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 15

Hadoop Cluster Profile Templates

The Hadoop cluster profile template specifies the number of nodes in the cluster. The template also takes care of provisioning and configuring the Hadoop cluster services. Apache Software Foundation projects around the world develop services for Hadoop deployment and integration. Some Hadoop distributions support only a subset of these services, or have their own distribution-specific services.

Each of the following supplies a dedicated function:

> **Note**

You cannot uncheck some of the services because they are necessary to create a Hadoop cluster. All mandatory services are checked by default.

<table>
<thead>
<tr>
<th>Hadoop Services</th>
<th>Cloudera</th>
<th>MapR</th>
<th>Hortonworks</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDFS</td>
<td>Yes</td>
<td>—</td>
<td>Yes</td>
</tr>
<tr>
<td>CLDB</td>
<td>—</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>YARN/MapReduce</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ZooKeeper</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>HBase</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Hive</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Oozie</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Creating a Hadoop Cluster Profile Template

Before You Begin
Create a Hadoop cluster Configuration 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:

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template Name field</td>
<td>A unique name for the template.</td>
</tr>
<tr>
<td>Template Description field</td>
<td>A short description for the template.</td>
</tr>
</tbody>
</table>
### Configuring Hadoop Cluster Profile Templates

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Node Count</strong> field</td>
<td>The number of nodes in the cluster. The default is four nodes.</td>
</tr>
<tr>
<td><strong>Hadoop Distribution</strong> drop-down list</td>
<td>The type of Hadoop distribution. The Hadoop cluster services are displayed based on the selected Hadoop distribution.</td>
</tr>
<tr>
<td><strong>Hadoop Distribution Version</strong></td>
<td>Choose the Hadoop distribution version.</td>
</tr>
<tr>
<td><strong>Hadoop Cluster Configuration Parameters Template</strong> drop-down list</td>
<td>Choose the cluster configuration parameters template.</td>
</tr>
<tr>
<td><strong>Storage Pool Disk Grouping Configuration</strong></td>
<td></td>
</tr>
<tr>
<td><strong>No.of disks per pool for the servers with disks 1-12</strong> field</td>
<td>The number of disks per pool for the servers with disks ranging from 1 to 12. This field is displayed only for MapR cluster. You need to enter the value before creating a new node in the MapR cluster.</td>
</tr>
<tr>
<td><strong>No.of disks per pool for the servers with disks 13-24</strong> field</td>
<td>The number of disks per pool for the servers with disks ranging from 13 to 24. This field is displayed only for MapR cluster. You need to enter the value before creating a new node in the MapR cluster.</td>
</tr>
<tr>
<td><strong>No.of disks per pool for the servers with disks &gt;24</strong> field</td>
<td>The number of disks per pool for the servers with disks ranging from 24 to 96. This field is displayed only for MapR cluster. You need to enter the value before creating a new node in the MapR cluster.</td>
</tr>
</tbody>
</table>

### 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.
The Service Selection Page displays the Hadoop cluster services, depending on the Hadoop distribution already selected on the Hadoop Cluster Profile Template page.

**Step 1**  
On the Services Selection Policy page of the Create Hadoop Cluster Profile Template wizard, do the following:

**Step 2**  
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.
- **YARN**—A resource-management platform responsible for managing compute resources in clusters and using them for scheduling your applications.
- **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 data sets 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.
- **ZooKeeper**—An infrastructure for cross-node synchronization. The applications ensure that tasks across the cluster are serialized or synchronized.
- **Hue**—An interface that aggregates the most common Hadoop components to improve user experience. This allows you to avoid the underlying complexity of the system, and bypasses the command-line interface.
- **Spark**—An open-source data analytics engine.
- **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 data stores, 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).

**Step 3**  
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, you can:

- Create one or more Hadoop node configuration policies. Click **Add (+)**, and continue with Step 2.
- Modify the default node configuration policy. Choose the default policy in the table. Click **Edit**, and continue with Step 2.

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

a) In the Rack Name field, enter the name of the rack server.

b) In the DataNodes field, click Select and check the check box 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 three nodes.

c) 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, do the following:

**Step 2**
Click the row in the table with the node for which you want to change the HDFS policy configuration, and click **Edit**.

**Step 3**
In the Edit HDFS Policy Entry dialog box, review and, if necessary, change the following fields:

a) Choose **Yes** from the DataNode drop-down list if you want the node to act as the DataNode for HDFS. The data nodes store and retrieve data on request by the name node or by the client.

b) Choose **Yes** from the Primary NameNode drop down-list if you want the node to act as the primary name node for HDFS. The primary name node maintains all the operations of the HDFS cluster. There can be only one primary name node for the HDFS cluster.

c) Choose **Yes** from the Secondary NameNode drop down-list if you want the node to act as a secondary name node for HDFS. 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 more 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.

d) Choose **Yes** from the Balancer drop down-list if you want the node to act as a balancer for HDFS.

e) Choose **Yes** from the HTTPFS drop down-list if you want the node to act as HTTPFS for HDFS. This service provides HTTP access to HDFS.
f) Choose Yes from the Fail Over Controller drop down-list if you want the node to act as Fail Over Controller for HDFS.

g) Choose Yes from the Gateway drop down-list if you want the node to act as Gateway for HDFS.

Step 4  Click Submit.
Step 5  Repeat Steps 1 and 2 to configure the other nodes for HDFS.
Step 6  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, do the following:
Step 2  Click the row in the table with the node for which you want to change the CLDB policy configuration, and click Edit.
Step 3  In the Edit CLDB Policy Entry dialog box, choose Yes if you want the node to act as a CLDB agent.
Step 4  Click Submit.
Step 5  Repeat Steps 1 and 2 to configure the other nodes for CLDB.
Step 6  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, do the following:
Step 2  Click the row in the table with the node for which you want to change the YARN policy configuration, and click Edit.
Step 3  In the Edit YARN Policy Entry dialog box, review and, if necessary, change the following fields:
    a) Choose Yes from the Resource Manager drop-down list if you want the node to act as a Resource Manager. The Resource Manager is the ultimate authority that allocates resources among all the applications in the system.
    b) Choose Yes from the Node Manager drop down-list if you want the node to act as a task Node Manager. The Node Manager is responsible for launching the applications' containers, monitoring their resource usage (CPU, memory, disk, network), and reporting to the Resource Manager.
    c) Choose Yes from the Gateway drop down-list if you want the node to act as a Gateway.
    d) Choose Yes from the JobHistory drop down-list if you want the node to preserve the Job History.
Configuring Hadoop Cluster Profile Templates

Configuring the ZooKeeper Policy

What to Do Next
Configure the ZooKeeper policy.

Configuring the HBase Policy

What to Do Next
Configure the HBase policy.
Configuring the Hive Policy

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

**Step 2**
Click the row in the table with the node for which you want to change the Hive policy configuration, and click Edit.

**Step 3**
In the Edit Hive Policy Entry dialog box, review and, if necessary, change the following fields:

- **a)** Choose Yes from the HiveServer2 drop-down list if you want the node to act as a HiveServer2.
- **b)** Choose Yes from the Hive Metastore Server drop-down list if you want the node to act as a Hive metastore.
- **c)** Choose Yes from the WebHCat drop-down list if you want the node to act as a WebHCat. WebHCat is the REST API for HCatalog, a table and storage management layer for Hadoop.
- **d)** Choose Yes from the Gateway drop-down list if you want the node to act as a Gateway for Hive.

**Step 4**
Click Submit.

**Step 5**
Repeat Steps 1 and 2 to configure the other nodes for HBase.

**Step 6**
Click Next.

**What to Do Next**
Configure the Hive policy.

Configuring the Oozie Policy

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

**Step 2**
Click the row in the table with the node for which you want to change the Oozie policy configuration, and click Edit.

**Step 3**
Click Yes in the Edit Oozie Policy Entry dialog box to make the node to act as an Oozie server.

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

**Step 5**
Click Next.

**What to Do Next**
Configure the Oozie policy.
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, do the following:

**Step 2**
Click the row in the table with the node for which you want to change the Hue policy configuration, and click **Edit**.

**Step 3**
In the **Edit Hue Policy Entry** dialog box, do the following:

a) Choose **Yes** from the **Hue Server** drop-down list if you want the node to act as a Hue server.
b) Choose **Yes** from the **BeesWax Server** drop-down list if you want the node to act as a BeesWax server.
c) Choose **Yes** from the **Kt Renewer** drop-down list if you want the node to act as a Kt Renewer.
d) Click **Submit**.

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

**Step 5**
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, do the following:

**Step 2**
Click the row in the table with the node for which you want to change the Spark policy configuration, and click **Edit**.

**Step 3**
In the **Edit Spark Policy Entry** dialog box, review and, if necessary, change the following fields:

a) Choose **Yes** from the **Spark Master** drop-down list if you want the node to act as a Spark master.
b) Choose **Yes** from the **Spark Worker** drop-down list if you want the node to act as a Spark worker.

c) 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 Key-Value Store Indexer 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, do the following:
Step 2: 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.
Step 3: Choose Yes in the Edit KSIndexer Policy Entry dialog box if you want the node to act as a KSIndexer server.
Step 4: Click Submit.
Step 5: Repeat Steps 1 and 2 to configure the other nodes for KSIndexer.
Step 6: 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, do the following:
Step 2: Click the row in the table with the node for which you want to change the Solr policy configuration, and click Edit.
Step 3: Choose Yes in the Edit Solr Policy Entry dialog box if you want the node to act as a Solr server.
Step 4: Click Submit.
Step 5: Repeat Steps 1 and 2 to configure the other nodes for Solr.
Step 6: 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, do the following:
Step 2 Click the row in the table with the node for which you want to change the Sqoop policy configuration, and click Edit.
Step 3 Choose Yes in the Edit Sqoop Policy Entry dialog box if you want the node to act as a Sqoop server.
Step 4 Click Submit.
Step 5 Repeat Steps 1 and 2 to configure the other nodes for Sqoop.
Step 6 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, do the following:
Step 2 Click the row in the table with the node for which you want to change the Impala policy configuration, and click Edit.
Step 3 In the Edit Impala Policy Entry dialog box, do the following:
   a) Choose Yes from the Impala Daemon drop-down list if you want the node to act as an Impala daemon.
   b) Choose Yes from the Impala StateStore drop-down list if you want the node to act as an Impala Statestore.
   c) Choose Yes from the Impala Catalog Server drop-down list if you want the node to act as an Impala catalog server.
      The other fields in this dialog box are for your information only.
   d) Click Submit.
Step 4 Repeat Steps 1 and 2 to configure the other nodes for Impala.
Step 5 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 do the following:
Step 2 Click the row in the table with the node for which you want to change the Flume policy configuration, and click Edit.
Step 3 Choose Yes in the Edit Flume Policy Entry dialog box if you want the node to act as a Flume agent.
Step 4 Click Submit.
Step 5 Repeat Steps 1 and 2 to configure the other nodes for Flume.
Step 6 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, do the following:
Step 2 Click the row in the table with the node for which you want to change the Pig policy configuration, and click Edit.
Step 3 Choose Yes in the Edit Pig Policy Entry dialog box if you want the node to act as a Pig agent.
Step 4 Click Submit.
Step 5 Repeat Steps 1 and 2 to configure the other nodes for Pig.
Step 6 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, do the following:

**Step 2** Click the row in the table with the node for which you want to change the MAHOUT policy configuration, and click **Edit**.

**Step 3** Choose **Yes** in the Edit MAHOUT Policy Entry dialog box if you want the node to act as a MAHOUT agent.

**Step 4** Click **Submit**.

**Step 5** Repeat Steps 1 and 2 to configure the other nodes for MAHOUT.

**Step 6** 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, do the following:

**Step 2** Click the row in the table with the node for which you want to change the Falcon policy configuration, and click **Edit**.

**Step 3** Choose **Yes** in the Edit Falcon Policy Entry dialog box. You can make the node act as a Falcon server from the **Falcon Server** and the Falcon client from **Falcon Client** drop-down lists.

**Step 4** Click **Submit**.

**Step 5** Repeat Steps 1 and 2 to configure the other nodes for Falcon.

**Step 6** 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, do the following:
Step 2  Click the row in the table with the node for which you want to change the Tez policy configuration, and click Edit.
Step 3  Choose Yes in the Edit Tez Policy Entry dialog box if you want the node to act as a Tez agent.
Step 4  Click Submit.
Step 5  Repeat Steps 1 and 2 to configure the other nodes for Tez.
Step 6  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, do the following:
Step 2  Click the row in the table with the node for which you want to change the Storm policy configuration, and click Edit.
Step 3  In the Edit Storm Policy Entry dialog box, do the following:
   a) Choose Yes in the DRPC Server drop-down list if you want the node to act as a DRPC server.
   b) Choose Yes in the Nimbus drop-down list if you want the node to act as a Nimbus server.
   c) Choose Yes in the Storm REST API Server drop-down list if you want the node to act as a Storm REST API server.
   d) Choose Yes in the Storm UI Server drop-down list if you want the node to act as a Storm UI server.
   e) Choose Yes in the Supervisor drop-down list if you want the node to act as a supervisor.
Step 4  Click Submit.
Step 5  Repeat Steps 1 and 2 to configure the other nodes for Storm.
Step 6  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, do the following:
Step 2  Click the row in the table with the node for which you want to change the Ganglia policy configuration, and click Edit.
Step 3  Choose Yes in the Edit Ganglia Policy Entry dialog box. You can make the node to act as a Ganglia server from the Ganglia Server and the Ganglia monitor from the Ganglia Monitor drop-down lists.
Step 4  Click Submit.
Step 5  Repeat Steps 1 and 2 to configure the other nodes for Ganglia.
Step 6  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 Templates 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:
a) Enter a unique name and description for the new Hadoop cluster profile template.
b) Click Next, review the information on each page, and modify, if necessary.
c) Click Submit.