

# Hortonworks Data Platform

## Ambari Reference Guide

(Aug 28, 2015)

## Hortonworks Data Platform : Ambari Reference Guide

Copyright © 2012-2015 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 .....	4
2. Customizing HDP Services .....	5
2.1. Defining Service Users and Groups for a HDP 2.x Stack .....	5
2.2. Setting Properties That Depend on Service Usernames/Groups .....	6
3. Configuring Storm for Supervision .....	8
4. Using Custom Host Names .....	10
4.1. How to Customize the name of a host .....	10
5. Moving the Ambari Server .....	12
5.1. Back up Current Data .....	12
5.2. Update all Agents .....	12
5.3. Install the New Ambari Server .....	13
6. Configuring LZO Compression .....	15
6.1. Configure core-site.xml for LZO .....	15
6.2. Running Compression with Hive Queries .....	15
6.2.1. Create LZO Files .....	16
6.2.2. Write Custom Java to Create LZO Files .....	16
7. Using Non-Default Databases .....	17
7.1. Using Non-Default Databases - Ambari .....	17
7.1.1. Using Ambari with Oracle .....	17
7.1.2. Using Ambari with MySQL .....	18
7.1.3. Using Ambari with PostgreSQL .....	19
7.1.4. Troubleshooting Non-Default Databases with Ambari .....	20
7.2. Using Non-Default Databases - Hive .....	22
7.2.1. Using Hive with Oracle .....	22
7.2.2. Using Hive with MySQL .....	23
7.2.3. Using Hive with PostgreSQL .....	25
7.2.4. Troubleshooting Non-Default Databases with Hive .....	26
7.3. Using Non-Default Databases - Oozie .....	28
7.3.1. Using Oozie with Oracle .....	28
7.3.2. Using Oozie with MySQL .....	28
7.3.3. Using Oozie with PostgreSQL .....	29
7.3.4. Troubleshooting Non-Default Databases with Oozie .....	30
8. Setting up an Internet Proxy Server for Ambari .....	32
8.1. How To Set Up an Internet Proxy Server for Ambari .....	32
9. Configuring Network Port Numbers .....	33
9.1. Default Network Port Numbers - Ambari .....	33
9.2. Optional: Changing the Default Ambari Server Port .....	33
10. Change the JDK Version .....	35
11. Using Ambari Blueprints .....	37
12. Configuring HDP Stack Repositories for Red Hat Satellite .....	38
12.1. How To Configure HDP Stack Repositories for Red Hat Satellite .....	38
13. Tuning Ambari Performance .....	40
14. Configuring and Tuning Ambari Metrics .....	41
14.1. Enabling HBase Region and Table Metrics .....	41
14.2. Setting Metrics Directories for Embedded Mode .....	41
14.3. Configuring Ambari Metrics for Distributed Mode .....	42

15. Moving the Ambari Metrics Collector ..... 43

# 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](#)
2. [Install Ambari Agents](#)

## 1.1. Download the Ambari Repo

Select the OS family running on your installation host.

### 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/2.1.1/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.2.1.1-2.x	Ambari 2.x	8
base	CentOS-6 - Base	6,518
extras	CentOS-6 - Extras	37
updates	CentOS-6 - Updates	785

4. Install the Ambari bits. This also installs the default PostgreSQL Ambari database.

```
yum install ambari-server
```

5. Enter `y` when prompted to to confirm transaction and dependency checks.

A successful installation displays output similar to the following:

```
Installing : postgresql-libs-8.4.20-1.el6_5.x86_64 1/4
Installing : postgresql-8.4.20-1.el6_5.x86_64 2/4
Installing : postgresql-server-8.4.20-1.el6_5.x86_64 3/4
Installing : ambari-server-2.1.1-147.noarch 4/4
Verifying : postgresql-server-8.4.20-1.el6_5.x86_64 1/4
Verifying : postgresql-libs-8.4.20-1.el6_5.x86_64 2/4
Verifying : postgresql-8.4.20-1.el6_5.x86_64 3/4
Verifying : ambari-server-2.1.1-147.noarch 4/4
Installed : ambari-server.noarch-2.1.1-147.noarch

Dependency
Installed :

postgresql.x86_64 0:8.4.20-1.el6_5
postgresql-libs.x86_64 0:8.4.20-1.el6_5
postgresql-server.x86_64 0:8.4.20-1.el6_5

Complete!
```



### 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
```

## SLES 11

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/2.1.1/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
AMBARI.2.1.1-2.x	Ambari 2.x	Yes
http-demeter.uni-regensburg.de-c997c8f9	SUSE-Linux-Enterprise-Software-Development-Kit-11-SP1 11.1.1-1.57	Yes
opensuse	OpenSuse	Yes

4. Install the Ambari bits. This also installs PostgreSQL.

```
zypper install ambari-server
```

5. Enter `y` when prompted to confirm transaction and dependency checks.

A successful installation displays output similar to the following:

```
Retrieving package postgresql-libs-8.3.5-1.12.x86_64 (1/4), 172.0 KiB (571.0 KiB
unpacked)
```

```
Retrieving: postgresql-libs-8.3.5-1.12.x86_64.rpm [done (47.3 KiB/s)]
```

```
Installing: postgresql-libs-8.3.5-1.12 [done]
```

```
Retrieving package postgresql-8.3.5-1.12.x86_64 (2/4), 1.0 MiB (4.2 MiB unpacked)
```

```
Retrieving: postgresql-8.3.5-1.12.x86_64.rpm [done (148.8 KiB/s)]
```

```
Installing: postgresql-8.3.5-1.12 [done]
```

```
Retrieving package postgresql-server-8.3.5-1.12.x86_64 (3/4), 3.0 MiB (12.6 MiB
unpacked)
```

```
Retrieving: postgresql-server-8.3.5-1.12.x86_64.rpm [done (452.5 KiB/s)]
```

```
Installing: postgresql-server-8.3.5-1.12 [done]
```

```
Updating etc/sysconfig/postgresql...
```

```
Retrieving package ambari-server-2.1.1-59.noarch (4/4), 99.0 MiB (126.3 MiB unpacked)
```

```
Retrieving: ambari-server-2.1.1-59.noarch.rpm [done (3.0 MiB/s)]
```

```
Installing: ambari-server-2.1.1-59 [done]
```

```
ambari-server 0:off 1:off 2:off 3:on 4:off 5:on 6:off
```

## 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 11

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.



## 2. Customizing HDP Services

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

### 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  $\geq 1000$ .

#### Service Users

Service*	Component	Default User Account
Accumulo	Accumulo Tracer, Accumulo Monitor, Accumulo GC, Accumulo Master	accumulo (HDP 2.2 and 2.3 Stack)
Ambari Metrics	Metrics Collector, Metrics Monitor	ams
Atlas	Atlas Metadata Server	atlas (HDP 2.3 Stack)
Falcon	Falcon Server	falcon (HDP 2.1, 2.2 and 2.3 Stack)
Flume	Flume Agents	flume
HBase	MasterServer RegionServer	hbase
HDFS	NameNode SecondaryNameNode DataNode	hdfs
Hive	Hive Metastore, HiveServer2	hive
Kafka	Kafka Broker	kafka
Knox	Knox Gateway	knox

Service*	Component	Default User Account
Mahout	Mahout clients	mahout (HDP 2.2 and 2.3 Stack)
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 (Ranger is available with HDP 2.2 Stack)
Ranger KMS	Ranger KMS Server	kms (HDP 2.3 Stack)
Spark	Spark History Server	spark (Spark is available with HDP 2.2 Stack)
Sqoop	Sqoop	sqoop
Storm	Masters (Nimbus, DRPC Server, Storm REST API, Server, Storm UI Server) Slaves (Supervisors, Logviewers)	storm (HDP 2.1, 2.2 and 2.3 Stack)
Tez	Tez clients	tez (HDP 2.1, 2.2 and 2.3 Stack)
WebHCat	WebHCat Server	hcat
YARN	NodeManager ResourceManager	yarn
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"

Property Name	Value
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. Configuring Storm for Supervision

If you have installed a cluster with HDP 2.2 Stack that includes the Storm service, you can configure the Storm components to operate under supervision. This section describes those steps:

1. Stop all Storm components.

Using Ambari Web, browse to `Services > Storm > Service Actions`, choose `Stop`. Wait until the Storm service stop completes.

2. Stop Ambari Server.

```
ambari-server stop
```

3. Change Supervisor and Nimbus command scripts in the Stack definition.

On Ambari Server host, run:

```
sed -ir "s/scripts\/supervisor.py/scripts\/supervisor_prod.py/g" /var/lib/ambari-server/resources/common-services/STORM/0.9.1.2.1/metainfo.xml  
sed -ir "s/scripts\/nimbus.py/scripts\/nimbus_prod.py/g" /var/lib/ambari-server/resources/common-services/STORM/0.9.1.2.1/metainfo.xml
```

4. Install supervisord on all Nimbus and Supervisor hosts.

- Install EPEL repository.

```
yum install epel-release -y
```

- Install supervisor package for supervisord.

```
yum install supervisor -y
```

- Enable supervisord on autostart.

```
chkconfig supervisord on
```

- Change supervisord configuration file permissions.

```
chmod 600 /etc/supervisord.conf
```

5. Configure supervisord to supervise Nimbus Server and Supervisors by appending the following to `/etc/supervisord.conf` on all Supervisor host and Nimbus hosts accordingly.

```
[program:storm-nimbus]  
command=env PATH=$PATH:/bin:/usr/bin:/usr/jdk64/jdk1.7.0_67/bin/ JAVA_HOME=  
/usr/jdk64/jdk1.7.0_67 /usr/hdp/current/storm-nimbus/bin/storm nimbus  
user=storm  
autostart=true  
autorestart=true  
startsecs=10  
startretries=999  
log_stdout=true
```

```
log_stderr=true
logfile=/var/log/storm/nimbus.out
logfile_maxbytes=20MB
logfile_backups=10

[program:storm-supervisor]
command=env PATH=$PATH:/bin:/usr/bin:/usr/jdk64/jdk1.7.0_67/bin/ JAVA_HOME=
/usr/jdk64/jdk1.7.0_67 /usr/hdp/current/storm-supervisor/bin/storm
supervisor
user=storm
autostart=true
autorestart=true
startsecs=10
startretries=999
log_stdout=true
log_stderr=true
logfile=/var/log/storm/supervisor.out
logfile_maxbytes=20MB
logfile_backups=10
```



### Note

Change `/usr/jdk64/jdk1.7.0_67` accordingly to the location of the JDK being used by Ambari in your environment.

6. Start Supervisor service on all Supervisor and Nimbus hosts.

```
service supervisord start
```

7. Start Ambari Server.

```
ambari-server start
```

8. Start all the other Storm components.

Using Ambari Web, browse to `Services > Storm > Service Actions`, choose `Start`.

## 4. 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](#)

### 4.1. 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 <ambari_hostname>
```

where `<ambari_hostname>` is the host name to use for Agent registration.

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 name of your custom echo 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. Be sure to `chmod` the script so it is executable by the Agent.

```
#!/bin/sh <hostname> -f
```

where `<hostname>` is the host name to use for Agent registration.

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
```

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
```

## 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. 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](#).

### 6.1. 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.

### 6.2. 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](#)

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

## 6.2.1. 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;
hive.exec.compress.output=true;SET
mapreduce.output.fileoutputformat.compress=true;"
```

## 6.2.2. 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>"
```

## 7. 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](#)
- [Using Non-Default Databases - Hive](#)
- [Using Non-Default Databases - Oozie](#)

### 7.1. 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](#)
- [Using Ambari with MySQL](#)
- [Using Ambari with PostgreSQL](#)
- [Troubleshooting Non-Default Databases with Ambari](#)

#### 7.1.1. 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. Select Oracle Database 11g Release 2 - `ojdbc6.jar`.
  - c. Copy the .jar file to the Java share directory.

```
cp ojdbc6.jar /usr/share/java
```
  - d. Make sure the .jar file has the appropriate permissions - 644.
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.

## 7.1.2. Using Ambari with MySQL

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
```
  - 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 database using the schema script.

```
mysql -u <AMBARIUSER> -p

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 and enter the credentials you defined in Step 2. for user name, password and database name.

## 7.1.3. 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.

## 7.1.4. Troubleshooting Non-Default Databases with Ambari

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

### 7.1.4.1. 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.



### 7.1.4.1.1. 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](#).

### 7.1.4.2. 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.

#### 7.1.4.2.1. 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
```

### 7.1.4.3. 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.

#### 7.1.4.3.1. Solution

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

### 7.1.4.4. 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.

#### 7.1.4.4.1. 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](#).

## 7.2. 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](#)
- [Using Hive with MySQL](#)
- [Using Hive with PostgreSQL](#)
- [Troubleshooting Non-Default Databases with Hive](#)

### 7.2.1. 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. Select Oracle Database 11g Release 2 - ojdbc6.jar and download the file.
- c. Make sure the .jar file has the appropriate permissions - 644.
- d. Execute the following command, adding the path to the downloaded .jar file:

```
ambari-server setup --jdbc-db=oracle --jdbc-driver=/path/to/downloaded/ojdbc6.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.

3. Load the Hive database schema.

- For **HDP 2.2 or later Stacks**



## Important

Ambari sets up the Hive Metastore database schema automatically.

You do not need to pre-load the Hive Metastore database schema into your Oracle database for a HDP 2.2 Stack.

- For a **HDP 2.1 Stack**

You must pre-load the Hive database schema into your Oracle database using the schema script, as follows: `sqlplus <HIVEUSER>/<HIVEPASSWORD> <hive-schema-0.13.0.oracle.sql`

Find the `hive-schema-0.13.0.oracle.sql` file in the `/var/lib/ambari-server/resources/stacks/HDP/2.1/services/HIVE/etc/` directory of the Ambari Server host after you have installed Ambari Server.

- For a **HDP 2.0 Stack**

You must pre-load the Hive database schema into your Oracle database using the schema script, as follows: `sqlplus <HIVEUSER>/<HIVEPASSWORD> <hive-schema-0.12.0.oracle.sql`

Find the `hive-schema-0.12.0.oracle.sql` file in the `/var/lib/ambari-server/resources/stacks/HDP/2.0.6/services/HIVE/etc/` directory of the Ambari Server host after you have installed Ambari Server.

- For a **HDP 1.3 Stack**

You must pre-load the Hive database schema into your Oracle database using the schema script, as follows: `sqlplus <HIVEUSER>/<HIVEPASSWORD> <hive-schema-0.10.0.oracle.sql`

Find the `hive-schema-0.10.0.oracle.sql` file in the `/var/lib/ambari-server/resources/stacks/HDP/1.3.2/services/HIVE/etc/` directory of the Ambari Server host after you have installed Ambari Server.

## 7.2.2. 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*
```

- 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.

4. Load the Hive database schema.

- For a **HDP 2.2 or later Stacks**:



## Important

Ambari sets up the Hive Metastore database schema automatically.

You do not need to pre-load the Hive Metastore database schema into your MySQL database for a HDP 2.2 Stack.

- For **HDP 2.1 Stack**:

You must pre-load the Hive database schema into your MySQL database using the schema script, as follows. `mysql -u root -p <HIVEDATABASE> hive-schema-0.13.0.mysql.sql`

Find the `hive-schema-0.13.0.mysql.sql` file in the `/var/lib/ambari-server/resources/stacks/HDP/2.1/services/HIVE/etc/` directory of the Ambari Server host after you have installed Ambari Server.

## 7.2.3. 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
```

- 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.

### 3. Load the Hive database schema.

- For a **HDP 2.2 or later Stacks**:



#### Important

Ambari sets up the Hive Metastore database schema automatically.

You do not need to pre-load the Hive Metastore database schema into your PostgreSQL database for a HDP 2.2 Stack.

- For **HDP 2.1 Stack**:

You must pre-load the Hive database schema into your PostgreSQL database using the schema script, as follows:

```
# psql -U <HIVEUSER> -d <HIVEDATABASE> \connect <HIVEDATABASE>;  
\i hive-schema-0.13.0.postgres.sql;
```

Find the `hive-schema-0.13.0.postgres.sql` file in the `/var/lib/ambari-server/resources/stacks/HDP/2.1/services/HIVE/etc/` directory of the Ambari Server host after you have installed Ambari Server.

- For **HDP 2.0 Stack**:

You must pre-load the Hive database schema into your PostgreSQL database using the schema script, as follows:

```
# sudo -u postgres psql \connect <HIVEDATABASE>; \i hive-  
schema-0.12.0.postgres.sql;
```

Find the `hive-schema-0.12.0.postgres.sql` file in the `/var/lib/ambari-server/resources/stacks/HDP/2.0.6/services/HIVE/etc/` directory of the Ambari Server host after you have installed Ambari Server.

## 7.2.4. Troubleshooting Non-Default Databases with Hive

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

### 7.2.4.1. 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.

### 7.2.4.1.1. Solution

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

### 7.2.4.2. 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.

### 7.2.4.2.1. 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.

### 7.2.4.3. 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.

### 7.2.4.3.1. 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](#) in [Hive Authorization](#) and the HCatalog document [Storage Based Authorization](#).

#### 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>

[Metastore Server Security](#)[Hive Authorization](#)[Storage Based Authorization](#)

## 7.3. 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](#)
- [Using Oozie with MySQL](#)
- [Using Oozie with PostgreSQL](#)
- [Troubleshooting Non-Default Databases with Oozie](#)

### 7.3.1. Using Oozie with Oracle

To set up Oracle for use with Oozie:

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. Select Oracle Database 11g Release 2 - ojdbc6.jar.
  - c. Make sure the .jar file has the appropriate permissions - 644.
  - d. Execute the following command, adding the path to the downloaded.jar file:

```
ambari-server setup --jdbc-db=oracle --jdbc-driver=/path/to/  
downloaded/ojdbc6.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.

### 7.3.2. 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*
```

- 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.

### 7.3.3. 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
```

- 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.

## 7.3.4. Troubleshooting Non-Default Databases with Oozie

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

### 7.3.4.1. 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.

#### 7.3.4.1.1. Solution

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

### 7.3.4.2. 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.

#### 7.3.4.2.1. 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
```

## 8. Setting up an Internet Proxy Server for Ambari

If you plan to use the public repositories for installing the Stack, Ambari Server must have Internet access to confirm access to the repositories and validate the repositories. If your machine requires use of a proxy server for Internet access, you must configure Ambari Server to use the proxy server.

[How To Set Up an Internet Proxy Server for Ambari](#)

### 8.1. How To Set Up an Internet Proxy Server for Ambari

1. On the Ambari Server host, add proxy settings to the following script: `/var/lib/ambari-server/ambari-env.sh`.  
`-Dhttp.proxyHost=<yourProxyHost> -Dhttp.proxyPort=<yourProxyPort>`
2. 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>`
3. If your proxy server requires authentication, add the user name and password, as follows:  
`-Dhttp.proxyUser=<username> -Dhttp.proxyPassword=<password>`
4. Restart the Ambari Server to pick up this change.

If you plan to use local repositories, see [Using a Local Repository](#). Configuring Ambari to use a proxy server and have Internet access is not required. The Ambari Server must have access to your local repositories.

## 9. 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](#)
- [Optional: Changing the Default Ambari Server Port](#)

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

### 9.1. 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 <a href="#">Optional: Change the Ambari Server Port</a> for instructions on changing the default port.	http See <a href="#">Configure Ambari Server for Authenticated HTTP</a> 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	

### 9.2. 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. 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	JDK 1.7 or JDK 1.8
HDP 2.0, 2.1 or 2.2	JDK 1.7



### Important

If you plan to upgrade between Stacks (for example, go from HDP 2.2 -> HDP 2.3) **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.
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).

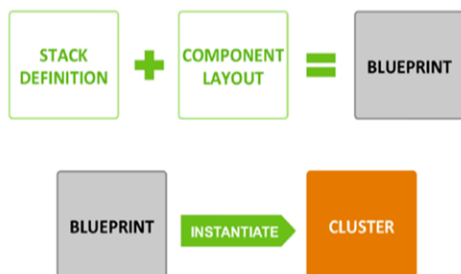
6. After setup completes, you must restart each component for the new JDK to be used by the Hadoop services.
7. Using the Ambari Web UI, do the following tasks:
  - Restart each component
  - Restart each host
  - Restart all services

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



# 11. 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](#).

## 12. Configuring HDP Stack Repositories for Red Hat Satellite

As part of installing HDP Stack with Ambari, `HDP.repo` and `HDP-UTILS.repo` files are generated and distributed to the cluster hosts based on the Base URL user input from the Cluster Install Wizard during the Select Stack step. In cases where you are using Red Hat Satellite to manage your Linux infrastructure, you can disable the repositories defined in the HDP Stack `.repo` files and instead leverage Red Hat Satellite.

[How To Configure HDP Stack Repositories for Red Hat Satellite](#)

### 12.1. How To Configure HDP Stack Repositories for Red Hat Satellite

To disable the repositories created and distributed by Ambari in the `.repo` files:

1. **Before installing your cluster**, on the Ambari Server host, browse to the following:

```
/var/lib/ambari-server/resources/stacks/HDP/2.0.6/configuration
```

2. Modify the `cluster-env.xml` file

```
vi cluster-env.xml
```

3. Search for the property name `repo_suse_rhel_template` (for RHEL/ CentOS /SLES) or `repo_ubuntu_template` (for Ubuntu). For example:

```
<name>repo_suse_rhel_template</name>
```

4. Modify the `<value>` of the property by setting `"enabled=0"`. **The `.repo` files will still be created and distributed by Ambari Server but the repositories will not be enabled.**

5. Save and exit.

6. Restart the Ambari Server and proceed with your cluster install.



#### Important

You must configure Red Hat Satellite to define and enable the Stack repositories. Please refer to the Red Hat Satellite documentation for more information.



#### Important

When using RedHat Satellite or Spacewalk for repository management, it's very important to note that the name of the repository must match our HDP naming convention. When adding repositories, please ensure their names match those found in the `.repo` files. Example: `HDP-2.2.4.2`



## Note

If you need to modify the template after cluster install, you can modify the property on cluster-env configuration using the Ambari REST API or the `/var/lib/ambari-server/resources/scripts/configs.sh` script. The example below shows setting the `repo_suse_rhel_template` property to "enabled=0" using the script.

```
configs.sh -u admin -p admin set ambari.server cluster.name
cluster-env repo_suse_rhel_template "[{{repo_id}}-DISABLED]\nname=
{{repo_id}}-DISABLED\n{% if mirror_list %}mirrorlist={{mirror_list}}
{% else %}baseurl={{base_url}}{% endif %}\n\ncpath=/\nenabled=0\
ngpgcheck=0"
```

where **ambari.server** is the Ambari Server hostname and **cluster.name** is the name of your cluster.

## 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, increase the timeout in MySQL from 15 minutes to 8 hours and the max connections from 32 to 128.

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 [Configuring and Tuning Ambari Metrics](#) for more information.

# 14. Configuring and Tuning Ambari Metrics

- [Enabling HBase Region and Table Metrics](#)
- [Setting Metrics Directories for Embedded Mode](#)
- [Configuring Ambari Metrics for Distributed Mode](#)

## 14.1. Enabling HBase Region and Table Metrics

By default, Ambari Metrics does not collect metrics related to HBase regions, tables and RegionServers. These metrics can be numerous and can cause performance issues.

If you want these metrics to be collected by Ambari, you can do the following. It is **highly recommended** that you test turning on this option and confirm your AMS performance is acceptable.

1. On the Ambari Server, browse to:

```
/var/lib/ambari-server/resources/common-services/  
HBASE/0.96.0.2.0/package/templates
```

2. Edit the following template files:

```
hadoop-metrics2-hbase.properties-GANGLIA-MASTER.j2
```

```
hadoop-metrics2-hbase.properties-GANGLIA-RS.j2
```

3. Edit the `hadoop-metrics2.properties.j2` template file.

4. Comment out (or remove) the following lines:

```
*.source.filter.class=org.apache.hadoop.metrics2.filter.GlobFilter  
  
hbase.*.source.filter.exclude=*Regions*
```

5. Save the template files and restart Ambari Server for the changes to take effect.



### Important

If you upgrade Ambari to a newer version, you will need to re-apply this change to the template file.

## 14.2. Setting Metrics Directories for Embedded Mode

By default, the Ambari Metrics Collector runs in **embedded** mode. The Collector will capture and write metrics to the local file system on the host where the Collector is

running. When configured for **distributed** mode, Metric Collector writes go to HDFS of the cluster.



### Note

Refer to [Configuring Ambari Metrics for Distributed Mode](#) for more information on switching to **distributed** mode.

When running in **embedded** mode, you should confirm the "hbase.rootdir" and "hbase.tmp.dir" directory configurations in **Ambari Metrics > Configs > Advanced ams-hbase-site** are using a large partition, such as:

```
file:///grid/0/var/lib/ambari-metrics-collector/hbase
```

## 14.3. Configuring Ambari Metrics for Distributed Mode

The following section describes how to switch the Ambari Metrics Collector from embedded mode to distributed mode.

In **Ambari Web**, browse to **Ambari Metrics > Configs**, make the following changes, then restart the Collector.

Section	Property	Value
General	Metrics Service operation mode (timeline.metrics.service.operation.mode)	distributed
Advanced ams-hbase-site	hbase.cluster.distributed	true
Advanced ams-hbase-site	hbase.rootdir *	hdfs://c6403.ambari.apache.org:8020/ amsmetrics

\* If using NameNode HA, set the hbase.rootdir value to `hdfs://hdfsnameservice/amsmetrics`

## 15. Moving the Ambari Metrics Collector

For clusters larger than 200 nodes, calculate and set a larger task cache size on the Ambari server.

1. In **Ambari Web**, stop the **Ambari Metrics** service.
2. Execute the following API call to delete the **current** Metric Collector component.

```
curl -u admin:admin -H "X-Requested-By:ambari" -i -X
DELETE http://ambari.server:8080/api/v1/clusters/cluster.name/
hosts/metrics.collector.hostname/host_components/METRICS_COLLECTOR
```

where **ambari.server** is the Ambari Server host, **cluster.name** is your Cluster Name, and **metrics.collector.hostname** is the host running the Metrics Collector.

3. Execute the following API call to add Metrics Collector to a new host.

```
curl -u admin:admin -H "X-Requested-By:ambari" -i -X
POST http://ambari.server:8080/api/v1/clusters/cluster.name/
hosts/metrics.collector.hostname/host_components/METRICS_COLLECTOR
```

where **ambari.server** is the Ambari Server host, **cluster.name** is your Cluster Name, and **metrics.collector.hostname** is the host that will run the Metrics Collector.

4. In Ambari Web, go the Host page where you installed the new Metrics Collector. Click to Install the Metrics Collector component from the Host page.
5. On every host, point the Metric Monitors to the new Collector by editing this file: `/etc/ambari-metrics-monitor/conf/metric_monitor.ini`. Change property **metrics\_server** and change the value to the hostname of the new host. The following sed command can be used directly (replace hostnames):

```
sed -i 's/old.metrics.collector.host/
new.metrics.collector.host/' /etc/ambari-metrics-monitor/conf/
metric_monitor.ini
```

6. In **Ambari Web**, start the **Ambari Metrics** service.