Policy

-
- Step 1** On the **Spark Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Spark policy configuration and click **Edit**.
If you do not see a node you need for Spark on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.
- Step 2** In the **Edit Spark Policy Entry** dialog box, review and, if required, change the following fields:
- | Name | Description |
|-----------------------------|--|
| Spark Master drop-down list | Choose Yes if you want the node to act as a Spark master. Otherwise, choose No . |
| Spark Worker drop-down list | Choose Yes if you want the node to act as a Spark worker. Otherwise, choose No . |
- Step 3** Click **Submit**.
- Step 4** Repeat Steps 1 and 2 to configure the other nodes for Spark.
- Step 5** Click **Next**.
-

What to Do Next

Configure the KSIIndexer policy.

Configuring the Key-Value Store Indexer Policy

-
- Step 1** On the **Key-Value Store Indexer Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Key-Value Store Indexer policy configuration and click **Edit**.
If you do not see a node you need for KSIIndexer on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.
- Step 2** In the **Edit KSIIndexer Policy Entry** dialog box, choose **Yes** if you want the node to act as a KSIIndexer server.
- Step 3** Click **Submit**.
- Step 4** Repeat Steps 1 and 2 to configure the other nodes for KSIIndexer.
- Step 5** Click **Next**.
-

What to Do Next

Configure the Solr policy.

Configuring the Solr Policy

-
- Step 1** On the **Solr Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Solr policy configuration and click **Edit**.
If you do not see a node you need for Solr on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.
- Step 2** In the **Edit Solr Policy Entry** dialog box, choose **Yes** if you want the node to act as a Solr server.
- Step 3** Click **Submit**.
- Step 4** Repeat Steps 1 and 2 to configure the other nodes for Solr.
- Step 5** Click **Next**.
-

What to Do Next

Configure the Sqoop policy.

Configuring the Sqoop Policy

-
- Step 1** On the **Sqoop Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Sqoop policy configuration and click **Edit**.
If you do not see a node you need for Sqoop on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.
- Step 2** In the **Edit Sqoop Policy Entry** dialog box, choose **Yes** if you want the node to act as a Sqoop server.
- Step 3** Click **Submit**.
- Step 4** Repeat Steps 1 and 2 to configure the other nodes for Sqoop.
- Step 5** Click **Next**.
-

What to Do Next

Configure the Impala policy.

Configuring the Impala Policy

-
- Step 1** On the **Impala Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Impala policy configuration and click **Edit**.
If you do not see a node you need for Impala on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.
- Step 2** In the **Edit Impala Policy Entry** dialog box, do the following:
- from the **Impala StateStore** drop-down list, choose **Yes** if you want the node to act as an Impala Statestore.
 - From the **Impala Catalog Server** drop-down list, choose **Yes** if you want the node to act as an Impala catalog server.
The other fields in this dialog box are for your information only.
 - Click **Submit**.
- Step 3** Repeat Steps 1 and 2 to configure the other nodes for Impala.
- Step 4** Click **Submit**.
-

What to Do Next

Configure the Flume policy.

Configuring the Flume Policy

- Step 1** On the **Flume Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Flume policy configuration and click **Edit**.
If you do not see a node you need for Flume on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.
- Step 2** In the **Edit Flume Policy Entry** dialog box, choose **Yes** if you want the node to act as a Flume agent.
- Step 3** Click **Submit**.
- Step 4** Repeat Steps 1 and 2 to configure the other nodes for Flume.
- Step 5** Click **Next**.
-

What to Do Next

Configure the PIG Policy.

Configuring the PIG Policy

- Step 1** On the **Pig Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Pig policy configuration and click **Edit**.
If you do not see a node you need for Pig on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.
- Step 2** In the **Edit Pig Policy Entry** dialog box, choose **Yes** if you want the node to act as a Pig agent.
- Step 3** Click **Submit**.
- Step 4** Repeat Steps 1 and 2 to configure the other nodes for Pig.
- Step 5** Click **Next**.
-

What to Do Next

Configure the MAHOUT Policy.

Configuring the MAHOUT Policy

- Step 1** On the **MAHOUT Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the MAHOUT policy configuration and click **Edit**.
If you do not see a node you need for MAHOUT on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

- Step 2** In the **Edit MAHOUT Policy Entry** dialog box, choose **Yes** if you want the node to act as a MAHOUT agent.
- Step 3** Click **Submit**.
- Step 4** Repeat Steps 1 and 2 to configure the other nodes for MAHOUT.
- Step 5** Click **Submit**.
-

What to Do Next

Configure a Falcon Policy.

Configuring the Falcon Policy

- Step 1** On the **Falcon Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Falcon policy configuration and click **Edit**.
If you do not see a node you need for Falcon on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.
- Step 2** In the **Edit Falcon Policy Entry** dialog box, choose **Yes** if you want the node to act as a Falcon server and as the Falcon client from the **Falcon Server** and **Falcon Client** drop-down lists.
- Step 3** Click **Submit**.
- Step 4** Repeat Steps 1 and 2 to configure the other nodes for Falcon.
- Step 5** Click **Submit**.
-

What to Do Next

Configure the Tez Policy.

Configuring the Tez Policy

- Step 1** On the **Tez Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Tez policy configuration and click **Edit**.
If you do not see a node you need for Tez on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.
- Step 2** In the **Edit Tez Policy Entry** dialog box, choose **Yes** if you want the node to act as a Tez agent.
- Step 3** Click **Submit**.
- Step 4** Repeat Steps 1 and 2 to configure the other nodes for Tez.
- Step 5** Click **Submit**.
-

What to Do Next

Configure the Storm Policy.

Configuring the Storm Policy

- Step 1** On the **Storm Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Storm policy configuration and click **Edit**.
If you do not see a node you need for Storm on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.
- Step 2** In the **Edit Storm Policy Entry** dialog box, do the following:
- In the **DRPC Server** drop-down list, choose **Yes** if you want the node to act as a DRPC server.
 - In the **Nimbus** drop-down list, choose **Yes** if you want the node to act as a Nimbus server.
 - In the **Storm REST API Server** drop-down list, choose **Yes** if you want the node to act as a Storm REST API server.
 - In the **Storm UI Server** drop-down list, choose **Yes** if you want the node to act as a Storm UI server.
 - In the **Supervisor** drop-down list, choose **Yes** if you want the node to act as a supervisor.
- Step 3** Click **Submit**.
- Step 4** Repeat Steps 1 and 2 to configure the other nodes for Storm.
- Step 5** Click **Submit**.
-

What to Do Next

Configure the Ganglia Policy.

Configuring the Ganglia Policy

- Step 1** On the **Ganglia Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Ganglia policy configuration and click **Edit**.
If you do not see a node you need for Ganglia on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.
- Step 2** In the **Edit Ganglia Policy Entry** dialog box, choose **Yes** if you want the node to act as a Ganglia server and as Ganglia monitor from the **Ganglia Server** and **Ganglia Monitor** drop-down lists.
- Step 3** Click **Submit**.
- Step 4** Repeat Steps 1 and 2 to configure the other nodes for Ganglia.
- Step 5** Click **Submit**.
-

Cloning a Hadoop Cluster Profile Template

- Step 1** On the menu bar, choose **Solutions > Big Data Containers**.
- Step 2** Click the **Hadoop Cluster Profile Template** tab.
- Step 3** Click the row for the template that you want to clone.
- Step 4** Click **Clone** .
- Step 5** In the **Clone Hadoop Cluster Profile Template** dialog box, do the following:
- Enter a unique name and description for the new Hadoop cluster profile template.
 - Click **Next**, review the information on each page, and modify, if required.
 - Click **Submit**.
-