

Hortonworks Data Platform

Apache Ambari Reference

(November 28, 2016)

Hortonworks Data Platform: Apache Ambari Reference

Copyright © 2012-2016 Hortonworks, Inc. All rights reserved.

The Hortonworks Data Platform, powered by Apache Hadoop, is a massively scalable and 100% open source platform for storing, processing and analyzing large volumes of data. It is designed to deal with data from many sources and formats in a very quick, easy and cost-effective manner. The Hortonworks Data Platform consists of the essential set of Apache Hadoop projects including MapReduce, Hadoop Distributed File System (HDFS), HCatalog, Pig, Hive, HBase, ZooKeeper and Ambari. Hortonworks is the major contributor of code and patches to many of these projects. These projects have been integrated and tested as part of the Hortonworks Data Platform release process and installation and configuration tools have also been included.

Unlike other providers of platforms built using Apache Hadoop, Hortonworks contributes 100% of our code back to the Apache Software Foundation. The Hortonworks Data Platform is Apache-licensed and completely open source. We sell only expert technical support, [training](#) and partner-enablement services. All of our technology is, and will remain free and open source. Please visit the [Hortonworks Data Platform](#) page for more information on Hortonworks technology. For more information on Hortonworks services, please visit either the [Support](#) or [Training](#) page. Feel free to [Contact Us](#) directly to discuss your specific needs.

Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at

<http://www.apache.org/licenses/LICENSE-2.0>

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

Table of Contents

1. Installing Ambari Agents Manually	1
1.1. Download the Ambari Repo	1
1.2. Install the Ambari Agents Manually	6
2. Customizing HDP Services	8
2.1. Defining Service Users and Groups for a HDP 2.x Stack	8
2.2. Setting Properties That Depend on Service Usernames/Groups	9
3. Using Custom Host Names	11
3.1. How to Customize the name of a host	11
4. Changing Host Names	13
5. Moving the Ambari Server	15
5.1. Back up Current Data	15
5.2. Update all Agents	15
5.3. Install the New Ambari Server	16
6. Moving the ZooKeeper Server	18
7. Configuring LZO Compression	20
7.1. Configure core-site.xml for LZO	20
7.2. Running Compression with Hive Queries	20
7.2.1. Create LZO Files	21
7.2.2. Write Custom Java to Create LZO Files	21
8. Using Non-Default Databases	22
8.1. Using Non-Default Databases - Ambari	22
8.1.1. Using Ambari with Oracle	22
8.1.2. Using Ambari with MySQL/MariaDB	23
8.1.3. Using Ambari with PostgreSQL	25
8.1.4. Troubleshooting Non-Default Databases with Ambari	26
8.2. Using Non-Default Databases - Hive	27
8.2.1. Using Hive with Oracle	27
8.2.2. Using Hive with MySQL	28
8.2.3. Using Hive with PostgreSQL	29
8.2.4. Troubleshooting Non-Default Databases with Hive	30
8.3. Using Non-Default Databases - Oozie	31
8.3.1. Using Oozie with Oracle	32
8.3.2. Using Oozie with MySQL	33
8.3.3. Using Oozie with PostgreSQL	34
8.3.4. Troubleshooting Non-Default Databases with Oozie	35
9. Setting up Ambari to use an Internet Proxy Server	36
10. Configuring Network Port Numbers	38
10.1. Default Network Port Numbers - Ambari	38
10.2. Optional: Changing the Default Ambari Server Port	38
10.3. Optional: Changing the Ambari Server-Agent Port	39
11. Change the JDK Version	40
12. Using Ambari Blueprints	42
13. Tuning Ambari Performance	43
14. Customizing Ambari Log + PID Directories	44
14.1. Customizing Ambari Server Log + PID Directories	44
14.2. Customizing Ambari Agent Log + PID Directories	45

List of Tables

8.1. Hive Security Authorization Settings 31

1. Installing Ambari Agents Manually

In cases where you do not have SSH for Ambari to automatically install the Agents or you want to pre-install the Agents, you can perform a manual agent setup. This involves two steps:

1. [Download the Ambari Repo \[1\]](#)
2. [Install the Ambari Agents Manually \[6\]](#)

1.1. Download the Ambari Repo

Select the OS family running on your installation host.

RHEL/CentOS/Oracle Linux 7

On a server host that has Internet access, use a command line editor to perform the following steps:

1. Log in to your host as `root`.
2. Download the Ambari repository file to a directory on your installation host.

```
wget -nv http://public-repo-1.hortonworks.com/ambari/centos7/2.x/updates/2.4.2.0/ambari.repo -O /etc/yum.repos.d/ambari.repo
```



Important

Do not modify the `ambari.repo` file name. This file is expected to be available on the Ambari Server host during Agent registration.

3. Confirm that the repository is configured by checking the repo list.

```
yum repolist
```

You should see values similar to the following for Ambari repositories in the list.

Version values vary, depending on the installation.

repo id	repo name	status
AMBARI.2.4.2.0-2.x	Ambari 2.x	8
base	CentOS-7 - Base	6,518
extras	CentOS-7 - Extras	37
updates	CentOS-7 - Updates	785

4. Proceed to [Install the Ambari Agents manually](#).



Note

Accept the warning about trusting the Hortonworks GPG Key. That key will be automatically downloaded and used to validate packages from Hortonworks. You will see the following message:

```
Importing GPG key 0x07513CAD: Userid: "Jenkins (HDP
Builds) <jenkin@hortonworks.com>" From : http://
s3.amazonaws.com/dev.hortonworks.com/ambari/centos7/RPM-
GPG-KEY/RPM-GPG-KEY-Jenkins
```

RHEL/CentOS/Oracle Linux 6

On a server host that has Internet access, use a command line editor to perform the following steps:

1. Log in to your host as `root`.
2. Download the Ambari repository file to a directory on your installation host.

```
wget -nv http://public-repo-1.hortonworks.com/ambari/
centos6/2.x/updates/2.4.2.0/ambari.repo -O /etc/yum.repos.d/
ambari.repo
```



Important

Do not modify the `ambari.repo` file name. This file is expected to be available on the Ambari Server host during Agent registration.

3. Confirm that the repository is configured by checking the repo list.

```
yum repolist
```

You should see values similar to the following for Ambari repositories in the list.

Version values vary, depending on the installation.

repo id	repo name	status
AMBARI.2.4.2.0-2.x	Ambari 2.x	8
base	CentOS-6 - Base	6,518
extras	CentOS-6 - Extras	37
updates	CentOS-6 - Updates	785

4. Proceed to [Install the Ambari Agents manually](#).



Note

Accept the warning about trusting the Hortonworks GPG Key. That key will be automatically downloaded and used to validate packages from Hortonworks. You will see the following message:

```
Importing GPG key 0x07513CAD: Userid: "Jenkins (HDP
Builds) <jenkin@hortonworks.com>" From : http://
s3.amazonaws.com/dev.hortonworks.com/ambari/centos6/RPM-
GPG-KEY/RPM-GPG-KEY-Jenkins
```

On a server host that has Internet access, use a command line editor to perform the following steps:

1. Log in to your host as `root`.
2. Download the Ambari repository file to a directory on your installation host.

```
wget -nv http://public-repo-1.hortonworks.com/ambari/suse11/2.x/updates/2.4.2.0/ambari.repo -O /etc/zypp/repos.d/ambari.repo
```



Important

Do not modify the `ambari.repo` file name. This file is expected to be available on the Ambari Server host during Agent registration.

3. Confirm the downloaded repository is configured by checking the repo list.

```
zypper repos
```

You should see the Ambari repositories in the list.

Version values vary, depending on the installation.

Alias	Name	Enabled	Refresh
AMBARI.2.4.2.0-2.x	Ambari 2.x	Yes	No
http-demeter.uni-regensburg.de-c997c8f9	SUSE-Linux-Enterprise-Software-Development-Kit-11-SP1 11.1.1-1.57	Yes	Yes
opensuse	OpenSuse	Yes	Yes

4. Proceed to [Install the Ambari Agents manually](#).

SLES 12

On a server host that has Internet access, use a command line editor to perform the following steps:

1. Log in to your host as `root`.
2. Download the Ambari repository file to a directory on your installation host.

```
wget -nv http://public-repo-1.hortonworks.com/ambari/sles12/2.x/updates/2.4.0.1/ambari.repo -O /etc/zypp/repos.d/ambari.repo
```



Important

Do not modify the `ambari.repo` file name. This file is expected to be available on the Ambari Server host during Agent registration.

3. Confirm the downloaded repository is configured by checking the repo list.

```
zypper repos
```

You should see the Ambari repositories in the list.

Version values vary, depending on the installation.

Alias	Name	Enabled	Refresh
AMBARI.2.4.2.0-2.x	Ambari 2.x	Yes	No
http-demeter.uni-regensburg.de-c997c8f9	SUSE-Linux-Enterprise-Software-Development-Kit-12 12.1	Yes	Yes
opensuse	OpenSuse	Yes	Yes

4. Proceed to [Install the Ambari Agents manually](#).

Ubuntu 12

On a server host that has Internet access, use a command line editor to perform the following steps:

1. Log in to your host as `root`.
2. Download the Ambari repository file to a directory on your installation host.

```
wget -nv http://public-repo-1.hortonworks.com/ambari/
ubuntu12/2.x/updates/2.4.2.0/ambari.list -O /etc/apt/
sources.list.d/ambari.list
```

```
apt-key adv --recv-keys --keyserver keyserver.ubuntu.com
B9733A7A07513CAD
```

```
apt-get update
```



Important

Do not modify the `ambari.list` file name. This file is expected to be available on the Ambari Server host during Agent registration.

3. Confirm that Ambari packages downloaded successfully by checking the package name list.

```
apt-cache showpkg ambari-server
```

```
apt-cache showpkg ambari-agent
```

```
apt-cache showpkg ambari-metrics-assembly
```

You should see the Ambari packages in the list.

4. Proceed to [Install the Ambari Agents manually](#).

Ubuntu 14

On a server host that has Internet access, use a command line editor to perform the following steps:

1. Log in to your host as `root`.
2. Download the Ambari repository file to a directory on your installation host.


```
wget -nv http://public-repo-1.hortonworks.com/ambari/
ubuntu14/2.x/updates/2.4.2.0/ambari.list -O /etc/apt/
sources.list.d/ambari.list

apt-key adv --recv-keys --keyserver keyserver.ubuntu.com
B9733A7A07513CAD

apt-get update
```



Important

Do not modify the `ambari.list` file name. This file is expected to be available on the Ambari Server host during Agent registration.

3. Confirm that Ambari packages downloaded successfully by checking the package name list.

```
apt-cache showpkg ambari-server

apt-cache showpkg ambari-agent

apt-cache showpkg ambari-metrics-assembly
```

You should see the Ambari packages in the list.

4. Proceed to [Install the Ambari Agents manually](#).

Debian 7

On a server host that has Internet access, use a command line editor to perform the following steps:

1. Log in to your host as `root`.
2. Download the Ambari repository file to a directory on your installation host.

```
wget -nv http://public-repo-1.hortonworks.com/ambari/
debian7/2.x/updates/2.4.2.0/ambari.list -O /etc/apt/
sources.list.d/ambari.list

apt-key adv --recv-keys --keyserver keyserver.debian.com
B9733A7A07513CAD

apt-get update
```



Important

Do not modify the `ambari.list` file name. This file is expected to be available on the Ambari Server host during Agent registration.

3. Confirm that Ambari packages downloaded successfully by checking the package name list.

```
apt-cache showpkg ambari-server
```

```
apt-cache showpkg ambari-agent
apt-cache showpkg ambari-metrics-assembly
```

You should see the Ambari packages in the list.

4. Proceed to [Install the Ambari Agents manually](#).

1.2. Install the Ambari Agents Manually

Use the instructions specific to the OS family running on your agent hosts.

RHEL/CentOS/Oracle Linux

1. Install the Ambari Agent on every host in your cluster.

```
yum install ambari-agent
```

2. Using a text editor, configure the Ambari Agent by editing the `ambari-agent.ini` file as shown in the following example:

```
vi /etc/ambari-agent/conf/ambari-agent.ini

[server]

hostname=<your.ambari.server.hostname>

url_port=8440

secured_url_port=8441
```

3. Start the agent on every host in your cluster.

```
ambari-agent start
```

The agent registers with the Server on start.

SLES

1. Install the Ambari Agent on every host in your cluster.

```
zypper install ambari-agent
```

2. Configure the Ambari Agent by editing the `ambari-agent.ini` file as shown in the following example:

```
vi /etc/ambari-agent/conf/ambari-agent.ini

[server]

hostname=<your.ambari.server.hostname>

url_port=8440
```

```
secured_url_port=8441
```

3. Start the agent on every host in your cluster.

```
ambari-agent start
```

The agent registers with the Server on start.

Debian/Ubuntu

1. Install the Ambari Agent on every host in your cluster.

```
apt-get install ambari-agent
```

2. Configure the Ambari Agent by editing the `ambari-agent.ini` file as shown in the following example:

```
vi /etc/ambari-agent/conf/ambari-agent.ini
```

```
[server]
```

```
hostname=<your.ambari.server.hostname>
```

```
url_port=8440
```

```
secured_url_port=8441
```

3. Start the agent on every host in your cluster.

```
ambari-agent start
```

The agent registers with the Server on start.

2. Customizing HDP Services

- [Defining Service Users and Groups for a HDP 2.x Stack \[8\]](#)
- [Setting Properties That Depend on Service Usernames/Groups \[9\]](#)

2.1. Defining Service Users and Groups for a HDP 2.x Stack

The individual services in Hadoop run under the ownership of their respective Unix accounts. These accounts are known as service users. These service users belong to a special Unix group. "Smoke Test" is a service user dedicated specifically for running smoke tests on components during installation using the `Services` View of the Ambari Web GUI. You can also run service checks as the "Smoke Test" user on-demand after installation. You can customize any of these users and groups using the `Misc` tab during the `Customize Services` installation step.



Note

Use the `Skip Group Modifications` option to not modify the Linux groups in the cluster. Choosing this option is typically required if your environment manages groups using LDAP and not on the local Linux machines.

If you choose to customize names, Ambari checks to see if these custom accounts already exist. If they do not exist, Ambari creates them. The default accounts are always created during installation whether or not custom accounts are specified. These default accounts are not used and can be removed post-install.



Note

All new service user accounts, and any existing user accounts used as service users, must have a UID ≥ 1000 .

Service Users

Service*	Component	Default User Account
Accumulo	Accumulo Tracer, Accumulo Monitor, Accumulo GC, Accumulo Master	accumulo (HDP 2.2 or later)
Ambari Metrics	Metrics Collector, Metrics Monitor	ams
Ambari Infra	Infra Solr Instance	infra-solr
Atlas	Atlas Metadata Server	atlas (HDP 2.3 or later)
Falcon	Falcon Server	falcon
Flume	Flume Agents	flume
HBase	MasterServer RegionServer	hbase
HDFS	NameNode SecondaryNameNode DataNode	hdfs
Hive	Hive Metastore, HiveServer2	hive
HUE	HUE	hue

Service*	Component	Default User Account
Kafka	Kafka Broker	kafka
Knox	Knox Gateway	knox
Mahout	Mahout clients	mahout (HDP 2.2 or later)
MapReduce2	HistoryServer	mapred
Oozie	Oozie Server	oozie
PostgreSQL	PostgreSQL (with Ambari Server)	postgres (Created as part of installing the default PostgreSQL database with Ambari Server. If you are not using the Ambari PostgreSQL database, this user is not needed.)
Ranger	Ranger Admin, Ranger Usersync	ranger (HDP 2.2 or later)
Ranger KMS	Ranger KMS Server	kms (HDP 2.3 or later)
Slider	Slider clients	slider
Spark	Lively Servers	livy
Spark	Spark History Server	spark (HDP 2.2 or later)
Sqoop	Sqoop	sqoop
Storm	Masters (Nimbus, DRPC Server, Storm REST API, Server, Storm UI Server) Slaves (Supervisors, Logviewers)	storm
Tez	Tez clients	tez
WebHCat	WebHCat Server	hcat
YARN	NodeManager ResourceManager	yarn
Zeppelin Notebook	Zeppelin Notebook	zeppelin
ZooKeeper	ZooKeeper	zookeeper

*For all components, the Smoke Test user performs smoke tests against cluster services as part of the install process. It also can perform these on-demand, from the Ambari Web UI. The default user account for the smoke test user is ambari-qa.

Service Groups

Service	Components	Default Group Account
All	All	hadoop
Atlas	Atlas Metadata Server	atlas
Knox	Knox Gateway	knox
Ranger	Ranger Admin, Ranger Usersync	ranger
Ranger KMS	Ranger KMS Server	kms
Spark	Spark History Server	spark

2.2. Setting Properties That Depend on Service Usernames/Groups

Some properties must be set to match specific service user names or service groups. If you have set up non-default, customized service user names for the HDFS or HBase service or the Hadoop group name, you must edit the following properties, using `Services > Service.Name > Configs > Advanced`:

HDFS Settings: Advanced

Property Name	Value
dfs.permissions.superusergroup	The same as the HDFS username. The default is "hdfs"
dfs.cluster.administrators	A single space followed by the HDFS username.
dfs.block.local-path-access.user	The HBase username. The default is "hbase".

MapReduce Settings: Advanced

Property Name	Value
mapreduce.cluster.administrators	A single space followed by the Hadoop group name.

3. Using Custom Host Names

You can customize the agent registration host name and the public host name used for each host in Ambari. Use this capability when "hostname" does not return the public network host name for your machines.

[How to Customize the name of a host \[11\]](#)

3.1. How to Customize the name of a host

1. At the `Install Options` step in the Cluster Installer wizard, select `Perform Manual Registration for Ambari Agents`.
2. Install the Ambari Agents manually on each host, as described in [Install the Ambari Agents Manually](#).
3. To echo the customized name of the host to which the Ambari agent registers, for every host, create a script like the following example, named `/var/lib/ambari-agent/hostname.sh`. Be sure to `chmod` the script so it is executable by the Agent.

```
#!/bin/sh
```

```
echo 'hostname -f'
```

4. Open `/etc/ambari-agent/conf/ambari-agent.ini` on every host, using a text editor.
5. Add to the `[agent]` section the following line:

```
hostname_script=/var/lib/ambari-agent/hostname.sh
```

where `/var/lib/ambari-agent/hostname.sh` is the path to your custom hostname script.

6. To generate a public host name for every host, create a script like the following example, named `/var/lib/ambari-agent/public_hostname.sh` to show the name for that host in the UI. MAke sure that you `chmod` the script so it is executable by the Agent.

```
#!/bin/sh
```

```
echo 'hostname -f'
```

7. Open `/etc/ambari-agent/conf/ambari-agent.ini` on every host, using a text editor.
8. Add to the `[agent]` section the following line:

```
public_hostname_script=/var/lib/ambari-agent/public_hostname.sh
```

where `/var/lib/ambari-agent/public_hostname.sh` is the path to your custom public hostname script.

9. If applicable, add the host names to `/etc/hosts` on every host.

10. Restart the Agent on every host for these changes to take effect.

```
ambari-agent restart
```


4. Changing Host Names

In some situations, after you have created your cluster you might require changing the hostnames of the hosts. In addition to any infrastructure and environment changes you need to make, you also need to change the host names that Ambari is using to manage the HDP cluster. Use this procedure to change the host names in Ambari.



Important

It is **critical** that you make a backup of your Ambari database prior to executing this procedure.

1. You **must** disable Kerberos prior to executing this procedure.

Using **Ambari Web**, browse to **Admin > Kerberos** and click **Disable Kerberos**.

2. In the Ambari Background Operations dialog, stop all pending commands and jobs.
3. Stop all services.
4. Backup the Ambari database.
5. Stop ambari-server and ambari-agents on all hosts.

```
ambari-server stop
```

```
ambari-agent stop
```

6. Create *.json file with host names changes.

Example: host_names_changes.json

```
{
  "cluster1" : {
    "c6400.ambari.apache.org" : "c6410.ambari.apache.org",
    "c6401.ambari.apache.org" : "c6411.ambari.apache.org",
    ....
  }
}
```

where **cluster1** is cluster name and **"c6400.ambari.apache.org"** : **"c6410.ambari.apache.org"** is the host names pair in the format "current_host_name" : "new_host_name".

7. Execute the following command on the ambari-server host:

```
ambari-server update-host-names host_names_changes.json
```

8. After successful end of this action, please update host names for all nodes, according to changes that you added to *.json file.
9. If you changed the host name for the node on which the ambari server resides, then you must update that name for every ambari-agent.

In `/etc/ambari-agent/conf/ambari-agent.ini`, update the "hostname" field to the new host name for node on which the ambari-server resides.

10 Start ambari-server and ambari-agents on all hosts.

```
ambari-server start
```

```
ambari-agent start
```

11 Start all services, using Ambari Web.

For each, browse to **Services** > `<service_name>` > **Service Actions**, choose **Start**.



Note

If you have NameNode HA enabled, after starting the ZooKeeper service, you must:

- a. Start all ZooKeeper components.
- b. Execute the following command on both NameNode hosts:

```
hdfs zkfc -formatZK -force
```

12. If you disabled Kerberos before starting this procedure, you must [Enable Kerberos Security](#) by working through either the automated or manual setup documented in Hortonworks Data Platform Apache Ambari Security.



Note

If you enable Kerberos with the manual option, you **must** be sure to generate and deploy new keytabs that contain the new host names.

5. Moving the Ambari Server

To transfer an Ambari Server that uses the default, embedded, PostgreSQL database from one host to a new host, use the following instructions:

1. [Back up current data](#) - from the original Ambari Server database.
2. [Update all Agents](#) - to point to the new Ambari Server.
3. [Install the New Ambari Server](#) - on the new host and populate databases with information from the original Server.



Note

If your Ambari Server is using one of the non-default databases (such as MySQL, Oracle, or an existing PostgreSQL instance) then be sure to follow backup, restore, and stop/start procedures that match that database type.

5.1. Back up Current Data

1. On the Ambari Server host, stop the original Ambari Server.

```
ambari-server stop
```

2. Create a directory to hold the database backups.

```
cd /tmp
```

```
mkdir dbdumps/
```

```
cd dbdumps/
```

3. Create the database backups.

```
pg_dump -U {ambari.db.username} -f ambari.sql
```

```
Password: {ambari.db.password}
```

where the following:

Variable	Description	Default
ambari.db.username	The database username.	ambari
ambari.db.password	The database password.	bigdata

4. Create a backup of the Ambari Server meta info.

```
ambari-server backup
```

5.2. Update all Agents

1. On each agent host, stop the agent.

```
ambari-agent stop
```

2. Remove old agent certificates (if any exist).

```
rm /var/lib/ambari-agent/keys/*
```

3. Using a text editor, edit `/etc/ambari-agent/conf/ambari-agent.ini` to point to the new host.

```
[server]

hostname={new.ambari.server.fqdn}

url_port=8440

secured_url_port=8441
```

5.3. Install the New Ambari Server

1. Install the new Ambari Server on the new host.

```
yum install ambari-server
```

2. Run setup the Ambari Server and setup similar to how the original Ambari Server is configured.

```
ambari-server setup
```

3. Restart the PostgreSQL instance.

```
service postgresql restart
```

4. Open the PostgreSQL interactive terminal.

```
su - postgres
```

```
psql
```

5. Using the interactive terminal, drop the "ambari" database created by the new ambari setup and install.

```
drop database ambari;
```

6. Check to make sure the databases have been dropped. The "ambari" databases should not be listed.

```
\l
```

7. Create new "ambari" database to hold the transferred data.

```
create database ambari;
```

8. Exit the PostgreSQL interactive terminal.

\q

9. Copy the saved data (/tmp/dbdumps/ambari.sql) from [Back up Current Data](#) to the new Ambari Server host.

10. Load the saved data into the new database.

```
psql -d ambari -f /tmp/dbdumps/ambari.sql
```

11. Start the new Server.

```
ambari-server start
```

12. On each Agent host, start the Ambari Agent.

```
ambari-agent start
```

13. Open Ambari Web. Point your browser to:

```
<new.Ambari.Server>:8080
```

The new Ambari Server is ready to use.

6. Moving the ZooKeeper Server

Use the following steps so move the ZooKeeper server to a new host:

1. Stop the ZooKeeper server.
2. Select **Hosts** on the Ambari dashboard, then select the host on which to install the new ZooKeeper server.
3. On the Summary page of the new ZooKeeper host, select **Add > ZooKeeper Server** and add the new ZooKeeper server.
4. Update the following properties on the new ZooKeeper server (use the existing ZooKeeper server settings as a reference).
 - `ha.zookeeper.quorum`
 - `hbase.zookeeper.quorum`
 - `templeton.zookeeper.hosts`
 - `yarn.resourcemanager.zk-address`
 - `hive.zookeeper.quorum`
 - `hive.cluster.delegation.token.store.zookeeper.connectString`
5. Select **Hosts** on the Ambari dashboard, then select the original ZooKeeper server host. Select **ZooKeeper > Service Actions > Delete Service** to delete the original ZooKeeper server.
6. Save the HDFS namespace.
7. Restart the new ZooKeeper server and the Hive service.



Note

In Ambari 2.4.0.0, adding or removing ZooKeeper servers requires manually editing the following Atlas properties. Select **Atlas > Configs > Advanced**, then select **Advanced application-properties** and edit the following properties to reflect the new ZooKeeper server settings:

- `atlas.graph.index.search.solr.zookeeper-url`

Example format:

```
host1:2181/infra-solr,host2:2181/infra-solr,host3:2181/infra-solr
```

- `atlas.kafka.zookeeper.connect`

Example format:

```
host1:2181,host2:2181,host3:2181
```

- `atlas.audit.hbase.zookeeper.quorum`

Example format:

`host1,host2,host3`

After updating these properties, restart Atlas and the following services that contain Atlas hooks (in order to refresh the configuration files):

- Hive
- Storm
- Falcon
- Sqoop

7. Configuring LZO Compression

LZO is a lossless data compression library that favors speed over compression ratio. Ambari does not install nor enable LZO Compression by default. To enable LZO compression in your HDP cluster, you must [Configure core-site.xml for LZO](#).

Optionally, you can implement LZO to optimize Hive queries in your cluster for speed. For more information about using LZO compression with Hive, see [Running Compression with Hive Queries](#).

7.1. Configure core-site.xml for LZO

1. Browse to Ambari Web > Services > HDFS > Configs, then expand Advanced core-site.
2. Find the `io.compression.codecs` property key.
3. Append to the `io.compression.codecs` property key, the following value:
`com.hadoop.compression.lzo.LzoCodec`
4. Add a description of the config modification, then choose Save.
5. Expand the `Custom core-site.xml` section.
6. Select Add Property.
7. Add to `Custom core-site.xml` the following property key and value

Property Key	Property Value
<code>io.compression.codec.lzo.class</code>	<code>com.hadoop.compression.lzo.LzoCodec</code>

8. Choose Save.
9. Add a description of the config modification, then choose Save.
10. Restart the HDFS, MapReduce2 and YARN services.



Note

If performing a Restart or a Restart All does not start the required package install, you may need to stop, then start the HDFS service to install the necessary LZO packages. Restart is only available for a service in the "Running" or "Started" state.

7.2. Running Compression with Hive Queries

Running Compression with Hive Queries requires creating LZO files. To create LZO files, use one of the following procedures:

- [Create LZO Files \[21\]](#)

- [Write Custom Java to Create LZO Files \[21\]](#)

7.2.1. Create LZO Files

1. Create LZO files as the output of the Hive query.
2. Use `lzo` command utility or your custom Java to generate `lzo.index` for the `.lzo` files.

Hive Query Parameters

Prefix the query string with these parameters:

```
SET mapreduce.output.fileoutputformat.compress.codec=com.hadoop.compression.lzo.LzoCodec
SET hive.exec.compress.output=true
SET mapreduce.output.fileoutputformat.compress=true
```

For example:

```
hive -e "SET mapreduce.output.fileoutputformat.compress.codec=com.hadoop.compression.lzo.LzoCodec;SET hive.exec.compress.output=true;SET mapreduce.output.fileoutputformat.compress=true;"
```

7.2.2. Write Custom Java to Create LZO Files

1. Create text files as the output of the Hive query.
2. Write custom Java code to
 - convert Hive query generated text files to `.lzo` files
 - generate `lzo.index` files for the `.lzo` files

Hive Query Parameters

Prefix the query string with these parameters:

```
SET hive.exec.compress.output=false
SET mapreduce.output.fileoutputformat.compress=false
```

For example:

```
hive -e "SET hive.exec.compress.output=false;SET mapreduce.output.fileoutputformat.compress=false;<query-string>"
```

8. Using Non-Default Databases

Use the following instructions to prepare a non-default database for Ambari, Hive, or Oozie. You must complete these instructions before you set up the Ambari Server by running `ambari-server setup`.

- [Using Non-Default Databases - Ambari \[22\]](#)
- [Using Non-Default Databases - Hive \[27\]](#)
- [Using Non-Default Databases - Oozie \[31\]](#)



Important

Using the **Microsoft SQL Server** or **SQL Anywhere** database options are not supported.

8.1. Using Non-Default Databases - Ambari

The following sections describe how to use Ambari with an existing database, other than the embedded PostgreSQL database instance that Ambari Server uses by default.

- [Using Ambari with Oracle \[22\]](#)
- [Using Ambari with MySQL/MariaDB \[23\]](#)
- [Using Ambari with PostgreSQL \[25\]](#)
- [Troubleshooting Non-Default Databases with Ambari \[26\]](#)



Important

Using the **Microsoft SQL Server** or **SQL Anywhere** database options are not supported.



Important

For High Availability (HA) purposes, it is **required** that the relational database used with Ambari is also made highly available following best practices for the given database type.

8.1.1. Using Ambari with Oracle

To set up Oracle for use with Ambari:

1. On the Ambari Server host, install the appropriate `JDBC.jar` file.
 - a. Download the Oracle JDBC (OJDBC) driver from <http://www.oracle.com/technetwork/database/features/jdbc/index-091264.html>.
 - b. For **Oracle Database 11g**: `select Oracle Database 11g Release 2 drivers > ojdbc6.jar.`

c. For **Oracle Database 12c**: select Oracle Database 12c Release 1 drivers > ojdbc7.jar.

d. Copy the .jar file to the Java share directory. For example:

```
cp ojdbc7.jar /usr/share/java/
```

e. Make sure the .jar file has the appropriate permissions. For example:

```
chmod 644 /usr/share/java/ojdbc7.jar
```

2. Create a user for Ambari and grant that user appropriate permissions.

For example, using the Oracle database admin utility, run the following commands:

```
# sqlplus sys/root as sysdba

CREATE USER <AMBARIUSER> IDENTIFIED BY <AMBARIPASSWORD> default
tablespace "USERS" temporary tablespace "TEMP";

GRANT unlimited tablespace to <AMBARIUSER>;

GRANT create session to <AMBARIUSER>;

GRANT create TABLE to <AMBARIUSER>;

GRANT create SEQUENCE to <AMBARIUSER>;

QUIT;
```

Where <AMBARIUSER> is the Ambari user name and <AMBARIPASSWORD> is the Ambari user password.

3. Load the Ambari Server database schema.

a. You must pre-load the Ambari database schema into your Oracle database using the schema script.

```
sqlplus <AMBARIUSER>/<AMBARIPASSWORD> < Ambari-DDL-Oracle-
CREATE.sql
```

b. Find the Ambari-DDL-Oracle-CREATE.sql file in the /var/lib/ambari-server/resources/ directory of the Ambari Server host after you have installed Ambari Server.

4. When setting up the Ambari Server, select Advanced Database Configuration > Option [2] Oracle and respond to the prompts using the username/password credentials you created in step 2.

8.1.2. Using Ambari with MySQL/MariaDB

To set up MySQL for use with Ambari:

1. On the Ambari Server host, install the connector.

a. Install the connector

RHEL/CentOS/Oracle Linux

```
yum install mysql-connector-java
```

SLES

```
zypper install mysql-connector-java
```

Debian/Ubuntu

```
apt-get install libmysql-java
```

b. Confirm that `.jar` is in the Java share directory.

```
ls /usr/share/java/mysql-connector-java.jar
```

c. Make sure the `.jar` file has the appropriate permissions - 644.

2. Create a user for Ambari and grant it permissions.

- For example, using the MySQL database admin utility:

```
# mysql -u root -p
```

```
CREATE USER '<AMBARIUSER>'@'%' IDENTIFIED BY  
'<AMBARIPASSWORD>';
```

```
GRANT ALL PRIVILEGES ON *.* TO '<AMBARIUSER>'@'%' ;
```

```
CREATE USER '<AMBARIUSER>'@'localhost' IDENTIFIED BY  
'<AMBARIPASSWORD>';
```

```
GRANT ALL PRIVILEGES ON *.* TO '<AMBARIUSER>'@'localhost' ;
```

```
CREATE USER '<AMBARIUSER>'@'<AMBARISERVERFQDN>' IDENTIFIED BY  
'<AMBARIPASSWORD>';
```

```
GRANT ALL PRIVILEGES ON *.* TO  
'<AMBARIUSER>'@'<AMBARISERVERFQDN>';
```

```
FLUSH PRIVILEGES;
```

- Where `<AMBARIUSER>` is the Ambari user name, `<AMBARIPASSWORD>` is the Ambari user password and `<AMBARISERVERFQDN>` is the Fully Qualified Domain Name of the Ambari Server host.

3. Load the Ambari Server database schema.

- You must pre-load the Ambari database schema into your MySQL/MariaDB database using the schema script.

```
CREATE DATABASE <AMBARIDATABASE>;

USE <AMBARIDATABASE>;

SOURCE Ambari-DDL-MySQL-CREATE.sql;
```

- Where <AMBARIUSER> is the Ambari user name and <AMBARIDATABASE> is the Ambari database name.

Find the `Ambari-DDL-MySQL-CREATE.sql` file in the `/var/lib/ambari-server/resources/` directory of the Ambari Server host after you have installed Ambari Server.

4. When setting up the Ambari Server, select Advanced Database Configuration > Option [3] MySQL/MariaDB and enter the credentials you defined in Step 2. for user name, password and database name.

8.1.3. Using Ambari with PostgreSQL

To set up PostgreSQL for use with Ambari:

1. Create a user for Ambari and grant it permissions.

- Using the PostgreSQL database admin utility:

```
# sudo -u postgres psql

CREATE DATABASE <AMBARIDATABASE>;

CREATE USER <AMBARIUSER> WITH PASSWORD '<AMBARIPASSWORD>';

GRANT ALL PRIVILEGES ON DATABASE <AMBARIDATABASE> TO
<AMBARIUSER>;

\connect <AMBARIDATABASE>;

CREATE SCHEMA <AMBARISCHEMA> AUTHORIZATION <AMBARIUSER>;

ALTER SCHEMA <AMBARISCHEMA> OWNER TO <AMBARIUSER>;

ALTER ROLE <AMBARIUSER> SET search_path to '<AMBARISCHEMA>',
'public';
```

- Where <AMBARIUSER> is the Ambari user name <AMBARIPASSWORD> is the Ambari user password, <AMBARIDATABASE> is the Ambari database name and <AMBARISCHEMA> is the Ambari schema name.

2. Load the Ambari Server database schema.

- You must pre-load the Ambari database schema into your PostgreSQL database using the schema script.

```
# psql -U <AMBARIUSER> -d <AMBARIDATABASE>
```

```
\connect <AMBARIDATABASE>;  
  
\i Ambari-DDL-Postgres-CREATE.sql;
```

- Find the `Ambari-DDL-Postgres-CREATE.sql` file in the `/var/lib/ambari-server/resources/` directory of the Ambari Server host after you have installed Ambari Server.

3. When setting up the Ambari Server, select `Advanced Database Configuration > Option[4] PostgreSQL` and enter the credentials you defined in Step 2. for user name, password, and database name.

8.1.4. Troubleshooting Non-Default Databases with Ambari

Use these topics to help troubleshoot any issues you might have installing Ambari with an existing Oracle database.

8.1.4.1. Problem: Ambari Server Fails to Start: No Driver

Check `/var/log/ambari-server/ambari-server.log` for the following error:

```
ExceptionDescription:ConfigurationError.Class[oracle.jdbc.driver.OracleDriver]  
not found.
```

The Oracle JDBC.jar file cannot be found.

8.1.4.1.1. Solution

Make sure the file is in the appropriate directory on the Ambari server and re-run `ambari-server setup`. Review the load database procedure appropriate for your database type in [Using Non-Default Databases - Ambari](#).

8.1.4.2. Problem: Ambari Server Fails to Start: No Connection

Check `/var/log/ambari-server/ambari-server.log` for the following error:

```
The Network Adapter could not establish the connection Error Code:  
17002
```

Ambari Server cannot connect to the database.

8.1.4.2.1. Solution

Confirm that the database host is reachable from the Ambari Server and is correctly configured by reading `/etc/ambari-server/conf/ambari.properties`.
`server.jdbc.url=jdbc:oracle:thin:@oracle.database.hostname:1521/
ambaridb`
`server.jdbc.rca.url=jdbc:oracle:thin:@oracle.database.hostname:1521/
ambari`

8.1.4.3. Problem: Ambari Server Fails to Start: Bad Username

Check `/var/log/ambari-server/ambari-server.log` for the following error:

```
Internal Exception: java.sql.SQLException:ORA01017: invalid
username/password; logon denied
```

You are using an invalid username/password.

8.1.4.3.1. Solution

Confirm the user account is set up in the database and has the correct privileges. See Step 3 above.

8.1.4.4. Problem: Ambari Server Fails to Start: No Schema

Check `/var/log/ambari-server/ambari-server.log` for the following error:

```
Internal Exception: java.sql.SQLSyntaxErrorException: ORA00942:
table or view does not exist
```

The schema has not been loaded.

8.1.4.4.1. Solution

Confirm you have loaded the database schema. Review the load database schema procedure appropriate for your database type in [Using Non-Default Databases - Ambari](#).

8.2. Using Non-Default Databases - Hive

The following sections describe how to use Hive with an existing database, other than the MySQL database instance that Ambari installs by default.

- [Using Hive with Oracle \[27\]](#)
- [Using Hive with MySQL \[28\]](#)
- [Using Hive with PostgreSQL \[29\]](#)
- [Troubleshooting Non-Default Databases with Hive \[30\]](#)



Important

Using the **Microsoft SQL Server** or **SQL Anywhere** database options are not supported.

8.2.1. Using Hive with Oracle

To set up Oracle for use with Hive:

1. On the Ambari Server host, stage the appropriate JDBC driver file for later deployment.
 - a. Download the Oracle JDBC (OJDBC) driver from <http://www.oracle.com/technetwork/database/features/jdbc/index-091264.html>.

- b. For **Oracle Database 11g**: select Oracle Database 11g Release 2 drivers > ojdbc6.jar.
- c. For **Oracle Database 12c**: select Oracle Database 12c Release 1 drivers > ojdbc7.jar.

d. Make sure the .jar file has the appropriate permissions. For example:

```
chmod 644 ojdbc7.jar
```

e. Execute the following command, adding the path to the downloaded .jar file:

```
ambari-server setup --jdbc-db=oracle --jdbc-driver=/path/to/  
downloaded/ojdbc7.jar
```

2. Create a user for Hive and grant it permissions.

- Using the Oracle database admin utility:

```
# sqlplus sys/root as sysdba  
  
CREATE USER <HIVEUSER> IDENTIFIED BY <HIVEPASSWORD>;  
  
GRANT SELECT_CATALOG_ROLE TO <HIVEUSER>;  
  
GRANT CONNECT, RESOURCE TO <HIVEUSER>;  
  
QUIT;
```

- Where <HIVEUSER> is the Hive user name and <HIVEPASSWORD> is the Hive user password.

8.2.2. Using Hive with MySQL

To set up MySQL for use with Hive:

1. On the Ambari Server host, stage the appropriate MySQL connector for later deployment.

a. Install the connector.

RHEL/CentOS/Oracle Linux

```
yum install mysql-connector-java*
```

SLES

```
zypper install mysql-connector-java*
```

Debian/Ubuntu

```
apt-get install libmysql-java
```

b. Confirm that mysql-connector-java.jar is in the Java share directory.


```
ls /usr/share/java/mysql-connector-java.jar
```

c. Make sure the .jar file has the appropriate permissions - 644.

d. Execute the following command:

```
ambari-server setup --jdbc-db=mysql --jdbc-driver=/usr/share/
java/mysql-connector-java.jar
```

2. Create a user for Hive and grant it permissions.

- Using the MySQL database admin utility:

```
# mysql -u root -p

CREATE USER '<HIVEUSER>'@'localhost' IDENTIFIED BY
'<HIVEPASSWORD>';

GRANT ALL PRIVILEGES ON *.* TO '<HIVEUSER>'@'localhost';

CREATE USER '<HIVEUSER>'@'%' IDENTIFIED BY '<HIVEPASSWORD>';

GRANT ALL PRIVILEGES ON *.* TO '<HIVEUSER>'@'%';

CREATE USER '<HIVEUSER>'@'<HIVEMETASTOREFQDN>' IDENTIFIED BY
'<HIVEPASSWORD>';

GRANT ALL PRIVILEGES ON *.* TO
'<HIVEUSER>'@'<HIVEMETASTOREFQDN>';

FLUSH PRIVILEGES;
```

- Where <HIVEUSER> is the Hive user name, <HIVEPASSWORD> is the Hive user password and <HIVEMETASTOREFQDN> is the Fully Qualified Domain Name of the Hive Metastore host.

3. Create the Hive database.

The Hive database must be created before loading the Hive database schema.

```
# mysql -u root -p

CREATE DATABASE <HIVEDATABASE>
```

Where <HIVEDATABASE> is the Hive database name.

8.2.3. Using Hive with PostgreSQL

To set up PostgreSQL for use with Hive:

1. On the Ambari Server host, stage the appropriate PostgreSQL connector for later deployment.

a. Install the connector.

RHEL/CentOS/Oracle Linux

```
yum install postgresql-jdbc*
```

SLES

```
zypper install -y postgresql-jdbc
```

Debian/Ubuntu

```
apt-get install -y libpostgresql-jdbc-java
```

- b. Confirm that .jar is in the Java share directory.

```
ls /usr/share/java/postgresql-jdbc.jar
```

- c. Change the access mode of the .jar file to 644.

```
chmod 644 /usr/share/java/postgresql-jdbc.jar
```

- d. Execute the following command:

```
ambari-server setup --jdbc-db=postgres --jdbc-driver=/usr/
share/java/postgresql-jdbc.jar
```

2. Create a user for Hive and grant it permissions.

- Using the PostgreSQL database admin utility:

```
echo "CREATE DATABASE <HIVEDATABASE>;" | psql -U postgres
```

```
echo "CREATE USER <HIVEUSER> WITH PASSWORD '<HIVEPASSWORD>';" |
psql -U postgres
```

```
echo "GRANT ALL PRIVILEGES ON DATABASE <HIVEDATABASE> TO
<HIVEUSER>;" | psql -U postgres
```

- Where <HIVEUSER> is the Hive user name, <HIVEPASSWORD> is the Hive user password and <HIVEDATABASE> is the Hive database name.

8.2.4. Troubleshooting Non-Default Databases with Hive

Use these entries to help you troubleshoot any issues you might have installing Hive with non-default databases.

8.2.4.1. Problem: Hive Metastore Install Fails Using Oracle

Check the install log:

```
cp /usr/share/java/${jdbc_jar_name} ${target}] has failures: true
```

The Oracle JDBC.jar file cannot be found.

8.2.4.1.1. Solution

Make sure the file is in the appropriate directory on the Hive Metastore server and click **Retry**.

8.2.4.2. Problem: Install Warning when "Hive Check Execute" Fails Using Oracle

Check the install log:

```
java.sql.SQLException: ORA-01754: a table may contain
only one column of type LONG
```

The Hive Metastore schema was not properly loaded into the database.

8.2.4.2.1. Solution

Ignore the warning, and complete the install. Check your database to confirm the Hive Metastore schema is loaded. In the Ambari Web GUI, browse to **Services > Hive**. Choose **Service Actions > Service Check** to check that the schema is correctly in place.

8.2.4.3. Problem: Hive Check Execute may fail after completing an Ambari upgrade to version 1.4.2

For secure and non-secure clusters, with Hive security authorization enabled, the Hive service check may fail. Hive security authorization may not be configured properly.

8.2.4.3.1. Solution

Two workarounds are possible. Using Ambari Web, in **HiveConfigsAdvanced**:

- Disable `hive.security.authorization`, by setting the `hive.security.authorization.enabled` value to `false`.

or

- Properly configure Hive security authorization. For example, set the following properties:

For more information about configuring Hive security, see [Metastore Server Security](#) and [Hive Authorization](#).

Table 8.1. Hive Security Authorization Settings

Property	Value
<code>hive.security.authorization.manager</code>	<code>org.apache.hadoop.hive.ql.security.authorization.StorageBasedAuthorizationProvider</code>
<code>hive.security.metastore.authorization.manager</code>	<code>org.apache.hadoop.hive.ql.security.authorization.StorageBasedAuthorizationProvider</code>
<code>hive.security.authenticator.manager</code>	<code>org.apache.hadoop.hive.ql.security.ProxyUserAuthenticator</code>

8.3. Using Non-Default Databases - Oozie

The following sections describe how to use Oozie with an existing database, other than the Derby database instance that Ambari installs by default.

- [Using Oozie with Oracle \[32\]](#)
- [Using Oozie with MySQL \[33\]](#)
- [Using Oozie with PostgreSQL \[34\]](#)
- [Troubleshooting Non-Default Databases with Oozie \[35\]](#)



Important

Using the **Microsoft SQL Server** or **SQL Anywhere** database options are not supported.

8.3.1. Using Oozie with Oracle

To set up Oracle for use with Oozie:

1. On the Ambari Server host, install the appropriate JDBC driver file.
 - a. Download the Oracle JDBC (OJDBC) driver from <http://www.oracle.com/technetwork/database/features/jdbc/index-091264.html>.
 - b. For **Oracle Database 11g**: select Oracle Database 11g Release 2 drivers > ojdbc6.jar.
 - c. For **Oracle Database 12c**: select Oracle Database 12c Release 1 drivers > ojdbc7.jar.
 - d. Make sure the .jar file has the appropriate permissions. For example:

```
chmod 644 ojdbc7.jar
```
 - e. Execute the following command, adding the path to the downloaded .jar file:

```
ambari-server setup --jdbc-db=oracle --jdbc-driver=/path/to/downloaded/ojdbc7.jar
```
2. Create a user for Oozie and grant it permissions.

Using the Oracle database admin utility, run the following commands:

```
# sqlplus sys/root as sysdba

CREATE USER <OOZIEUSER> IDENTIFIED BY <OOZIEPASSWORD>;

GRANT ALL PRIVILEGES TO <OOZIEUSER>;

GRANT CONNECT, RESOURCE TO <OOZIEUSER>;

QUIT;
```

Where <OOZIEUSER> is the Oozie user name and <OOZIEPASSWORD> is the Oozie user password.

8.3.2. Using Oozie with MySQL

To set up MySQL for use with Oozie:

1. On the Ambari Server host, stage the appropriate MySQL connector for later deployment.

- a. Install the connector.

RHEL/CentOS/Oracle Linux

```
yum install mysql-connector-java*
```

SLES

```
zypper install mysql-connector-java*
```

Debian/Ubuntu

```
apt-get install libmysql-java
```

- b. Confirm that `mysql-connector-java.jar` is in the Java share directory.

```
ls /usr/share/java/mysql-connector-java.jar
```

- c. Make sure the `.jar` file has the appropriate permissions - 644.

- d. Execute the following command:

```
ambari-server setup --jdbc-db=mysql --jdbc-driver=/usr/share/
java/mysql-connector-java.jar
```

2. Create a user for Oozie and grant it permissions.

- Using the MySQL database admin utility:

```
# mysql -u root -p
```

```
CREATE USER '<OOZIEUSER>'@'%' IDENTIFIED BY '<OOZIEPASSWORD>';
```

```
GRANT ALL PRIVILEGES ON *.* TO '<OOZIEUSER>'@'%' ;
```

```
FLUSH PRIVILEGES;
```

- Where `<OOZIEUSER>` is the Oozie user name and `<OOZIEPASSWORD>` is the Oozie user password.

3. Create the Oozie database.

- The Oozie database must be created prior.

```
# mysql -u root -p
```

```
CREATE DATABASE <OOZIEDATABASE>
```

- Where <OOZIEDATABASE> is the Oozie database name.

8.3.3. Using Oozie with PostgreSQL

To set up PostgreSQL for use with Oozie:

1. On the Ambari Server host, stage the appropriate PostgreSQL connector for later deployment.

- a. Install the connector.

RHEL/CentOS/Oracle Linux

```
yum install postgresql-jdbc
```

SLES

```
zypper install -y postgresql-jdbc
```

Debian/Ubuntu

```
apt-get install -y libpostgresql-jdbc-java
```

- b. Confirm that .jar is in the Java share directory.

```
ls /usr/share/java/postgresql-jdbc.jar
```

- c. Change the access mode of the .jar file to 644.

```
chmod 644 /usr/share/java/postgresql-jdbc.jar
```

- d. Execute the following command:

```
ambari-server setup --jdbc-db=postgres --jdbc-driver=/usr/share/java/postgresql-jdbc.jar
```

2. Create a user for Oozie and grant it permissions.

- Using the PostgreSQL database admin utility:

```
echo "CREATE DATABASE <OOZIEDATABASE>;" | psql -U postgres
```

```
echo "CREATE USER <OOZIEUSER> WITH PASSWORD '<OOZIEPASSWORD>';"  
| psql -U postgres
```

```
echo "GRANT ALL PRIVILEGES ON DATABASE <OOZIEDATABASE> TO  
<OOZIEUSER>;" | psql -U postgres
```

- Where <OOZIEUSER> is the Oozie user name, <OOZIEPASSWORD> is the Oozie user password and <OOZIEDATABASE> is the Oozie database name.

8.3.4. Troubleshooting Non-Default Databases with Oozie

Use these entries to help you troubleshoot any issues you might have installing Oozie with non-default databases.

8.3.4.1. Problem: Oozie Server Install Fails Using MySQL

Check the install log:

```
cp /usr/share/java/mysql-connector-java.jar usr/lib/oozie/libext/  
mysql-connector-java.jar has failures: true
```

The MySQL JDBC.jar file cannot be found.

8.3.4.1.1. Solution

Make sure the file is in the appropriate directory on the Oozie server and click **Retry**.

8.3.4.2. Problem: Oozie Server Install Fails Using Oracle or MySQL

Check the install log:

```
Exec[exec cd /var/tmp/oozie && /usr/lib/oozie/bin/ooziedb.sh  
create -sqlfile oozie.sql -run ] has failures: true
```

Oozie was unable to connect to the database or was unable to successfully setup the schema for Oozie.

8.3.4.2.1. Solution

Check the database connection settings provided during the `Customize Services` step in the install wizard by browsing back to `Customize Services > Oozie`. After confirming and adjusting your database settings, proceed forward with the install wizard.

If the Install Oozie Server wizard continues to fail, get more information by connecting directly to the Oozie server and executing the following command as `<OOZIEUSER>`:

```
su oozie /usr/lib/oozie/bin/ooziedb.sh create -sqlfile oozie.sql -  
run
```

9. Setting up Ambari to use an Internet Proxy Server

If you plan to use the **public repositories** (i.e. available on the Internet) for installing the cluster software, you need to make sure Ambari and the hosts in the cluster have Internet access to obtain the software from those repositories. Specifically:

- **Ambari Server:** uses Internet access to validate the repositories.
- **yum** (or equivalent package manager depending on your operating system): performs the software installation from the repositories.

Therefore, if your environment requires use of an Internet proxy server for access, you must configure Ambari Server component and “yum” on all the hosts to use the proxy server.



Note

Ambari can install software if you have no Internet access. If you have no Internet access (via a proxy server or otherwise), you can use local repositories for installing the cluster software. In that case, configuring Ambari to use a proxy server is not required. However, Ambari and the hosts in the cluster must have access to your local repositories. See [Using a Local Repository](#) for more information on setting up and using local repositories.

Configure Internet Proxy Settings for Ambari Server

1. On the Ambari Server host, stop Ambari Server:

```
ambari-server stop
```

2. Add proxy settings to the following script: `/var/lib/ambari-server/ambari-env.sh`.

```
-Dhttp.proxyHost=<yourProxyHost> -  
Dhttp.proxyPort=<yourProxyPort>
```

3. Optionally, to prevent some host names from accessing the proxy server, define the list of excluded hosts, as follows:

```
-Dhttp.nonProxyHosts=<pipe|separated|list|of|hosts>
```

4. If your proxy server requires authentication, add the username and password, as follows:

```
-Dhttp.proxyUser=<username> -Dhttp.proxyPassword=<password>
```

5. Restart the Ambari Server to pick up this change.

Configure yum for Internet Proxy Settings for All Hosts

Setting up yum to use a proxy server depends a lot on your environment and operating system. The instructions below provide some guidance but we **strongly recommend**

you consult with your System Administrators and Operating System documentation for assistance & specific instructions.

1. On each host in the cluster, specify the proxy settings in `/etc/yum.conf` by adding the following entry:

```
proxy=http://<yourProxyHost>:<yourProxyPort>
```

2. If your proxy server requires authentication, add the username and password, as follows:

```
enableProxyAuth=1
```

```
proxy_username=<username>
```

```
proxy_password=<password>
```

3. Save the yum configuration file.

It is important to highlight that defining a proxy server, username and password in `/etc/yum.conf` means **all users of yum connect to the proxy server with those details**. Please consult your System Administrators and refer to your Operating System documentation for more details on this configuration and possible alternatives.

Operating System	Reference
CentOS / Red Hat	https://www.centos.org/docs/5/html/yum/sn-yum-proxy-server.html
Oracle Linux	https://docs.oracle.com/cd/E37670_01/E37355/html/ol_proxy_config.html
Ubuntu / Debian	https://help.ubuntu.com/community/AptGet/Howto

10. Configuring Network Port Numbers

This chapter lists port number assignments required to maintain communication between Ambari Server, Ambari Agents, and Ambari Web.

- [Default Network Port Numbers - Ambari \[38\]](#)
- [Optional: Changing the Default Ambari Server Port \[38\]](#)
- [Optional: Changing the Ambari Server-Agent Port \[39\]](#)

For more information about configuring port numbers for Stack components, see [Configuring Ports](#) in the HDP Stack documentation.

10.1. Default Network Port Numbers - Ambari

The following table lists the default ports used by Ambari Server and Ambari Agent services.

Service	Servers	Default Ports Used	Protocol	Description	Need End User Access?	Configuration Parameters
Ambari Server	Ambari Server host	8080 See Optional: Changing the Ambari Server Port for instructions on changing the default port.	http See Configure Ambari Server for Authentication HTTP for instructions.	Interface to Ambari Web and Ambari REST API	No	
Ambari Server	Ambari Server host	8440	https	Handshake Port for Ambari Agents to Ambari Server	No	
Ambari Server	Ambari Server host	8441	https	Registration and Heartbeat Port for Ambari Agents to Ambari Server	No	
Ambari Agent	All hosts running Ambari Agents	8670 You can change the Ambari Agent ping port in the Ambari Agent configuration.	tcp	Ping port used for alerts to check the health of the Ambari Agent	No	

10.2. Optional: Changing the Default Ambari Server Port

By default, Ambari Server uses port 8080 to access the Ambari Web UI and the REST API. To change the port number, you must edit the Ambari properties file.

Ambari Server should not be running when you change port numbers. Edit `ambari.properties` before you start Ambari Server the first time or stop Ambari Server before editing properties.

1. On the Ambari Server host, open `/etc/ambari-server/conf/ambari.properties` with a text editor.
2. Add the client API port property and set it to your desired port value:

```
client.api.port=<port_number>
```

3. Start or re-start the Ambari Server. Ambari Server now accesses Ambari Web via the newly configured port:

```
http://<your.ambari.server>:<port_number>
```

10.3. Optional: Changing the Ambari Server-Agent Port

By default, Ambari Server uses port 8187 to communicate with Ambari Agents. To change the port number, you must edit the Ambari properties file.

Ambari Server should not be running when you change port numbers. Edit `ambari.properties` before you start Ambari Server the first time or stop Ambari Server before editing properties.

1. On the Ambari Server host, open `/etc/ambari-server/conf/ambari.properties` with a text editor.
2. Add the following properties and set them to your desired port values:

```
security.server.two_way_ssl.port=5222  
security.server.one_way_ssl.port=5223
```

3. On every Ambari Agent host, open `/etc/ambari-agent/conf/ambari-agent.ini` with a text editor.
4. Add the following properties and set them to your desired port values:

```
url_port=5223  
secured_url_port=5222
```

5. Start or re-start the Ambari Server. Ambari Server now accesses Ambari Agents via the newly configured port:

```
http://<your.ambari.server>:<port_number>
```

11. Change the JDK Version

During your initial Ambari Server Setup, you selected the JDK to use or provided a path to a custom JDK already installed on your hosts. After setting up your cluster, you may change the JDK version using the following procedure.

The choice of JDK is dependent on which HDP Stack you plan to install in your cluster. The following table indicates which JDKs work with which Stacks.

Stack	JDKs
HDP 2.3, 2.4 or 2.5	JDK 1.7 or JDK 1.8
HDP 2.2	JDK 1.7



Important

If you plan to upgrade between Stacks (for example, go from HDP 2.2 -> HDP 2.3, 2.4 or 2.5) **do not change the JDK** until you have successfully upgraded the Stack and are running the cluster on the target Stack. For example, the high-level process should follow:

1. Running HDP 2.2 with JDK 1.7.
2. Perform Stack upgrade to HDP 2.3, 2.4 or 2.5.
3. Change JDK from 1.7 to 1.8 (using the following procedure).

How to change the JDK version for an existing cluster

1. Re-run Ambari Server Setup.

```
ambari-server setup
```

2. At the prompt to change the JDK, Enter **y**.

```
Do you want to change Oracle JDK [y/n] (n)? y
```

3. At the prompt to choose a JDK, Enter **1** to change the JDK to v1.8.

```
[1] - Oracle JDK 1.8
```

```
[2] - Oracle JDK 1.7
```

```
[3] - Custom JDK
```

4. If you choose Oracle JDK 1.8 or Oracle JDK 1.7, the JDK you choose downloads and installs automatically on the Ambari Server host. This option requires that you have an internet connection. You must install this JDK on all hosts in the cluster to this same path.
5. If you choose `Custom` JDK, verify or add the custom JDK path on all hosts in the cluster. Use this option if you want to use OpenJDK or do not have an internet connection (and have pre-installed the JDK on all hosts).

6. After setup completes, you must restart each component for the new JDK to be used by the Hadoop services.
7. Using the Ambari Web UI, do the following tasks:
 - Restart each component
 - Restart all services

For more information about managing services in your cluster, see [Managing Services](#).

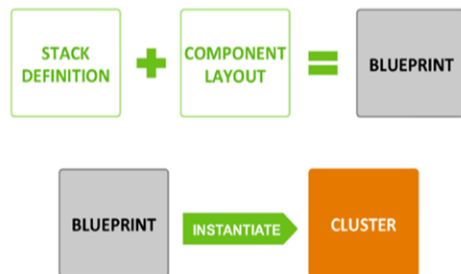


Important

You **must** also update your JCE security policy files on the Ambari Server and all hosts in the cluster **to match the new JDK version**. If you are running Kerberos and do not update the JCE to match the JDK, you will have issues starting services. Refer to [Hortonworks Data Platform Apache Ambari Security](#) for more information on [Installing the JCE](#).

12. Using Ambari Blueprints

Ambari Blueprints provide an API to perform cluster installations. You can build a reusable “blueprint” that defines which Stack to use, how Service Components should be laid out across a cluster and what configurations to set.



After setting up a blueprint, you can call the API to instantiate the cluster by providing the list of hosts to use. The Ambari Blueprint framework promotes reusability and facilitates automating cluster installations without UI interaction.

Learn more about Ambari Blueprints API on the [Ambari Wiki](#).

13. Tuning Ambari Performance

For clusters larger than 200 nodes, consider the following tuning options:

1. Calculate the new, larger cache size, using the following relationship:

```
ecCacheSizeValue=60*<cluster_size>
```

where <cluster_size> is the number of nodes in the cluster.

2. On the Ambari Server host, in `/etc/ambari-server/conf/ambari-properties`, add the following property and value:

```
server.ecCacheSize=<ecCacheSizeValue>
```

where <ecCacheSizeValue> is the value calculated previously, based on the number of nodes in the cluster.

3. Add the following properties to adjust the JDBC connection pool settings:

```
server.jdbc.connection-pool.acquisition-size=5
```

```
server.jdbc.connection-pool.max-age=0
```

```
server.jdbc.connection-pool.max-idle-time=14400
```

```
server.jdbc.connection-pool.max-idle-time-excess=0
```

```
server.jdbc.connection-pool.idle-test-interval=7200
```

4. If using MySQL as the Ambari database, in your MySQL configuration, increase the `wait_timeout` and `interactive_timeout` to 8 hours (28800) and max. connections from 32 to 128.



Important

It is **critical** that the Ambari configuration for `server.jdbc.connection-pool.max-idle-time` and `server.jdbc.connection-pool.idle-test-interval` must be lower than the MySQL `wait_timeout` and `interactive_timeout` set on the MySQL side. If you choose to decrease these timeout values, adjust `downserver.jdbc.connection-pool.max-idle-time` and `server.jdbc.connection-pool.idle-test-interval` accordingly in the Ambari configuration so that they are less than `wait_timeout` and `interactive_timeout`.

5. Restart Ambari Server.

```
ambari-server restart
```

6. If you are using the Ambari Metrics service, you might want to consider switching from the default embedded mode to distributed mode, as well as other tuning options. See [Performance Tuning](#) for more information.

14. Customizing Ambari Log + PID Directories

Ambari Server and Agents write log activity output to log files and use a PID-file that contains the process identification number (PID) for their running process. The log files and PID-file are found in following default locations (on their respective hosts):

	LOG file location	PID file location
Ambari Server	/var/log/ambari-server/ambari-server.log	/var/run/ambari-server/ambari-server.pid
Ambari Agent	/var/log/ambari-agent/ambari-agent.log	/var/run/ambari-agent/ambari-agent.pid

You can configure the logging level for `ambari-server.log` by modifying `/etc/ambari-server/conf/log4j.properties` on the Ambari Server host. For the Ambari Agents, you can set the `loglevel` in `/etc/ambari-agent/conf/ambari-agent.ini` on each host running an Ambari Agent. In general, you should also consider setting log-rotate policies for your systems. Refer to your operating system documentation and <http://linuxconfig.org/logrotate-8-manual-page> for more information on setting up log-rotate in your environment.

You can also modify these locations. Use the following instructions:

- [Customizing Ambari Server Log + PID Directories \[44\]](#)
- [Customizing Ambari Agent Log + PID Directories \[45\]](#)

14.1. Customizing Ambari Server Log + PID Directories

Use the following instructions to modify the Ambari Server Log and PID locations.

1. On the Ambari Server host, stop the Ambari Server:

```
ambari-server stop
```

2. To modify the PID location, edit the Ambari Server properties file:

```
vi /etc/ambari-server/conf/ambari.properties
```

3. Modify the `pid.dir` property and save the file:

```
pid.dir=/var/run/ambari-server
```

4. To modify the Log location, edit the Ambari Server `log4j` file:

```
vi /etc/ambari-server/conf/log4j.properties
```

5. Modify the `ambari.log.dir` property and save the file:


```
ambari.log.dir=${ambari.root.dir}/var/log/ambari-server
```

6. Create the new directories and be sure to set the directory ownership and permissions to allow the Ambari Server process access.

7. Restart the Ambari Server.

```
ambari-server start
```

14.2. Customizing Ambari Agent Log + PID Directories

Use the following instructions to modify the Ambari Agent Log and PID locations.

1. On each host running an Ambari Agent, stop the Ambari Agent:

```
ambari-agent stop
```

2. Edit the Ambari Agent properties file:

```
vi /etc/ambari-agent/conf/ambari-agent.ini
```

3. In the [agent] section, modify the piddir and logdir properties:

```
[agent]
```

```
logdir=/var/log/ambari-agent
```

```
piddir=/var/run/ambari-agent
```

4. Save the file.

5. Create the new directories and be sure to set the directory ownership and permissions to allow the Ambari Agent process access.

6. Restart the Ambari Agent.

```
ambari-agent start
```