

Hortonworks Data Platform

Automated Install with Ambari

(Dec 21, 2015)

Hortonworks Data Platform : Automated Install with Ambari

Copyright © 2012-2015 Hortonworks, Inc. Some 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.



Except where otherwise noted, this document is licensed under
Creative Commons Attribution ShareAlike 3.0 License.
<http://creativecommons.org/licenses/by-sa/3.0/legalcode>

Table of Contents

1. Getting Ready	1
1.1. Determine Stack Compatibility	1
1.2. Meet Minimum System Requirements	1
1.2.1. Operating Systems Requirements	2
1.2.2. Browser Requirements	2
1.2.3. Software Requirements	3
1.2.4. JDK Requirements	3
1.2.5. Database Requirements	4
1.2.6. Memory Requirements	4
1.2.7. Package Size and Inode Count Requirements	5
1.2.8. Check the Maximum Open File Descriptors	5
1.3. Collect Information	6
1.4. Prepare the Environment	6
1.4.1. Set Up Password-less SSH	7
1.4.2. Set Up Service User Accounts	8
1.4.3. Enable NTP on the Cluster and on the Browser Host	8
1.4.4. Check DNS and NSCD	8
1.4.5. Configuring iptables	10
1.4.6. Disable SELinux and PackageKit and check the umask Value	10
1.5. Using a Local Repository	11
1.5.1. Obtaining the Repositories	11
1.5.2. Setting Up a Local Repository	14
1.5.3. Getting Started Setting Up a Local Repository	14
1.5.4. Preparing The Ambari Repository Configuration File	19
2. Installing Ambari	21
2.1. Download the Ambari Repository	21
2.1.1. RHEL/CentOS/Oracle Linux 6	21
2.1.2. RHEL/CentOS/Oracle Linux 7	23
2.1.3. SLES 11	24
2.1.4. Ubuntu 12	26
2.1.5. Ubuntu 14	27
2.1.6. Debian 7	28
2.2. Set Up the Ambari Server	29
2.2.1. Setup Options	31
2.3. Start the Ambari Server	31
3. Installing, Configuring, and Deploying a HDP Cluster	33
3.1. Log In to Apache Ambari	33
3.2. Launching the Ambari Install Wizard	33
3.3. Name Your Cluster	34
3.4. Select Stack	34
3.5. Install Options	35
3.6. Confirm Hosts	36
3.7. Choose Services	37
3.8. Assign Masters	37
3.9. Assign Slaves and Clients	37
3.10. Customize Services	38
3.11. Review	39
3.12. Install, Start and Test	39

3.13. Complete	39
----------------------	----

1. Getting Ready

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](#)

1.1. Determine Stack Compatibility

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

Ambari*	HDP 2.3	HDP 2.2	HDP 2.1 (deprecated)
2.2	x	x	x
2.1	x	x	x
2.0		x	x
1.7		x	x
1.6.1			x
1.6.0			x
1.5.1			x

* Ambari does not install Hue or Solr.

For more information about:

- Installing Hue, see [Installing HDP Manually](#).
- Installing HDP Search (Solr) see [HDP Search](#).
- Installing Ranger, see [Installing Ranger](#).

1.2. Meet Minimum System Requirements

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

- [Operating Systems Requirements](#)
- [Browser Requirements](#)
- [Software Requirements](#)

- [JDK Requirements](#)
- [Database Requirements](#)
- [Recommended Maximum Open File Descriptors](#)

1.2.1. Operating Systems Requirements

The following, 64-bit operating systems are supported:

- Red Hat Enterprise Linux (RHEL) v7.x
- Red Hat Enterprise Linux (RHEL) v6.x
- CentOS v7.x
- CentOS v6.x
- Debian v7.x
- Oracle Linux v7.x
- Oracle Linux v6.x
- SUSE Linux Enterprise Server (SLES) v11 SP3
- Ubuntu Precise v12.04
- Ubuntu Precise v14.04



Note

Ambari can be used to install and manage HDP 2.2 or 2.1 on SLES 11 SP1.



Important

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 [Using a Local Repository](#).

1.2.2. 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 (7, 8)
 - Internet Explorer 10
 - Firefox 18

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

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

1.2.3. Software Requirements

On each of your hosts:

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



Important

- 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.
- Python v2.7.9 or later is **not supported** due to changes in how Python performs certificate validation.

1.2.4. JDK Requirements

The following Java runtime environments are supported:

- Oracle JDK 1.8 64-bit (minimum JDK 1.8_60) (default)
- Oracle JDK 1.7 64-bit (minimum JDK 1.7_67)
- OpenJDK 8 64-bit (not supported on SLES)
- OpenJDK 7 64-bit (not supported on SLES)



Note

JDK support is dependent on your choice of HDP Stack. Refer to Ambari Reference Guide > [Changing the JDK Version](#) for more information.

1.2.5. 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	Databases	Description
Ambari	<ul style="list-style-type: none"> - PostgreSQL 8 - PostgreSQL 9.1.13+, 9.3 - MySQL 5.6 - Oracle 11gr2, 12c 	By default, Ambari 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 - Ambari .
Hive	<ul style="list-style-type: none"> - PostgreSQL 8 - PostgreSQL 9.1.13+, 9.3 - MySQL 5.6 - Oracle 11gr2, 12c 	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 - Hive for more information.
Oozie	<ul style="list-style-type: none"> - PostgreSQL 8 - PostgreSQL 9.1.13+, 9.3 - MySQL 5.6 - Oracle 11gr2, 12c 	<p>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 - Oozie for more information.</p> <p>The default instance of Derby should not be used for a production environment. If you plan to use Derby for a demo, development or test environment, migration of the Oozie database from Derby to a new database is only available in the community.</p>
Ranger	<ul style="list-style-type: none"> - PostgreSQL 9.1.13+, 9.3 - MySQL 5.6 - Oracle 11gr2, 12c 	You must have an existing instance of PostgreSQL, MySQL or Oracle available for Ranger. Refer to Installing Ranger for more information.



Important

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. Alternatively, refer to the Ambari Reference Guide for information on [Changing the Default Ambari Server Port](#).



Important

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

1.2.6. Memory Requirements

The Ambari host should have at least 1 GB RAM, with 500 MB free.

To check available memory on any host, run:

```
free -m
```

If you plan to install the Ambari Metrics Service (AMS) into your cluster, you should review the [Tuning Ambari Metrics](#) section in the [Ambari Reference Guide](#) for guidelines on resources requirements. In general, the host you plan to run the Ambari Metrics Collector host should have the following memory and disk space available based on cluster size:

Number of hosts	Memory Available	Disk Space
1	1024 MB	10 GB
10	1024 MB	20 GB
50	2048 MB	50 GB
100	4096 MB	100 GB
300	4096 MB	100 GB
500	8096 MB	200 GB
1000	12288 MB	200 GB
2000	16384 MB	500 GB



Note

The above is offered as guidelines. Be sure to test for your particular environment. Also refer to [Package Size and Inode Count Requirements](#) for more information on package size and Inode counts.

1.2.7. Package Size and Inode Count Requirements

*Size and Inode values are approximate

	Size	Inodes
Ambari Server	100MB	5,000
Ambari Agent	8MB	1,000
Ambari Metrics Collector	225MB	4,000
Ambari Metrics Monitor	1MB	100
Ambari Metrics Hadoop Sink	8MB	100
After Ambari Server Setup	N/A	4,000
After Ambari Server Start	N/A	500
After Ambari Agent Start	N/A	200

1.2.8. 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
```

If the output is not greater than 10000, run the following command to set it to a suitable default:

```
ulimit -n 10000
```

1.3. 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.



Note

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 a [Typical 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
 - YARN data
 - ZooKeeper data, if you install ZooKeeper
 - Various log, pid, and db files, depending on your install type



Important

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.

1.4. Prepare the Environment

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

- [Set up Password-less SSH](#)
- [Set Up Service User Accounts](#)
- [Enable NTP on the Cluster](#)

- [Check DNS and NSCD](#)
- [Configure iptables](#)
- [Disable SELinux, PackageKit and Check umask Value](#)

1.4.1. 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.



Note

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.



Note

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

1.4.2. Set Up Service User Accounts

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 [Defining Service Users and Groups for a HDP 2.x Stack](#).

1.4.3. 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.

To check that the NTP service will be automatically started upon boot, run the following command on each host:

RHEL/CentOS/Oracle 6

```
chkconfig --list ntpd
```

RHEL/CentOS/Oracle 7

```
systemctl is-enabled ntpd
```

To set the NTP service to auto-start on boot, run the following command on each host:

RHEL/CentOS/Oracle 6

```
chkconfig ntpd on
```

RHEL/CentOS/Oracle 7

```
systemctl enable ntpd
```

To start the NTP service, run the following command on each host:

RHEL/CentOS/Oracle 6

```
service ntpd start
```

RHEL/CentOS/Oracle 7

```
systemctl start ntpd
```

1.4.4. Check DNS and NSCD

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

If you are unable to configure DNS in this way, you should 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 are provided as an overview and cover a basic network setup for generic Linux hosts. Different versions and flavors of Linux might require

slightly different commands and procedures. Please refer to the documentation for the operating system(s) deployed in your environment.

Hadoop relies heavily on DNS, and as such performs many DNS lookups during normal operation. To reduce the load on your DNS infrastructure, it's highly recommended to use the Name Service Caching Daemon (NSCD) on cluster nodes running Linux. This daemon will cache host, user, and group lookups and provide better resolution performance, and reduced load on DNS infrastructure.

1.4.4.1. 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>
```



Important

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
```

1.4.4.2. Set the Hostname

1. Confirm that the hostname is set by running the following command:

```
hostname -f
```

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

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

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

1.4.4.3. 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
```

```
HOSTNAME=<fully.qualified.domain.name>
```

1.4.5. 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:

RHEL/CentOS/Oracle Linux 6

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

RHEL/CentOS/Oracle Linux 7

```
systemctl disable firewalld  
  
service firewalld 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.

1.4.6. Disable SELinux and PackageKit and check the umask Value

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

```
setenforce 0
```



Note

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
```



Note

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

3. UMASK (User Mask or User file creation MASK) sets the default permissions or base permissions granted when a new file or folder is created on a Linux machine. Most Linux distros set 022 as the default umask value. A umask value of 022 grants read, write, execute permissions of 755 for new files or folders. A umask value of 027 grants read, write, execute permissions of 750 for new files or folders. Ambari supports a umask value of 022 or 027. For example, to set the umask value to 022, run the following command as root on all hosts, `vi /etc/profile` then, append the following line:
`umask 022`

1.5. Using a Local Repository

Local repositories are frequently used in enterprise clusters that have limited outbound internet access. In these scenarios, having packages available locally provides more governance, and better installation performance. These repositories are used heavily during installation for package distribution, as well as post-install for routine cluster operations such as service start/restart operations. The following section describes the steps required to setup and use a local repository:

- [Obtain the repositories](#)
- Set up a local repository having:
 - [No Internet Access](#)
 - [Temporary Internet Access](#)
- [Prepare the Ambari repository configuration file](#)

1.5.1. Obtaining the Repositories

This section describes how to obtain:

- [Ambari Repositories](#)
- [HDP Stack Repositories](#)

1.5.1.1. Ambari Repositories

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

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

Ambari 2.2 Repositories

OS	Format	URL
RedHat 6	Base URL	http://public-repo-1.hortonworks.com/ambari/centos6/2.x/updates/2.2.0.0
CentOS 6	Repo File	http://public-repo-1.hortonworks.com/ambari/centos6/2.x/updates/2.2.0.0/ambari.repo
	Tarball md5 asc	http://public-repo-1.hortonworks.com/ambari/centos6/2.x/updates/2.2.0.0/ambari-2.2.0.0-centos6.tar.gz
RedHat 7	Base URL	http://public-repo-1.hortonworks.com/ambari/centos7/2.x/updates/2.2.0.0
	Repo File	http://public-repo-1.hortonworks.com/ambari/centos7/2.x/updates/2.2.0.0/ambari.repo

CentOS 7 Oracle Linux 7	Tarball md5 asc	http://public-repo-1.hortonworks.com/ambari/centos7/2.x/updates/2.2.0.0/ambari-2.2.0.0-centos7.tar.gz
SLES 11	Base URL	http://public-repo-1.hortonworks.com/ambari/suse11/2.x/updates/2.2.0.0
	Repo File	http://public-repo-1.hortonworks.com/ambari/suse11/2.x/updates/2.2.0.0/ambari.repo
	Tarball md5 asc	http://public-repo-1.hortonworks.com/ambari/suse11/2.x/updates/2.2.0.0/ambari-2.2.0.0-suse11.tar.gz
Ubuntu 12	Base URL	http://public-repo-1.hortonworks.com/ambari/ubuntu12/2.x/updates/2.2.0.0
	Repo File	http://public-repo-1.hortonworks.com/ambari/ubuntu12/2.x/updates/2.2.0.0/ambari.list
	Tarball md5 asc	http://public-repo-1.hortonworks.com/ambari/ubuntu12/2.x/updates/2.2.0.0/ambari-2.2.0.0-ubuntu12.tar.gz
Ubuntu 14	Base URL	http://public-repo-1.hortonworks.com/ambari/ubuntu14/2.x/updates/2.2.0.0
	Repo File	http://public-repo-1.hortonworks.com/ambari/ubuntu14/2.x/updates/2.2.0.0/ambari.list
	Tarball md5 asc	http://public-repo-1.hortonworks.com/ambari/ubuntu14/2.x/updates/2.2.0.0/ambari-2.2.0.0-ubuntu14.tar.gz
Debian 7	Base URL	http://public-repo-1.hortonworks.com/ambari/debian7/2.x/updates/2.2.0.0
	Repo File	http://public-repo-1.hortonworks.com/ambari/debian7/2.x/updates/2.2.0.0/ambari.list
	Tarball md5 asc	http://public-repo-1.hortonworks.com/ambari/debian7/2.x/updates/2.2.0.0/ambari-2.2.0.0-debian7.tar.gz

1.5.1.2. HDP Stack Repositories

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

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

HDP 2.3 Repositories

OS	Version Number	Repository Name	Format	URL
RedHat 6 CentOS 6 Oracle Linux 6	HDP-2.3.4.0	HDP	Base URL	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/
			Repo File	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/
			Tarball md5 asc	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/HDP-2.3.4.0-centos6-rpm.tar.gz
		HDP-UTILS	Base URL	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/
RedHat 7 CentOS 7 Oracle Linux 7	HDP-2.3.4.0	HDP	Tarball md5 asc	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/UTILS-1.1.0.20-centos6.tar.gz
			Base URL	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/
			Repo File	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/
			Tarball md5 asc	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/HDP-2.3.4.0-centos7-rpm.tar.gz
SLES 11	HDP-2.3.4.0	HDP	Base URL	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/
			Repo File	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/
			Tarball md5 asc	http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/HDP-2.3.4.0-centos7-rpm.tar.gz
		HDP-UTILS	Base URL	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/

			Tarball md5 asc	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/UTILS-1.1.0.20-suse11sp3.tar.gz
Ubuntu 12	HDP-2.3.4.0	HDP	Base URL	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.2.8/ubuntu12-rpm.tar.gz
			Repo File	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.2.8/ubuntu12-rpm.tar.gz
			Tarball md5 asc	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.2.8/HDP-2.3.4.0-ubuntu12-deb.tar.gz
		HDP-UTILS	Base URL	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/UTILS-1.1.0.20-ubuntu12.tar.gz
			Tarball md5 asc	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/UTILS-1.1.0.20-ubuntu12.tar.gz
Ubuntu 14	HDP-2.3.4.0	HDP	Base URL	http://public-repo-1.hortonworks.com/HDP/ubuntu14/2.x/updates/2.2.8/ubuntu14-rpm.tar.gz
			Repo File	http://public-repo-1.hortonworks.com/HDP/ubuntu14/2.x/updates/2.2.8/ubuntu14-rpm.tar.gz
			Tarball md5 asc	http://public-repo-1.hortonworks.com/HDP/ubuntu14/2.x/updates/2.2.8/HDP-2.3.4.0-ubuntu14-deb.tar.gz
		HDP-UTILS	Base URL	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/UTILS-1.1.0.20-ubuntu14.tar.gz
			Tarball md5 asc	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/UTILS-1.1.0.20-ubuntu14.tar.gz
Debian7	HDP-2.3.4.0	HDP	Base URL	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.2.8/debian7-rpm.tar.gz
			Repo File	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.2.8/debian7-rpm.tar.gz
			Tarball md5 asc	http://public-repo-1.hortonworks.com/HDP/debian7/2.x/updates/2.2.8/HDP-2.3.4.0-debian7-deb.tar.gz
		HDP-UTILS	Base URL	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/UTILS-1.1.0.20-debian7.tar.gz
			Tarball md5 asc	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/UTILS-1.1.0.20-debian7.tar.gz

HDP 2.2 Repositories

OS	Version Number	Repository Name	Format	URL
RedHat 6 CentOS 6 Oracle Linux 6	HDP-2.2.8.0	HDP	Base URL	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.2.8/centos6-rpm.tar.gz
			Repo File	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.2.8/centos6-rpm.tar.gz
			Tarball	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.2.8/centos6-rpm.tar.gz
		HDP-UTILS	Base URL	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6.tar.gz
			Tarball	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6.tar.gz
SLES 11	HDP-2.2.8.0	HDP	Base URL	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.2.8/suse11sp3-rpm.tar.gz
			Repo File	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.2.8/suse11sp3-rpm.tar.gz
			Tarball	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.2.8/suse11sp3-rpm.tar.gz
		HDP-UTILS	Base URL	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/suse11sp3.tar.gz
			Tarball	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/suse11sp3.tar.gz
Ubuntu 12	HDP-2.2.8.0	HDP	Base URL	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.2.8/ubuntu12-rpm.tar.gz
			Repo File	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.2.8/ubuntu12-rpm.tar.gz
			Tarball	http://public-repo-1.hortonworks.com/HDP/ubuntu12/2.x/updates/2.2.8/ubuntu12-rpm.tar.gz
		HDP-UTILS	Base URL	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/ubuntu12.tar.gz
			Tarball	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/ubuntu12.tar.gz

HDP 2.1 Repositories

OS	Version Number	Repository Name	Format	URL
RedHat 6 CentOS 6 Oracle Linux 6	HDP-2.1.0.0	HDP	Base URL	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.1.2
			Repo File	http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.1.2
			Tarball	http://public-repo-1.hortonworks.com/HDP/centos6/HDP-2.1.20.0-cento
		HDP-UTILS	Base URL	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.19/repos/cento
SLES 11	HDP-2.1.0.0	HDP	Tarball	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.19/repos/cento
			UTILS-1.1.0.19-centos6.tar.gz	
			Base URL	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.1
			Repo File	http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.1
		HDP-UTILS	Tarball	http://public-repo-1.hortonworks.com/HDP/suse11sp3/HDP-2.1.20.0-su
			rpm.tar.gz	
		HDP-UTILS	Base URL	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.19/repos/suse1
			Tarball	http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.19/repos/suse1
			UTILS-1.1.0.19-suse11sp3.tar.gz	

1.5.2. 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](#)

1.5.3. 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 (Debian/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.

- a. On the mirror server, install an HTTP server (such as Apache httpd) using the instructions provided [here](#).
- b. Activate this web server.
- c. Ensure that any firewall settings allow inbound HTTP access from your cluster nodes to your mirror server.



Note

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 Debian/Ubuntu:**

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

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



Note

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.

1.5.3.1. 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.
 - a. 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 Debian/Ubuntu:**

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

- b. 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.

Untar Locations for a Local Repository - No Internet Access

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

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

URLs for a Local Repository - No Internet Access

Repository	Base URL
Ambari Base URL	http://<web.server>/ambari-2.2.0.0/<OS>
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>

where <web.server> = FQDN of the web server host, and <OS> is centos6, centos7, sles11, ubuntu12, ubuntu14, or debian7.



Important

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.

- Install the plug-in.

- **For RHEL and CentOS 7:**

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

- **For RHEL and CentOS 6:**

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

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

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

1.5.3.2. 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 Debian/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 Debian/Ubuntu:**

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

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

```
mkdir -p ambari/<OS>
```

```
cd ambari/<OS>
```

```
reposync -r Updates-ambari-2.2.0.0
```

where `<OS>` is `centos6`, `centos7`, `sles11`, `ubuntu12`, `ubuntu14`, or `debian7`.

- 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-2.2.0.0
```

- 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.

URLs for the New Repository

Repository	Base URL
Ambari Base URL	<code>http://<web.server>/ambari/<OS>/Updates-ambari-2.2.0.0</code>
HDP Base URL	<code>http://<web.server>/hdp/<OS>/HDP-<latest.version></code>
HDP-UTILS Base URL	<code>http://<web.server>/hdp/<OS>/HDP-UTILS-<version></code>

where <web.server> = FQDN of the web server host, and <OS> is centos6, centos7, sles11, ubuntu12, ubuntu14, or debian7.



Important

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.

- Install the plug-in.

- For RHEL and CentOS 7:

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

- For RHEL and CentOS 6:

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

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

```
[main]
```

```
enabled=1
```

```
gpgcheck=0
```

1.5.4. Preparing The Ambari Repository Configuration File

1. Download the `ambari.repo` file from the public repository.

```
http://public-repo-1.hortonworks.com/ambari/<OS>/2.x/updates/2.2.0.0/ambari.repo
```

where `<OS>` is `centos6`, `centos7`, `sles11`, `ubuntu12`, `ubuntu14`, or `debian7`.

2. Edit the `ambari.repo` file and replace the Ambari Base URL `baseurl` 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.



Note

You can disable the GPG check by setting `gpgcheck=0`. Alternatively, you can keep the check enabled but replace the `gpgkey` with the URL to the GPG-KEY in your local repository.

```
[Updates-ambari-2.2.0.0]
```

```
name=ambari-2.2.0.0-Updates
```

```
baseurl=INSERT-BASE-URL
```

```
gpgcheck=1
```

```
gpgkey=http://public-repo-1.hortonworks.com/ambari/centos6/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
```

```
enabled=1
```

```
priority=1
```

Base URL for a Local Repository

Local Repository	Base URL
Built with Repository Tarball (No Internet Access)	<code>http://<web.server>/ambari-2.2.0.0/<OS></code>
Built with Repository File (Temporary Internet Access)	<code>http://<web.server>/ambari/<OS>/Updates-ambari-2.2.0.0</code>

where `<web.server>` = FQDN of the web server host, and `<OS>` is `centos6`, `centos7`, `sles11`, `ubuntu12`, `ubuntu14`, or `debian7`.

3. 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 Debian/Ubuntu:**

```
/etc/apt/sources.list.d/ambari.list
```

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

```
[main]
```

```
enabled=1
```

```
gpgcheck=0
```

4. Proceed to [Installing Ambari Server](#) to install and setup Ambari Server.

2. Installing Ambari

To install Ambari server on a single host in your cluster, complete the following steps:

1. [Download the Ambari repository](#)
2. [Set Up the Ambari Server](#)
3. [Start the Ambari Server](#)

2.1. Download the Ambari Repository

Follow the instructions in the section for the operating system that runs your installation host.

- [RHEL/CentOS/Oracle Linux 6](#)
- [RHEL/CentOS/Oracle Linux 7](#)
- [SLES 11](#)
- [Ubuntu 12](#)
- [Ubuntu 14](#)
- [Debian 7](#)

Use a command line editor to perform each instruction.

2.1.1. 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/2.2.0.0/ambari.repo -O /etc/yum.repos.d/ambari.repo
```



Important

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

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

```
yum repolist
```

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

Version values vary, depending on the installation.

repo id	repo name	status
AMBARI.2.2.0.0-2.x	Ambari 2.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-8.4.20-3.el6_6.x86_64      1/4
Installing : postgresql-8.4.20-3.el6_6.x86_64         2/4
Installing : postgresql-server-8.4.20-3.el6_6.x86_64   3/4
Installing : ambari-server-2.2.0.0-1470.x86_64         4/4
Verifying  : ambari-server-2.2.0.0-1470.x86_64        1/4
Verifying  : postgresql-8.4.20-3.el6_6.x86_64         2/4
Verifying  : postgresql-server-8.4.20-3.el6_6.x86_64  3/4
Verifying  : postgresql-libs-8.4.20-3.el6_6.x86_64    4/4

Installed:
  ambari-server.x86_64 0:2.2.0.0-1470

Dependency Installed:
  postgresql.x86_64 0:8.4.20-3.el6_6
  postgresql-libs.x86_64 0:8.4.20-3.el6_6
  postgresql-server.x86_64 0:8.4.20-3.el6_6
```

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
```



Note

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 [Using a Local Repository](#).

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.

2.1.2. RHEL/CentOS/Oracle Linux 7

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

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

```
wget -nv http://public-repo-1.hortonworks.com/ambari/centos7/2.x/updates/2.2.0.0/ambari.repo -O /etc/yum.repos.d/ambari.repo
```



Important

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

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

```
yum repolist
```

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

Version values vary, depending on the installation.

repo id	repo name	status
AMBARI.2.2.0.0-2.x	Ambari 2.x	5
base	CentOS-7 - Base	6,518
extras	CentOS-7 - Extras	15
updates	CentOS-7 - 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-8.4.20-3.el6_6.x86_64      1/4
Installing : postgresql-8.4.20-3.el6_6.x86_64          2/4
Installing : postgresql-server-8.4.20-3.el6_6.x86_64    3/4
Installing : ambari-server-2.2.0.0-1470.x86_64          4/4
Verifying  : ambari-server-2.2.0.0-1470.x86_64          1/4
Verifying  : postgresql-8.4.20-3.el6_6.x86_64          2/4
Verifying  : postgresql-server-8.4.20-3.el6_6.x86_64    3/4
Verifying  : postgresql-libs-8.4.20-3.el6_6.x86_64      4/4

Installed:
  ambari-server.x86_64 0:2.2.0.0-1470
```

```
Dependency Installed:
postgresql.x86_64 0:8.4.20-3.el6_6
postgresql-libs.x86_64 0:8.4.20-3.el6_6
postgresql-server.x86_64 0:8.4.20-3.el6_6
```

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
```



Note

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 [Using a Local Repository](#).

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.

2.1.3. 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/2.2.0.0/ambari.repo -O /etc/zypp/repos.d/ambari.repo
```



Important

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

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

```
zypper repos
```

You should see the Ambari repositories in the list.

Version values vary, depending on the installation.

Alias	Name	Enabled	Refresh
AMBARI.2.2.0.0-2.x	Ambari 2.x	Yes	No
http-demeter.uni-regensburg.de-c997c8f9	SUSE-Linux-Enterprise-Software-Development-Kit-11-SP1 11.1.1-1.57	Yes	Yes
opensuse	OpenSuse	Yes	Yes

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

```
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-2.2.0.0-135.noarch (4/4), 99.0 MiB (126.3 MiB unpacked)
```

```
Retrieving: ambari-server-2.2.0.0-135.noarch.rpm [done (3.0 MiB/s)]
```

```
Installing: ambari-server-2.2.0.0-135 [done]
```

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



Note

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 [Using a Local Repository](#).

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.

2.1.4. Ubuntu 12

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

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

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

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

```
apt-get update
```



Important

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

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

```
apt-cache pkgnames ambari
```

You should see the Ambari packages in the list.

Version values vary, depending on the installation.

Name
ambari-agent
ambari-server
ambari-metrics-assembly

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

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



Note

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 [Using a Local Repository](#).

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.

2.1.5. Ubuntu 14

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

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

```
wget -nv http://public-repo-1.hortonworks.com/ambari/
ubuntu14/2.x/updates/2.2.0.0/ambari.list -O /etc/apt/
sources.list.d/ambari.list
```

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

```
apt-get update
```



Important

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

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

```
apt-cache pkgnames ambari
```

You should see the Ambari packages in the list.

Version values vary, depending on the installation.

Name
ambari-agent
ambari-server
ambari-metrics-assembly

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

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



Note

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 [Using a Local Repository](#).

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.

2.1.6. Debian 7

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

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

```
wget -nv http://public-repo-1.hortonworks.com/ambari/debian7/2.x/updates/2.2.0.0/ambari.list -O /etc/apt/sources.list.d/ambari.list
```

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

```
apt-get update
```



Important

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

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

```
apt-cache pkgnames ambari
```

You should see the Ambari packages in the list.

Version values vary, depending on the installation.

Name
ambari-agent
ambari-server
ambari-metrics-assembly

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

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



Note

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 [Using a Local Repository](#).

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.

2.2. Set Up the Ambari Server

Before starting the Ambari Server, you **must** set up the Ambari Server. Setup configures Ambari to talk to the Ambari database, installs the JDK and allows you to customize the user account the Ambari Server daemon will run as. The `ambari-server setup` command manages the setup process. Run the following command on the Ambari server host to start the setup process. You may also append [Setup Options](#) to the command.

```
ambari-server setup
```

Respond to the setup prompt:

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. Refer to the Ambari Security Guide > [Configuring Ambari for Non-Root](#), for more information about running the Ambari Server as non-root.
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 1.8.



Note

JDK support depends entirely on your choice of HDP Stack versions. Please refer to the [Ambari Reference Guide](#) to see which JDK versions are supported by the HDP Stack version you intend to install. By default, Ambari

Server setup downloads and installs Oracle JDK 1.8 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:



Important

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.



Important

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

- To use an existing Oracle 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 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 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.



Note

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.

2.2.1. Setup Options

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

Option	Description
-j (or --java-home)	<p>Specifies the JAVA_HOME path to use on the Ambari Server and all hosts in the cluster. By default when you specify this option, Ambari Server setup downloads the Oracle JDK 1.8 binary and accompanying Java Cryptography Extension (JCE) Policy Files to /var/lib/ambari-server/resources. Ambari Server then installs the JDK to /usr/jdk64.</p> <p>Use this option when you plan to use a JDK other than the default Oracle JDK 1.8. See JDK Requirements for 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 JCE on all hosts.</p> <p>This path must be valid on all hosts. For example:</p> <pre>ambari-server setup -j /usr/java/default</pre>
-jdbc-driver	Should be the path to the JDBC driver JAR file. Use this option to specify the location of the JDBC driver JAR file 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)	<p>Setup runs silently. Accepts all the default prompt values, such as:</p> <ul style="list-style-type: none"> User account "root" for the ambari-server Oracle 1.8 JDK (which is installed at /usr/jdk64). This can be overridden by adding the -j option and specifying an existing JDK path. Embedded PostgreSQL for Ambari DB (with database name "ambari") <p>If you want to run the Ambari Server as non-root, you must run setup in interactive mode. When prompted to customize the ambari-server user account, provide the account information. Refer to Configuring Ambari Server for Non-Root for more information.</p>
-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](#)

2.3. 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
```



Note

If you plan to use an existing database instance for Hive or for Oozie, you must complete the preparations described in [Using Non-Default Databases-Hive](#) and [Using Non-Default Databases-Oozie](#) **before** installing your Hadoop cluster.

Next Steps

[Install, configure and deploy an HDP cluster](#)

3. Installing, Configuring, and Deploying a HDP Cluster

Use the Ambari Install Wizard running in your browser to install, configure, and deploy your cluster, as follows:

- [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](#)

3.1. 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](#).

3.2. 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](#)

3.3. 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`.

3.4. 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 2.3, HDP 2.2, or HDP 2.1 radio button.

Select Stack

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

Stacks

☒ HDP 2.3
☐ HDP 2.2
☐ HDP 2.1

[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 [Using a Local Repository](#) 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
<input checked="" type="checkbox"/> redhat6	HDP-2.3	<input type="text" value="http://s3.amazonaws.com/dev.hortonworks.com/HDP/centos6/2.x/"/>
	HDP-UTILS-1.1.0.20	<input type="text" value="http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/c"/>
<input checked="" type="checkbox"/> redhat7	HDP-2.3	<input type="text" value="http://s3.amazonaws.com/dev.hortonworks.com/HDP/centos7/2.x/"/>
	HDP-UTILS-1.1.0.20	<input type="text" value="http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/c"/>
<input checked="" type="checkbox"/> suse11	HDP-2.3	<input type="text" value="http://s3.amazonaws.com/dev.hortonworks.com/HDP/suse11sp3/2"/>
	HDP-UTILS-1.1.0.20	<input type="text" value="http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/s"/>

☐ Skip Repository Base URL validation (Advanced)

[< Back](#) [Next >](#)



Note

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.

Operating Systems mapped to each OS Family

OS Family	Operating Systems
redhat7	Red Hat 7, CentOS 7, Oracle Linux 7
redhat6	Red Hat 6, CentOS 6, Oracle Linux 6
sles11	SUSE Linux Enterprise Server 11
ubuntu12	Ubuntu 12
ubuntu14	Ubuntu 14
debian7	Debian 7

3.5. 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`



Note

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.



Note

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.

3.6. 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`.



Note

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**.

3.7. 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.
3. After selecting the services to install now, choose **Next**.

3.8. 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**.

3.9. 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 an 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.
3. When you are satisfied with your assignments, choose **Next**.

3.10. Customize Services

The Customize Services step presents you with a set of tabs that let you review and modify your HDP cluster setup. The wizard attempts to set reasonable defaults for each of the options. You are **strongly encouraged** to review these settings as your requirements might be slightly different.

Browse through each service tab and by hovering your cursor over each of the properties, you can see a brief description of what the property does. The number of service tabs shown depends on the services you decided to install in your cluster. **Any tab that requires input shows a red badge with the number of properties that need attention.** Select each service tab that displays a red badge number and enter the appropriate information.

Directories

The choice of directories where HDP will store information is critical. Ambari will attempt to choose reasonable defaults based on the mount points available in your environment but you are **strongly encouraged** to review the default directory settings recommended by Ambari. In particular, confirm directories such as `/tmp` and `/var` are **not** being used for HDFS NameNode directories and DataNode directories under the **HDFS** tab.

Passwords

You must provide database passwords for the Hive and Oozie services and the Master Secret for Knox. Using Hive as an example, choose the **Hive** tab and expand the Advanced section. In Database Password field marked in red, provide a password, then retype to confirm it.



Note

By default, Ambari will install a new MySQL instance for the Hive Metastore and install a Derby instance for Oozie. If you plan to use existing databases for MySQL, Oracle or PostgreSQL, modify these options before proceeding. Refer to [Using Non-Default Databases](#) for more information on using existing databases.



Important

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

Service Account Users and Groups

The service account users and groups are available under the **Misc** tab. These are the **operating system accounts the service components will run as**. If these users do not exist on your hosts, Ambari will automatically create the users and groups locally on the hosts. If these users already exist, Ambari will use those accounts.

Depending on how your environment is configured, you might not allow groupmod or usermod operations. If this is the case, you **must** be sure all users and groups are already created and **be sure to** select the "Skip group modifications" option on the **Misc** tab. This tells Ambari to not modify group membership for the service users.

Refer to the Ambari Reference Guide [Customizing HDP Services](#) for more information on the service account users and groups that are needed for HDP.

After you complete Customizing Services, choose `Next`.

3.11. 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`.

3.12. 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`.

3.13. Complete

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