Ambari 1.7.0 Install Guide

Hortonworks Data Platform
Ambari 1.7.0
2014-12-02

Copyright

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.**

For more information on Hortonworks technology, Please visit the Hortonworks Data Platform page. 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.

Table of Contents

Installing HDP Using Ambari

This section describes the information and materials you should get ready to install a HDP cluster using Ambari. Ambari provides an end-to-end management and monitoring solution for your HDP cluster. Using the Ambari Web UI and REST APIs, you can deploy, operate, manage configuration changes, and monitor services for all nodes in your cluster from a central point.

- Determine Stack Compatibility
- Meet Minimum System Requirements
- Collect Information
- Prepare the Environment
- Optional: Configure Local Repositories for Ambari

Determine Stack Compatibility

Use this table to determine whether your Ambari and HDP stack versions are compatible.

Ambari	HDP 2.2[1]	HDP 2.1[2]	HDP 2.0[3]	HDP1.3
1.7.0	x	x	x	x
1.6.1		x	x	x
1.6.0		x	x	x
1.5.1		x	x	x
1.5.0			x	x
1.4.4.23			x	x
1.4.3.38			x	x
1.4.2.104			x	x
1.4.1.61			x	x
1.4.1.25			x	x
1.2.5.17				x

For more information about Installing Accumulo, Hue, Knox, Ranger, and Solr services, see Installing HDP Manually.

Meet Minimum System Requirements

To run Hadoop, your system must meet the following minimum requirements:

- Hardware Recommendations
- Operating Systems Requirements
- Browser Requirements
- Software Requirements

---

[1] Ambari 1.7x does not install Accumulo, Hue, Ranger, or Solr services for the HDP 2.2 Stack.

[2] Ambari 1.7x does not install Accumulo, Hue, Knox, or Solr services for the HDP 2.1 Stack.

[3] Ambari 1.7x does not install Hue for the HDP 2.0 Stack.

- JDK Requirements

- Database Requirements

- Recommended Maximum Open File Descriptors

Hardware Recommendations

There is no single hardware requirement set for installing Hadoop.

For more information about hardware components that may affect your installation, see Hardware Recommendations For Apache Hadoop.

Operating Systems Requirements

The following, 64-bit operating systems are supported:

- Red Hat Enterprise Linux (RHEL) v6.x

- Red Hat Enterprise Linux (RHEL) v5.x (deprecated)

- CentOS v6.x

- CentOS v5.x (deprecated)

- Oracle Linux v6.x

- Oracle Linux v5.x (deprecated)

- SUSE Linux Enterprise Server (SLES) v11, SP1 and SP3

- Ubuntu Precise v12.04

> If you plan to install HDP Stack on SLES 11 SP3, be sure to refer to Configuring Repositories in the HDP documentation for the HDP repositories specific for SLES 11 SP3. Or, if you plan to perform a Local Repository install, be sure to use the SLES 11 SP3 repositories.

> The installer pulls many packages from the base OS repositories. If you do not have a complete set of base OS repositories available to all your machines at the time of installation you may run into issues.
>
> If you encounter problems with base OS repositories being unavailable, please contact your system administrator to arrange for these additional repositories to be proxied or mirrored. For more information see Optional: Configure the Local Repositories.

Browser Requirements

The Ambari Install Wizard runs as a browser-based Web application. You must have a machine capable of running a graphical browser to use this tool.
The minimum required browser versions are:

- Windows (Vista, 7)

    - Internet Explorer 9.0

    - Firefox 18

- Google Chrome 26
- Mac OS X (10.6 or later)
    - Firefox 18
    - Safari 5
    - Google Chrome 26
- Linux (RHEL, CentOS, SLES, Oracle Linux, UBUNTU)
    - Firefox 18
    - Google Chrome 26

On any platform, we recommend updating your browser to the latest, stable version.

Software Requirements

On each of your hosts:

- yum and rpm (RHEL/CentOS/Oracle Linux)
- zypper and php_curl (SLES)
- apt (Ubuntu)
- scp, curl, unzip, tar, and wget
- OpenSSL (v1.01, build 16 or later)
- python (v2.6 or later)

> ⚠️ The Python version shipped with SUSE 11, 2.6.0-8.12.2, has a critical bug that may cause the Ambari Agent to fail within the first 24 hours. If you are installing on SUSE 11, please update all your hosts to Python version 2.6.8-0.15.1.

JDK Requirements

The following Java runtime environments are supported:

- Oracle JDK 1.7_67 64-bit (default)
- Oracle JDK 1.6_31 64-bit (DEPRECATED)
- OpenJDK 7 64-bit (not supported on SLES)
  To install OpenJDK 7 for RHEL, run the following command on all hosts:

```
yum install java-1.7.0-openjdk
```

Database Requirements

Ambari requires a relational database to store information about the cluster configuration and topology. If you install HDP Stack with Hive or Oozie, they also require a relational database. The following table outlines these database requirements:

Component	Description

Ambari	By default, will install an instance of PostgreSQL on the Ambari Server host. Optionally, to use an existing instance of PostgreSQL, MySQL or Oracle. For further information, see Using Non-Default Databases for Ambari.
Hive	By default (on RHEL/CentOS/Oracle Linux 6), Ambari will install an instance of MySQL on the Hive Metastore host. Otherwise, you need to use an existing instance of PostgreSQL, MySQL or Oracle. See Using Non-Default Databases for Hive for more information.
Oozie	By default, Ambari will install an instance of Derby on the Oozie Server host. Optionally, to use an existing instance of PostgreSQL, MySQL or Oracle, see Using Non-Default Databases for Oozie for more information.

> For the Ambari database, if you use an existing Oracle database, make sure the Oracle listener runs on a port other than 8080 to avoid conflict with the default Ambari port.

Check the Maximum Open File Descriptors

The recommended maximum number of open file descriptors is 10000, or more.
To check the current value set for the maximum number of open file descriptors, execute the following shell commands on each host:

```
ulimit -Sn
ulimit -Hn
```

Collect Information

Before deploying an HDP cluster, you should collect the following information:

- The fully qualified domain name (FQDN) of each host in your system.
  The Ambari install wizard supports using IP addresses. You can use `hostname -f` to check or verify the FQDN of a host.

> Deploying all HDP components on a single host is possible, but is appropriate only for initial evaluation purposes. Typically, you set up at least three hosts; one master host and two slaves, as a minimum cluster. For more information about deploying HDP components, see the descriptions for aTypical Hadoop Cluster.

- A list of components you want to set up on each host.

- The base directories you want to use as mount points for storing:

    - NameNode data

    - DataNodes data

    - Secondary NameNode data

    - Oozie data

    - MapReduce data (Hadoop version 1.x)

    - YARN data (Hadoop version 2.x)

    - ZooKeeper data, if you install ZooKeeper

- Various log, pid, and db files, depending on your install type

⚠️   You must use base directories that provide persistent storage locations for your HDP components and your Hadoop data. Installing HDP components in locations that may be removed from a host may result in cluster failure or data loss.
For example: Do Not use `/tmp` in a base directory path.

Prepare the Environment

To deploy your Hadoop instance, you need to prepare your deployment environment:

- Check Existing Package Versions

- Set up Password-less SSH

- Set up Service User Accounts

- Enable NTP on the Cluster

- Check DNS

- Configure iptables

- Disable SELinux, PackageKit and Check umask Value

Check Existing Package Versions

During installation, Ambari overwrites current versions of some packages required by Ambari to manage a Hadoop cluster. Package versions other than those that Ambari installs can cause problems running the installer. Remove any package versions that do not match the following ones:

Component - Description	Files and Versions
Ambari Server Database	postgresql 8.4.13-1.el6_3, postgresql-libs 8.4.13-1.el6_3, postgresql-server 8.4.13-1.el6_3
Ambari Agent - Installed on each host in your cluster. Communicates with the Ambari Server to execute commands.	None
Nagios Server - The host that runs the Nagios server.	nagios 3.5.0-99, nagios-devel 3.5.0-99, nagios-www 3.5.0-99, nagios-plugins 1.4.9-1
Ganglia Server - The host that runs the Ganglia Server.	ganglia-gmetad 3.5.0-99, ganglia-devel 3.5.0-99, libganglia 3.5.0-99, ganglia-web 3.5.7-99, rrdtool 1.4.5-1.el6
Ganglia Monitor - Installed on each host in the cluster. Sends metrics data to the Ganglia Collector.	ganglia-gmond 3.5.0-99, libganglia 3.5.0-99

Component - Description	Files and Versions
Ambari Server Database	postgresql 8.3.5-1, postgresql-server 8.3.5-1, postgresql-libs 8.3.5-1
Ambari Agent - Installed on each host in your cluster. Communicates with the Ambari Server to execute commands.	None

Nagios Server - The host that runs the Nagios server.   nagios 3.5.0-99, nagios-devel 3.5.0-99, nagios-www 3.5.0-99, nagios-plugins 1.4.9-1

Ganglia Server - The host that runs the Ganglia Server.   ganglia-gmetad 3.5.0-99 ganglia-devel 3.5.0-99 libganglia 3.5.0-99 ganglia-web 3.5.7-99 rrdtool 1.4.5-4.5.1

Ganglia Monitor - Installed on each host in the cluster. Sends metrics data to the Ganglia Collector.   ganglia-gmond 3.5.0-99, libganglia 3.5.0-99

Component - Description	Files and Versions
Ambari Server Database	libpq5 postgresql postgresql-9.1 postgresql-client-9.1 postgresql-client-common postgresql-common ssl-cert
Ambari Agent - Installed on each host in your cluster. Communicates with the Ambari Server to execute commands.	zlibc_0.9k-4.1_amd64
Nagios Server - The host that runs the Nagios Server.	nagios3
Ganglia Server - The host that runs the Ganglia Server.	gmetad ganglia-webfrontend ganglia-monitor-python rrdcached
Ganglia Monitor - Installed on each host in the cluster. Sends metrics data to the Ganglia Collector.	gmetad ganglia-webfrontend ganglia-monitor-python rrdcached

Component - Description	Files and Versions
Ambari Server Database	libffi 3.0.5-1.el5, python26 2.6.8-2.el5, python26-libs 2.6.8-2.el5, postgresql 8.4.13-1.el6_3, postgresql-libs 8.4.13-1.el6_3, postgresql-server 8.4.13-1.el6_3
Ambari Agent - Installed on each host in your cluster. Communicates with the Ambari Server to execute commands.	libffi 3.0.5-1.el5, python26 2.6.8-2.el5, python26-libs 2.6.8-2.el5
Nagios Server - The host that runs the Nagios server.	nagios 3.5.0-99, nagios-devel 3.5.0-99, nagios-www 3.5.0-99, nagios-plugins 1.4.9-1
Ganglia Server - The host that runs the Ganglia Server.	ganglia-gmetad 3.5.0-99, ganglia-devel 3.5.0-99, libganglia 3.5.0-99, ganglia-web 3.5.7-99, rrdtool 1.4.5-1.el5
Ganglia Monitor - Installed on each host in the cluster. Sends metrics data to the Ganglia Collector.	ganglia-gmond 3.5.0-99, libganglia 3.5.0-99

Set Up Password-less SSH

To have Ambari Server automatically install Ambari Agents on all your cluster hosts, you must set up password-less SSH connections between the Ambari Server host and all other hosts in the cluster. The Ambari Server host uses SSH public key authentication to remotely access and install the Ambari Agent.

> You can choose to manually install the Agents on each cluster host. In this case, you do not need to generate and distribute SSH keys.

1   Generate public and private SSH keys on the Ambari Server host.

```
ssh-keygen
```

2   Copy the SSH Public Key (id_rsa.pub) to the root account on your target hosts.

```
.ssh/id_rsa
.ssh/id_rsa.pub
```

3   Add the SSH Public Key to the authorized_keys file on your target hosts.

```
cat id_rsa.pub >> authorized_keys
```

4   Depending on your version of SSH, you may need to set permissions on the .ssh directory (to 700) and the authorized_keys file in that directory (to 600) on the target hosts.

```
chmod 700 ~/.ssh
chmod 600 ~/.ssh/authorized_keys
```

5   From the Ambari Server, make sure you can connect to each host in the cluster using SSH, without having to enter a password.

```
ssh root@<remote.target.host>
```
where `<remote.target.host>` has the value of each host name in your cluster.

6   If the following warning message displays during your first connection:
```
Are you sure you want to continue connecting (yes/no)?
```
Enter `Yes`.

7   Retain a copy of the SSH Private Key on the machine from which you will run the web-based Ambari Install Wizard.

> It is possible to use a non-root SSH account, if that account can execute `sudo` without entering a password.

Set up Service User Accounts

The Ambari install wizard creates one administrator-level user account for Ambari, admin. The credentials for the admin account are username/password = admin/admin. For more information about creating additional users and groups for your HDP cluster, see Users and Groups Overview in *Managing Users and Groups*.

Each HDP service requires a service user account. The Ambari Install wizard creates new and preserves any existing service user accounts, and uses these accounts when configuring Hadoop services. Service user account creation applies to service user accounts on the local operating system and to LDAP/AD accounts.

For more information about customizing service user accounts for each HDP service, see one of the following topics:

- Customizing Services for HDP 2.x Stack

- Customizing Services for HDP 1.x Stack

Enable NTP on the Cluster and on the Browser Host

The clocks of all the nodes in your cluster and the machine that runs the browser through which you access the Ambari Web interface must be able to synchronize with each other.

Install a network ttime protocol daem on each host:
```
yum install ntpd
```

To check that the NTP service is on, run the following command on each host:
```
chkconfig —list ntpd
```
To turn on the NTP service, run the following command on each host:
```
chkconfig ntpd
```

Check DNS

 All hosts in your system must be configured for both forward and and reverse DNS.

If you are unable to configure DNS in this way, you must edit the /etc/hosts file on every host in your cluster to contain the IP address and Fully Qualified Domain Name of each of your hosts. The following instructions cover a basic /etc/hosts setup for generic Linux hosts. Different versions and flavors of Linux might require slightly different commands. Please refer to the documentation for the operating system(s) deployed in your environment.

Edit the Host File

1   Using a text editor, open the hosts file on every host in your cluster. For example:

```
vi /etc/hosts
```

2   Add a line for each host in your cluster. The line should consist of the IP address and the FQDN.
     For example:

```
1.2.3.4 <fully.qualified.domain.name>
```

⚠   Do **not** remove the following two lines from your hosts file. Removing or editing the
          following lines may cause various programs that require network functionality to fail.

```
127.0.0.1 localhost.localdomain localhost
::1 localhost6.localdomain6 localhost6
```

Set the Hostname

1   Use the "hostname" command to set the hostname on each host in your cluster.
     For example:

```
hostname <fully.qualified.domain.name>
```

2   Confirm that the hostname is set by running the following command:

```
hostname -f
```

This should return the <fully.qualified.domain.name> you just set.

Edit the Network Configuration File

1   Using a text editor, open the network configuration file on every host and set the desired
     network configuration for each host. For example:

```
vi /etc/sysconfig/network
```

2   Modify the HOSTNAME property to set the fully qualified domain name.

```
NETWORKING=yes
NETWORKING_IPV6=yes
HOSTNAME=<fully.qualified.domain.name>
```

Configuring iptables

For Ambari to communicate during setup with the hosts it deploys to and manages, certain ports
must be open and available. The easiest way to do this is to temporarily disable iptables, as follows:

```
chkconfig iptables off
/etc/init.d/iptables stop
```

You can restart iptables after setup is complete. If the security protocols in your environment prevent
disabling iptables, you can proceed with iptables enabled, if all required ports are open and
available. For more information about required ports, see Configuring Network Port Numbers.

Ambari checks whether iptables is running during the Ambari Server setup process. If iptables is
running, a warning displays, reminding you to check that required ports are open and available. The
Host Confirm step in the Cluster Install Wizard also issues a warning for each host that has iptables
running.

Disable SELinux and PackageKit and check the umask Value

1   You must temporarily disable SELinux for the Ambari setup to function.
     On each host in your cluster,

```
setenforce 0
```

> ✏️   **To permanently disable SELinux**
>   set SELINUX=disabled in /etc/selinux/config
>   This ensures that SELinux does not turn itself on after you reboot the machine .

2   On an installation host running RHEL/CentOS with PackageKit installed,
     open /etc/yum/pluginconf.d/refresh-packagekit.conf using a text editor.
     Make the following change: enabled=0

> PackageKit is not enabled by default on SLES or Ubuntu systems. Unless you have specifically enabled PackageKit, you may skip this step for a SLES or Ubuntu installation host.

3   UMASK (User Mask or User file creation MASK) is the default permission or base permission given when a new file or folder is created on a Linux machine. Most Linux distros set 022 as the default umask. For a HDP cluster, make sure that umask is set to 022.
To set umask 022, run the following command as root on all hosts,
`vi /etc/profile`
then, append the following line:
`umask 022`

Using a Local Repository

If your cluster is behind a fire wall that prevents or limits Internet access, you can install Ambari and a Stack using local repositories. This section describes how to:

- Obtain the repositories
- Set up a local repository having:
    - No Internet Access
    - Temporary Internet Access
- Prepare the Ambari repository configuration file

Obtaining the Repositories

This section describes how to obtain:

- Ambari Repositories
- HDP Repositories

Ambari Repositories

If you do not have Internet access for setting up the Ambari repository, use the link appropriate for your OS family to download a tarball that contains the software.

RHEL/CentOS/Oracle Linux 6

```
wget -nv http://public-repo-1.hortonworks.com/ambari/centos6/ambari-
1.7.0-centos6.tar.gz
```

SLES 11

```
wget -nv http://public-repo-1.hortonworks.com/ambari/suse11/ambari-
1.7.0-suse11.tar.gz
```

UBUNTU 12

```
wget -nv http://public-repo-1.hortonworks.com/ambari/ubuntu12/ambari-
1.7.0-ubuntu12.tar.gz
```

RHEL/CentOS/ORACLE Linux 5 (DEPRECATED)

```
wget -nv http://public-repo-1.hortonworks.com/ambari/centos5/ambari-
1.7.0-centos5.tar.gz
```

If you have temporary Internet access for setting up the Ambari repository, use the link appropriate for your OS family to download a repository that contains the software.

RHEL/CentOS/Oracle Linux 6

```
wget -nv http://public-repo-
1.hortonworks.com/ambari/centos6/1.x/updates/1.7.0/ambari.repo -O
/etc/yum.repos.d/ambari.repo
```

SLES 11

```
wget -nv http://public-repo-
1.hortonworks.com/ambari/suse11/1.x/updates/1.7.0/ambari.repo -O
/etc/zypp/repos.d/ambari.repo
```

UBUNTU 12

```
wget -nv http://public-repo-
1.hortonworks.com/ambari/ubuntu12/1.x/updates/1.7.0/ambari.list -O
/etc/apt/sources.list/ambari.repo
```

RHEL/CentOS/ORACLE Linux 5 (DEPRECATED)

```
wget -nv http://public-repo-
1.hortonworks.com/ambari/centos5/1.x/updates/1.7.0/ambari.repo -O
/etc/yum.repos.d/ambari.repo
```

HDP Stack Repositories

If you do not have Internet access to set up the Stack repositories, use the link appropriate for your OS family to download a tarball that contains the HDP Stack version you plan to install.

RHEL/CentOS/Oracle Linux 6

```
wget -nv http://public-repo-1.hortonworks.com/HDP/centos6/HDP-2.2.0.0-
centos6-rpm.tar.gz
```

```
wget -nv http://public-repo-1.hortonworks.com/HDP-UTILS-
1.1.0.20/repos/centos6/HDP-UTILS-1.1.0.20-centos6.tar.gz
```

SLES 11SP3

```
wget -nv http://public-repo-1.hortonworks.com/HDP/suse11sp3/HDP-
2.2.0.0-suse11sp3-rpm.tar.gz
```

```
wget -nv http://public-repo-1.hortonworks.com/HDP-UTILS-
1.1.0.20/repos/suse11sp3/HDP-UTILS-1.1.0.20-suse11sp3.tar.gz
```

UBUNTU 12

```
wget -nv http://public-repo-1.hortonworks.com/HDP/ubuntu12/HDP-2.2.0.0-
ubuntu12-deb.tar.gz

wget -nv http://public-repo-1.hortonworks.com/HDP-UTILS-
1.1.0.20/repos/ubuntu12/HDP-UTILS-1.1.0.20-ubuntu12.tar.gz
```

RHEL/CentOS/ORACLE Linux 5 (DEPRECATED)

```
wget -nv http://public-repo-1.hortonworks.com/HDP/centos5/HDP-2.2.0.0-
centos5-rpm.tar.gz

wget -nv http://public-repo-1.hortonworks.com/HDP-UTILS-
1.1.0.20/repos/centos5/HDP-UTILS-1.1.0.20-centos5.tar.gz
```

RHEL/CentOS/Oracle Linux 6

```
wget -nv http://public-repo-1.hortonworks.com/HDP/centos6/HDP-2.1.5.0-
centos6-rpm.tar.gz
wget -nv http://public-repo-1.hortonworks.com/HDP-UTILS-
1.1.0.17/repos/centos6/HDP-UTILS-1.1.0.17-centos6.tar.gz
```

SLES 11

```
wget -nv http://public-repo-1.hortonworks.com/HDP/sles11sp1/HDP-
2.1.5.0-sles11sp1-rpm.tar.gz
wget -nv http://public-repo-1.hortonworks.com/HDP-UTILS-
1.1.0.17/repos/suse11/HDP-UTILS-1.1.0.17-suse11.tar.gz
```

UBUNTU 12

```
wget -nv http://public-repo-1.hortonworks.com/HDP/ubuntu12/HDP-2.1.5.0-
ubuntu12-tars-tarball.tar.gz
wget -nv http://public-repo-1.hortonworks.com/HDP-UTILS-
1.1.0.18/repos/ubuntu12/hdp.list
```

RHEL/CentOS/Oracle Linux 6

```
wget -nv http://public-repo-1.hortonworks.com/HDP/centos6/HDP-2.0.12.0-
centos6-rpm.tar.gz
wget -nv http://public-repo-1.hortonworks.com/HDP-UTILS-
1.1.0.17/repos/centos6/HDP-UTILS-1.1.0.17-centos6.tar.gz
```

SLES 11

```
wget -nv http://public-repo-1.hortonworks.com/HDP/suse11/HDP-2.0.12.0-
suse11-rpm.tar.gz
wget -nv http://public-repo-1.hortonworks.com/HDP-UTILS-
1.1.0.17/repos/suse11/HDP-UTILS-1.1.0.17-suse11.tar.gz
```

RHEL/CentOS/ORACLE Linux 5 (DEPRECATED)

```
wget -nv http://public-repo-1.hortonworks.com/HDP/centos5/HDP-2.0.12.0-
centos5-rpm.tar.gz
wget -nv http://public-repo-1.hortonworks.com/HDP-UTILS-
1.1.0.17/repos/centos5/HDP-UTILS-1.1.0.17-centos5.tar.gz
```

RHEL/CentOS/Oracle Linux 6

```
wget -nv http://public-repo-1.hortonworks.com/HDP/centos6/HDP-1.3.9.0-
centos6-rpm.tar.gz
wget -nv http://public-repo-1.hortonworks.com/HDP-UTILS-
1.1.0.16/repos/centos6/HDP-UTILS-1.1.0.16-centos6.tar.gz
```

SLES 11

```
wget -nv http://public-repo-1.hortonworks.com/HDP/suse11/HDP-1.3.9.0-
suse11-rpm.tar.gz
wget -nv http://public-repo-1.hortonworks.com/HDP-UTILS-
1.1.0.16/repos/suse11/HDP-UTILS-1.1.0.16-suse11.tar.gz
```

RHEL/CentOS/ORACLE Linux 5 (DEPRECATED)

```
wget -nv http://public-repo-1.hortonworks.com/HDP/centos5/HDP-1.3.9.0-
centos5-rpm.tar.gz
wget -nv http://public-repo-1.hortonworks.com/HDP-UTILS-
1.1.0.16/repos/centos5/HDP-UTILS-1.1.0.16-centos5.tar.gz
```

If you have temporary Internet access for setting up the Stack repositories, use the link appropriate for your OS family to download a repository that contains the HDP Stack version you plan to install.

RHEL/CentOS/Oracle Linux 6

```
wget -nv http://public-repo-
1.hortonworks.com/HDP/centos6/2.x/GA/2.2.0.0/hdp.repo -O
/etc/yum.repos.d/HDP.repo
```

SLES 11SP3

```
wget -nv http://public-repo-
1.hortonworks.com/HDP/suse11sp3/2.x/GA/2.2.0.0/hdp.repo -O
/etc/zypp/repos.d/HDP.repo
```

UBUNTU 12

```
wget -nv http://public-repo-
1.hortonworks.com/HDP/ubuntu12/2.x/GA/2.2.0.0/hdp.list -O
/etc/apt/sources.list.d/HDP.list
```

RHEL/CentOS/ORACLE Linux 5 (DEPRECATED)

```
wget -nv http://public-repo-
1.hortonworks.com/HDP/centos5/2.x/GA/2.2.0.0/hdp.repo -O
/etc/yum.repos.d/HDP.repo
```

RHEL/CentOS/Oracle Linux 6

```
wget -nv http://public-repo-
1.hortonworks.com/HDP/centos6/2.x/updates/2.1.5.0/hdp.repo -O
/etc/yum.repos.d/HDP.repo
```

SLES 11SP3

```
wget -nv http://public-repo-
1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.1.5.0/hdp.repo -O
/etc/zypp/repos.d/HDP.repo
```

UBUNTU 12

```
wget -nv http://public-repo-
1.hortonworks.com/HDP/ubuntu12/2.1.5.0/hdp.list
/etc/apt/sources.list.d/HDP.list
```

RHEL/CentOS/ORACLE Linux 5 (DEPRECATED)

```
wget -nv http://public-repo-
1.hortonworks.com/HDP/centos5/2.x/updates/2.1.5.0/hdp.repo -O
/etc/yum.repos.d/hdp.repo
```

RHEL/CentOS/Oracle Linux 6

```
wget -nv http://public-repo-
1.hortonworks.com/HDP/centos6/2.x/updates/2.0.12.0/hdp.repo -O
/etc/yum.repos.d/HDP.repo
```

SLES 11

```
wget -nv http://public-repo-
1.hortonworks.com/HDP/suse11/2.x/updates/2.0.12.0/hdp.repo -O
/etc/zypp/repos.d/HDP.repo
```

RHEL/CentOS/ORACLE 5 (DEPRECATED)

```
wget -nv http://public-repo-
1.hortonworks.com/HDP/centos5/2.x/updates/2.0.12.0/hdp.repo -O
/etc/yum.repos.d/HDP.repo
```

RHEL/CentOS/Oracle Linux 6

```
wget -nv http://public-repo-
1.hortonworks.com/HDP/centos6/1.x/updates/1.3.7.0/hdp.repo -O
/etc/yum.repos.d/HDP.repo
```

SLES 11

```
wget -nv http://public-repo-
1.hortonworks.com/HDP/suse11/1.x/updates/1.3.7.0/hdp.repo -O
/etc/zypp/repos.d/HDP.repo
```

RHEL/CentOS/ORACLE Linux 5 (DEPRECATED)

```
wget -nv http://public-repo-
1.hortonworks.com/HDP/centos5/1.x/updates/1.3.7.0/hdp.repo -O
/etc/yum.repos.d/HDP.repo
```

Setting Up a Local Repository

Based on your Internet access, choose one of the following options:

- No Internet Access

  This option involves downloading the repository tarball, moving the tarball to the selected mirror server in your cluster, and extracting to create the repository.

- Temporary Internet Access

  This option involves using your temporary Internet access to sync (using reposync) the software packages to your selected mirror server and creating the repository.

  Both options proceed in a similar, straightforward way. Setting up for each option presents some key differences, as described in the following sections:

  - Getting Started Setting Up a Local Repository

  - Setting Up a Local Repository with No Internet Access

  - Setting Up a Local Repository with Temporary Internet Access

Getting Started Setting Up a Local Repository

To get started setting up your local repository, complete the following prerequisites:

- Select an existing server in, or accessible to the cluster, that runs a supported operating system

- Enable network access from all hosts in your cluster to the mirror server

- Ensure the mirror server has a package manager installed such as yum (RHEL / CentOS / Oracle Linux), zypper (SLES), or apt-get (Ubuntu)

- **Optional:** If your repository has temporary Internet access, and you are using RHEL/CentOS/Oracle Linux as your OS, install yum utilities:

  ```
  yum install yum-utils createrepo
  ```

1 Create an HTTP server.

   1 On the mirror server, install an HTTP server (such as Apache httpd) using the instructions provided here .

   2 Activate this web server.

   3 Ensure that any firewall settings allow inbound HTTP access from your cluster nodes to your mirror server.

   If you are using Amazon EC2, make sure that SELinux is disabled.

2 On your mirror server, create a directory for your web server.

- For example, from a shell window, type:

- **For RHEL/CentOS/Oracle Linux:**

```
mkdir -p /var/www/html/
```
- **For SLES:**

```
mkdir -p /srv/www/htdocs/rpms
```
- **For Ubuntu:**

```
mkdir -p /var/www/html/
```

- If you are using a symlink, enable the `followsymlinks` on your web server.

> After you have completed the steps in Getting Started Setting up a Local Repository, move on to specific setup for your repository internet access type.

Setting Up a Local Repository with No Internet Access

After completing the Getting Started Setting up a Local Repository procedure, finish setting up your repository by completing the following steps:

1   Obtain the tarball for the repository you would like to create. For options, see Obtaining the Repositories.

2   Copy the repository tarballs to the web server directory and untar.

  1   Browse to the web server directory you created.

- **For RHEL/CentOS/Oracle Linux:**

```
cd /var/www/html/
```
- **For SLES:**

```
cd /srv/www/htdocs/rpms
```
- **For Ubuntu:**

```
cd /var/www/html/
```

  2   Untar the repository tarballs to the following locations:
      where `<web.server>`, `<web.server.directory>`, `<OS>`, `<version>`, and `<latest.version>` represent the name, home directory, operating system type, version, and most recent release version, respectively.

Repository Content	Repository Location
Ambari Repository	Untar under `<web.server.directory>`
HDP Stack Repositories	Create directory and untar under `<web.server.directory>`/hdp

Table 1.   Untar Locations for a Local Repository - No Internet Access

3   Confirm you can browse to the newly created local repositories.

Repository	URL
Ambari Base URL	http://\<web.server\>/ambari/\<OS\>/2.x/updates/1.7.0
HDP Base URL	http://\<web.server\>/hdp/HDP/\<OS\>/2.x/updates/\<latest.version\>
HDP-UTILS Base URL	http://\<web-server\>/hdp/HDP-UTILS-\<version\>/repos/\<OS\>

Table 2.   URLs for a Local Repository - No Internet Access

⚠️   Be sure to record these Base URLs. You will need them when installing Ambari and the cluster.

4   Optional: If you have multiple repositories configured in your environment, deploy the following plug-in on all the nodes in your cluster.

   1   Install the plug-in.

   •   **For RHEL and CentOS 6:**

```
yum install yum-plugin-priorities
```

   •   **For RHEL and CentOS 5:**

```
yum install yum-priorities
```

   2   **Edit the** `/etc/yum/pluginconf.d/priorities.conf` **file to add the following:**

```
[main]
enabled=1
gpgcheck=0
```

Setting up a Local Repository With Temporary Internet Access

After completing the Getting Started Setting up a Local Repository procedure, finish setting up your repository by completing the following steps:

1   Put the repository configuration files for Ambari and the Stack in place on the host. For options, see Obtaining the Repositories.

2   Confirm availability of the repositories.

        For RHEL/CentOS/Oracle Linux:

```
yum repolist
```

        For SLES:

```
zypper repos
```

        For Ubuntu:

```
dpkg -list
```

3   Synchronize the repository contents to your mirror server.

  • Browse to the web server directory:

    For RHEL/CentOS/Oracle Linux:

```
cd /var/www/html
```
    For SLES:

```
cd /srv/www/htdocs/rpms
```

  • For Ambari, create `ambari` directory and reposync.

```
mkdir -p ambari/<OS>
cd ambari/<OS>
reposync -r Updates-ambari-1.7.0
```

  • For HDP Stack Repositories, create `hdp` directory and reposync.

```
mkdir -p hdp/<OS>
cd hdp/<OS>
reposync -r HDP-<latest.version>
reposync -r HDP-UTILS-<version>
```

4   Generate the repository metadata.

  • For Ambari:

```
createrepo <web.server.directory>/ambari/<OS>/Updates-ambari-
1.7.0
```

  • For HDP Stack Repositories:

```
createrepo <web.server.directory>/hdp/<OS>/HDP-<latest.version>
createrepo <web.server.directory>/hdp/<OS>/HDP-UTILS-<version>
```

5   Confirm that you can browse to the newly created repository.

Repository	URL
Ambari Base URL	http://<web.server>/ambari/<OS>/Updates-ambari-1.7.0
HDP Base URL	http://<web.server>/hdp/<OS>/HDP-<latest.version>
HDP-UTILS Base URL	http://<web.server>/hdp/<OS>/HDP-UTILS-<version>

Table 3.   URLs for the New Repository

> ⚠   Be sure to record these Base URLs. You will need them when installing Ambari and the Cluster.

6   Optional. If you have multiple repositories configured in your environment, deploy the following plug-in on all the nodes in your cluster.

  1   Install the plug-in.

  • **For RHEL and CentOS 6:**

```
yum install yum-plugin-priorities
```

2   Edit the `/etc/yum/pluginconf.d/priorities.conf` file to add the following:

```
[main]
enabled=1
gpgcheck=0
```

Preparing The Ambari Repository Configuration File

1   Download the `ambari.repo` file from the mirror server you created in the preceding sections or from the public repository.

   •   From your mirror server:

   ```
   http://<web_server>/ambari/<OS>/1.x/updates/1.7.0/ambari.repo
   ```

   •   From the public repository:

   ```
   http://public-repo-
   1.hortonworks.com/ambari/<OS>/1.x/updates/1.7.0/ambari.repo
   ```

2   Edit the ambari.repo file using the Ambari repository Base URL obtained when setting up your local repository. Refer to step 3 in Setting Up a Local Repository with No Internet Access, or step 5 in Setting Up a Local Repository with Temporary Internet Access, if necessary.

Repository	URL
Ambari Base URL	http://<web-server>/ambari/<OS>/1.x/updates/1.7.0

Table 4.   Base URLs for a Local Repository

3   If this an Ambari updates release, disable the GA repository definition.

[ambari-1.x]
```
name=Ambari 1.x
baseurl=http://public-repo-1.hortonworks.com/ambari/centos6/1.x/GA
gpgcheck=1
gpgkey=http://public-repo-1.hortonworks.com/ambari/centos6/RPM-GPG-
KEY/RPM-GPG-KEY-Jenkins
 enabled=0
priority=1
```

[Updates-ambari-1.7.0]
```
name=ambari-1.7.0 - Updates
baseurl= <this.is.the.AMBARI.base.url>
gpgcheck=1
gpgkey=http://public-repo-1.hortonworks.com/ambari/centos6/RPM-GPG-
KEY/RPM-GPG-KEY-Jenkins
enabled=1
priority=1
```

4   Place the ambari.repo file on the machine you plan to use for the Ambari Server.

   For RHEL/CentOS/Oracle Linux:

```
/etc/yum.repos.d/ambari.repo
```

For SLES:

```
/etc/zypp/repos.d/ambari.repo
```

For Ubuntu:

```
/etc/apt-get install/list.d/ambari.list
```

Edit the `/etc/yum/pluginconf.d/priorities.conf` file to add the following:

```
[main]
enabled=1
gpgcheck=0
```

5   Proceed to Installing Ambari Server to install and setup Ambari Server.

Download the Ambari Repo

Select one of the following tabs that shows the OS family running on your installation host.

Follow instructions in the section for the operating system that runs on your installation host.
Use a command line editor to perform each instruction.

RHEL/CentOS/Oracle Linux 6

1   Log in to your host as `root`. You may `sudo` as `su` if your environment requires such access.
     For example, type:

```
<username> ssh <hostname.FQDN>
sudo su -
```
where `<username>` is your user name and `<hostname.FQDN>` is the fully qualified domain
name of your server host.

2   Download the Ambari repository file to a directory on your installation host.

```
wget -nv http://public-repo-
1.hortonworks.com/ambari/centos6/1.x/updates/1.7.0/ambari.repo -O
/etc/yum.repos.d/ambari.repo
```

⚠️   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.1.7.0-1.x	Ambari 1.x	5
base	CentOS-6 - Base	6,518
extras	CentOS-6 - Extras	15
updates	CentOS-6 - Updates	209

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-1.7.0-135.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  : ambari-server-1.7.0-135.noarch
3/4
Verifying  : postgresql-8.4.20-1.el6_5.x86_64
4/4

Installed:
  ambari-server.noarch 0:1.7.0-135

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!
```

✎   **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

1   Log in to your host as `root`. You may `sudo` as `su` if your environment requires such access.
    For example, type:

2   `<username> ssh <hostname.FQDN>`
    `sudo su -`
    where `<username>` is your user name and `<hostname.FQDN>` is the fully qualified domain
    name of your server host.

3   Download the Ambari repository file to a directory on your installation host.
    `wget -nv http://public-repo-`
    `1.hortonworks.com/ambari/suse11/1.x/updates/1.7.0/ambari.repo -O`
    `/etc/zypp/repos.d/ambari.repo`

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

4   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.1.7.0-1.x	Ambari 1.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

5   Install the Ambari bits. This also installs PostgreSQL.

    `zypper install ambari-server`

6   Enter `y` when prompted to 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-1.7.0-135.noarch (4/4), 99.0 MiB
(126.3 MiB unpacked)
Retrieving: ambari-server-1.7.0-135.noarch.rpm [done (3.0 MiB/s)]
Installing: ambari-server-1.7.0-135 [done]
ambari-server            0:off  1:off  2:off  3:on   4:off  5:on
6:off
```

UBUNTU 12

1   Log in to your host as `root`. You may `sudo` as `su` if your environment requires such access. For example, type:

2   `<username> ssh <hostname.FQDN>`
`sudo su -`

    where `<username>` is your user name and `<hostname.FQDN>` is the fully qualified domain name of your server host.

3   Download the Ambari repository file to a directory on your installation host.
`wget -nv http://public-repo-`
`1.hortonworks.com/ambari/ubuntu12/1.x/updates/1.7.0/ambari.list -O`
`/etc/apt/sources.list.d/ambari.list`

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

`apt-get update`

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

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

`apt-cache pkgnames`

You should see the Ambari packages in the list.

Version values vary, depending on the installation.

Alias   **Name**
AMBARI-dev-2.x   Ambari 2.x

    5   Install the Ambari bits. This also installs PostgreSQL.

```
apt-get install ambari-server
```

RHEL/CentOS/ORACLE Linux 5 (DEPRECATED)

    1   Log in to your host as `root`. You may `sudo` as `su` if your environment requires such access. For example, type:

    2   `<username> ssh <hostname.FQDN>`
           `sudo su -`
           where `<username>` is your user name and `<hostname.FQDN>` is the fully qualified domain name of your server host.

    3   Download the Ambari repository file to a directory on your installation host.
           `wget -nv http://public-repo-1.hortonworks.com/ambari/centos5/1.x/updates/1.7.0/ambari.repo -O /etc/yum.repos.d/ambari.repo`

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

    4   Confirm the repository is configured by checking the repo list.

```
yum repolist
```
You should see the Ambari repositories in the list.
```
AMBARI.1.7.0-1.x                          |  951 B     00:00
AMBARI.1.7.0-1.x/primary       | 1.6 kB      00:00
AMBARI.1.7.0-1.x                    5/5
epel                                       | 3.7 kB      00:00
epel/primary_db                 | 3.9 MB      00:01
```

repo Id	repo Name	status
AMBARI.1.7.0-1.x	Ambari 1.x	5
base	CentOS-5 - Base	3,667
epel	Extra Packages for Enterprise Linux 5 - x86_64	7,614
puppet	Puppet	433
updates	CentOS-5 - Updates	118

    5   Install the Ambari bits. This also installs PostgreSQL.

```
yum install ambari-server
```

> When deploying HDP on a cluster having limited or no Internet access, you should provide access to the bits using an alternative method.
>
> - For more information about setting up local repositories, see Optional: Configure Local Repositories.
> - For more information about obtaining JCE policy archives for secure authentication, see Deploying JCE Policy Archives on the Ambari Server.
>
> Ambari Server by default uses an embedded PostgreSQL database. When you install the Ambari Server, the PostgreSQL packages and dependencies must be available for install. These packages are typically available as part of your Operating System repositories. Please confirm you have the appropriate repositories available for the postgresql-server packages.

Set Up the Ambari Server

The `ambari-server` command manages the setup process. Run the following command on the Ambari server host:
You may append Setup Options to the command.

```
ambari-server setup
```

Respond to the following prompts:

1   If you have *not* temporarily disabled SELinux, you may get a warning. Accept the default (`y`), and continue.

2   By default, Ambari Server runs under `root`. Accept the default (n) at the **Customize user account for ambari-server daemon** prompt, to proceed as `root`.
    If you want to create a different user to run the Ambari Server, or to assign a previously created user, select `y` at the **Customize user account for ambari-server daemon** prompt, then provide a user name.

3   If you have not temporarily disabled `iptables` you may get a warning. Enter `y` to continue.

4   Select a JDK version to download. Enter 1 to download Oracle JDK 1.7.

> By default, Ambari Server setup downloads and installs Oracle JDK 1.7 and the accompanying Java Cryptography Extension (JCE) Policy Files. If you plan to use a different version of the JDK, see Setup Options for more information.

5   Accept the Oracle JDK license when prompted. You must accept this license to download the necessary JDK from Oracle. The JDK is installed during the deploy phase.

6   Select `n` at **Enter advanced database configuration** to use the default, embedded PostgreSQL database for Ambari. The default PostgreSQL database name is `ambari`. The default user name and password are `ambari/bigdata`.
    Otherwise, to use an existing PostgreSQL, MySQL or Oracle database with Ambari, select `y`.

    - If you are using an existing PostgreSQL, MySQL, or Oracle database instance, use one of the following prompts:

⚠️ You must prepare a non-default database instance, using the steps detailed in Using Non-Default Databases-Ambari, before running setup and entering advanced database configuration.

- To use an existing Oracle 11g r2 instance, and select your own database name, user name, and password for that database, enter 2.

  Select the database you want to use and provide any information requested at the prompts, including host name, port, Service Name or SID, user name, and password.

- To use an existing MySQL 5.x database, and select your own database name, user name, and password for that database, enter 3.

  Select the database you want to use and provide any information requested at the prompts, including host name, port, database name, user name, and password.

- To use an existing PostgreSQL 9.x database, and select your own database name, user name, and password for that database, enter 4.

  Select the database you want to use and provide any information requested at the prompts, including host name, port, database name, user name, and password.

7   At Proceed with configuring remote database connection properties [y/n] choose y.

8   Setup completes.

📝 If your host accesses the Internet through a proxy server, you must configure Ambari Server to use this proxy server. See How to Set Up an Internet Proxy Server for Ambari for more information.

Setup Options

The following table describes options frequently used for Ambari Server setup.

Option	Description
-j (or -- java-home)	Specifies the JAVA_HOME path to use on the Ambari Server and all hosts in the cluster. By default when you do not specify this option, Ambari Server setup downloads the Oracle JDK 1.7 binary and accompanying Java Cryptography Extension (JCE) Policy Files to /var/lib/ambari-server/resources. Ambari Server then installs the JDK to /usr/jdk64. Use this option when you plan to use a JDK other than the default Oracle JDK 1.7. See JDK Requirements for more information on the supported JDKs. If you are using an alternate JDK, you must manually install the JDK on all hosts and specify the Java Home path during Ambari Server setup. If you plan to use Kerberos, you must also install the JCE on all hosts. This path must be valid on all hosts. For example: `ambari-server setup –j /usr/java/default`

--jdbc-driver	Should be the path to the JDBC driver JAR file. Use this option to specify the location of the JDBC driver JAR and to make that JAR available to Ambari Server for distribution to cluster hosts during configuration. Use this option with the --jdbc-db option to specify the database type.
--jdbc-db	Specifies the database type. Valid values are: [postgres | mysql | oracle] Use this option with the --jdbc-driver option to specify the location of the JDBC driver JAR file.
-s (or --silent)	Setup runs silently. Accepts all default prompt values.
-v (or --verbose)	Prints verbose info and warning messages to the console during Setup.
-g (or --debug)	Start Ambari Server in debug mode

Next Steps

Start the Ambari Server

Start the Ambari Server

- Run the following command on the Ambari Server host:

  ```
  ambari-server start
  ```

- To check the Ambari Server processes:

  ```
  ambari-server status
  ```

- To stop the Ambari Server:

  ```
  ambari-server stop
  ```

> ✏️   If you plan to use an existing database instance for Hive/HCatalog or for Oozie, you must complete the preparations described in Using Non-Default Databases-Oozie **before** installing your Hadoop cluster.

Next Steps

Install, configure and deploy an HDP cluster

Install, Configure and Deploy a HDP Cluster

This section describes how to use the Ambari install wizard running in your browser to install, configure, and deploy your cluster.

- Log In to Apache Ambari

- Name Your Cluster

- Select Stack

- Install Options

- Confirm Hosts

- Choose Services

- Assign Masters

- Assign Slaves and Clients

- Customize Services

- Review

- Install, Start and Test

- Complete

Log In to Apache Ambari

After starting the Ambari service, open Ambari Web using a web browser.

1   Point your browser to `http://<your.ambari.server>:8080,` where `<your.ambari.server>` is the name of your ambari server host. For example, a default Ambari server host is located at http://c6401.ambari.apache.org:8080.

2   Log in to the Ambari Server using the default user name/password: admin/admin. You can change these credentials later.

For a new cluster, the Ambari install wizard displays a Welcome page from which you launch the Ambari Install wizard.

Launching the Ambari Install Wizard

From the Ambari Welcome page, choose Launch Install Wizard.

Welcome to Apache Ambari

Provision a cluster, manage who can access the cluster, and customize views for Ambari users.

Create a Cluster

Use the Install Wizard to select services and configure your cluster

Launch Install Wizard

Manage Users + Groups

Manage the users and groups that can access Ambari

Users   **Groups**

Deploy Views

Create view instances and grant permissions

Views

Name Your Cluster

1   In `Name your cluster`, type a name for the cluster you want to create. Use no white spaces or special characters in the name.

2   Choose `Next`.

Select Stack

The Service Stack (the Stack) is a coordinated and tested set of HDP components. Use a radio button to select the Stack version you want to install. To install an HDP 2x stack, select the HDP 2.2, HDP 2.1, or HDP 2.0 radio button.

Select Stack

Please select the service stack that you want to use to install your Hadoop cluster.

Stacks

◉ HDP 2.2
○ HDP 2.1
○ HDP 2.0
○ HDP 1.3

▶ **Advanced Repository Options**

← Back   **Next →**

Expand Advanced Repository Options to select the Base URL of a repository from which Stack software packages download. Ambari sets the default Base URL for each repository, depending on the Internet connectivity available to the Ambari server host, as follows:

- For an Ambari Server host having Internet connectivity, Ambari sets the repository Base URL for the latest patch release for the HDP Stack version. For an Ambari Server having NO Internet connectivity, the repository Base URL defaults to the latest patch release version available at the time of Ambari release.

- You can override the repository Base URL for the HDP Stack with an earlier patch release if you want to install a specific patch release for a given HDP Stack version. For example, the HDP 2.1 Stack will default to the HDP 2.1 Stack patch release 7, or HDP-2.1.7. If you want to install HDP 2.1 Stack patch release 2, or HDP-2.1.2 instead, obtain the Base URL from the HDP Stack documentation, then enter that location in Base URL.

- If you are using a local repository, see Optional: Configure Ambari for Local Repositories for information about configuring a local repository location, then enter that location as the Base URL instead of the default, public-hosted HDP Stack repositories.

▼ **Advanced Repository Options**

Customize the repository Base URLs for downloading the Stack software packages. If your hosts do not have access to the internet, you will have to create a local mirror of the Stack repository that is accessible by all hosts and use those Base URLs here.

Important: When using local mirror repositories, you only need to provide Base URLs for the Operating System you are installing for your Stack. Uncheck all other repositories.

OS	Name	Base URL
☑ redhat5	HDP-2.2	http://public-repo-1.hortonworks.com/HDP/cer ⊗
	HDP-UTILS-1.1.0.20	http://public-repo-1.hortonworks.com/HDP-UT ⊗
☑ redhat6	HDP-2.2	http://public-repo-1.hortonworks.com/HDP/cer ⊗
	HDP-UTILS-1.1.0.20	http://public-repo-1.hortonworks.com/HDP-UT ⊗
☑ suse11	HDP-2.2	http://public-repo-1.hortonworks.com/HDP/sus ⊗
	HDP-UTILS-1.1.0.20	http://public-repo-1.hortonworks.com/HDP-UT ⊗
☑ ubuntu12	HDP-2.2	http://public-repo-1.hortonworks.com/HDP/ubi ⊗
	HDP-UTILS-1.1.0.20	http://public-repo-1.hortonworks.com/HDP-UT ⊗

☐ Skip Repository Base URL validation (Advanced) ❓

← Back   Next →

> ✎ The UI displays repository Base URLs based on Operating System Family (OS Family). Be sure to set the correct OS Family based on the Operating System you are running. The following table maps the OS Family to the Operating Systems.

Table 5.   Operating Systems mapped to each OS Family

OS Family	Operating Systems
redhat6	Red Hat 6, CentOS 6, Oracle Linux 6
suse11	SUSE Linux Enterprise Server 11
ubuntu12	Ubuntu Precise 12.04
redhat5	Red Hat 5, CentOS 5, Oracle Linux 5

Install Options

In order to build up the cluster, the install wizard prompts you for general information about how you want to set it up. You need to supply the FQDN of each of your hosts. The wizard also needs to access the private key file you created in Set Up Password-less SSH. Using the host names and key file information, the wizard can locate, access, and interact securely with all hosts in the cluster.

1   Use the `Target Hosts` text box to enter your list of host names, one per line. You can use ranges inside brackets to indicate larger sets of hosts. For example, for host01.domain through host10.domain use `host[01-10].domain`

> ✎ If you are deploying on EC2, use the **internal Private DNS** host names.

2   If you want to let Ambari automatically install the Ambari Agent on all your hosts using SSH, select **Provide your SSH Private Key** and either use the `Choose File` button in the `Host Registration Information` section to find the private key file that matches the public key you installed earlier on all your hosts or cut and paste the key into the text box manually.

> ✎ If you are using IE 9, the **Choose File** button may not appear. Use the text box to cut and paste your private key manually.

Fill in the user name for the SSH key you have selected. If you do not want to use `root` , you must provide the user name for an account that can execute `sudo` without entering a password.

3   If you do not want Ambari to automatically install the Ambari Agents, select `Perform manual registration`. For further information, see Installing Ambari Agents Manually.

4   Choose `Register and Confirm` to continue.

Confirm Hosts

`Confirm Hosts` prompts you to confirm that Ambari has located the correct hosts for your cluster and to check those hosts to make sure they have the correct directories, packages, and processes required to continue the install.

If any hosts were selected in error, you can remove them by selecting the appropriate checkboxes and clicking the grey **Remove** `Selected` button. To remove a single host, click the small white `Remove` button in the Action column.

At the bottom of the screen, you may notice a yellow box that indicates some warnings were encountered during the check process. For example, your host may have already had a copy of `wget` or `curl`. Choose `Click here to see the warnings` to see a list of what was checked and what caused the warning. The warnings page also provides access to a python script that can help you clear any issues you may encounter and let you run `Rerun Checks`.

> If you are deploying HDP using Ambari 1.4 or later on RHEL 6.5 you will likely see Ambari Agents fail to register with Ambari Server during the "Confirm Hosts" step in the Cluster Install wizard. Click the "Failed" link on the Wizard page to display the Agent logs. The following log entry indicates the SSL connection between the Agent and Server failed during registration:
> ```
>  INFO 2014-04-02 04:25:22,669 NetUtil.py:55
> - Failed to connect to https://<ambari-server>:8440/cert/ca due to
> [Errno 1] _ssl.c:492:
> error:100AE081:elliptic curve
> routines:EC_GROUP_new_by_curve_name:unknown group
> ```
>
> For more information about this issue, see the Ambari Troubleshooting Guide.

When you are satisfied with the list of hosts, choose `Next`.

Choose Services

Based on the Stack chosen during Select Stack, you are presented with the choice of Services to install into the cluster. HDP Stack comprises many services. You may choose to install any other available services now, or to add services later. The install wizard selects all available services for installation by default.

1   Choose `none` to clear all selections, or choose `all` to select all listed services.

2   Choose or clear individual checkboxes to define a set of services to install now.

> To use Ambari for monitoring your cluster, you must select `Nagios` and `Ganglia`. Not selecting these services generates a warning message when you complete this section. If you monitor your cluster using other tools, ignore the warning.

3   After selecting the services to install now, choose `Next`.

Assign Masters

The Ambari install wizard assigns the master components for selected services to appropriate hosts in your cluster and displays the assignments in Assign Masters. The left column shows services and current hosts. The right column shows current master component assignments by host, indicating the number of CPU cores and amount of RAM installed on each host.

1   To change the host assignment for a service, select a host name from the drop-down menu for that service.

2   To remove a ZooKeeper instance, click the green minus icon next to the host address you want to remove.

3   When you are satisfied with the assignments, choose `Next`.

Assign Slaves and Clients

The Ambari installation wizard assigns the slave components (DataNodes, NodeManagers, and RegionServers) to appropriate hosts in your cluster. It also attempts to select hosts for installing the appropriate set of clients.

1   Use **all** or **none** to select all of the hosts in the column or none of the hosts, respectively.

If a host has a red asterisk next to it, that host is also running one or more master components. Hover your mouse over the asterisk to see which master components are on that host.

2   Fine-tune your selections by using the checkboxes next to specific hosts.

✎   As an option you can start the HBase REST server manually after the install process is complete. It can be started on any host that has the HBase Master or the Region Server installed. If you attempt to start it on the same host as the Ambari server, however, you need to start it with the -p option, as its default port is 8080 and that conflicts with the Ambari Web default port.
`/usr/lib/hbase/bin/hbase-daemon.sh start rest -p`
`<custom_port_number>`

3   When you are satisfied with your assignments, choose `Next`.

Customize Services

`Customize Services` presents you with a set of tabs that let you manage configuration settings for HDP components. The wizard sets reasonable defaults for each of the options here, but you can use this set of tabs to tweak those settings. You are strongly encouraged to do so, as your requirements may be slightly different. Pay particular attention to the directories suggested by the installer.

> To prevent out-of-memory errors during the install, at the Customize Services step in the Cluster Install wizard browse to `Hive > hive-site.xml`, then modify the following configuration settings:
>
> | Property Name | Purpose | Default Value | Required Value |
> |---|---|---|---|
> | fs.hdfs.impl.disable.cache | Disable HDFS filesystem cache | false | true |
> | fs.file.impl.disable.cache | Disable local filesystem cache | false | true |

> For the `HDFSServicesConfigsGeneral` configuration property, make sure to enter an integer value, in bytes, that sets the HDFS maximum edit log size for checkpointing. A typical value is 500000000.

Hover your cursor over each of the properties to see a brief description of what it does. The number of tabs you see is based on the type of installation you have decided to do. A typical installation has at least ten groups of configuration properties and other related options, such as database settings for Hive/HCat and Oozie, admin name/password, and alert email for Nagios.

The install wizard sets reasonable defaults for all properties. You must provide database passwords for the Hive, Nagios, and Oozie services, the Master Secret for Knox, and a valid email address to which system alerts will be sent. Select each service that displays a number highlighted red. Then, fill in the required field on the Service Config tab. Repeat this until the red flags disappear.

For example, Choose Hive. Expand the Hive Metastore section, if necessary. In Database Password, provide a password, then retype to confirm it, in the fields marked red and "This is required."

For more information about customizing specific services for a particular HDP Stack, see [Customizing HDP Services](#).

After you complete Customizing Services, choose `Next`.

Review

The assignments you have made are displayed. Check to make sure everything is correct. If you need to make changes, use the left navigation bar to return to the appropriate screen.

To print your information for later reference, choose `Print`.

When you are satisfied with your choices, choose `Deploy`.

Install, Start and Test

The progress of the install displays on the screen. Ambari installs, starts, and runs a simple test on each component. Overall status of the process displays in progress bar at the top of the screen and host-by-host status displays in the main section. Do not refresh your browser during this process. Refreshing the browser may interrupt the progress indicators.

To see specific information on what tasks have been completed per host, click the link in the `Message` column for the appropriate host. In the **Tasks** pop-up, click the individual task to see the related log files. You can select filter conditions by using the `Show` drop-down list. To see a larger version of the log contents, click the `Open` icon or to copy the contents to the clipboard, use the `Copy` icon.

When `Successfully installed and started the services` **appears, choose** `Next`.

Complete

The Summary page provides you a summary list of the accomplished tasks. Choose `Complete`. Ambari Web GUI displays.