

Hortonworks Data Platform

Automated Install with Ambari

(Jul 15, 2014)

Hortonworks Data Platform : Automated Install with Ambari

Copyright © 2012, 2014 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	2
1.2.1. Hardware Recommendations	2
1.2.2. Operating Systems Requirements	2
1.2.3. Browser Requirements	3
1.2.4. Software Requirements	3
1.2.5. JDK Requirements	4
1.2.6. Database Requirements	4
1.2.7. File System Partitioning Recommendations	4
1.2.8. Recommended Maximum Open File Descriptors	4
1.3. Collect Information	4
1.4. Prepare the Environment	5
1.4.1. Check Existing Installs	6
1.4.2. Set Up Password-less SSH	6
1.4.3. Set up Users and Groups	7
1.4.4. Enable NTP on the Cluster and on the Browser Host	8
1.4.5. Check DNS	8
1.4.6. Configuring iptables	9
1.4.7. Disable SELinux and PackageKit and check the UMASK Value	9
1.4.8. Disable Transparent Huge Pages (THP)	10
1.4.9. Make /tmp Executable	10
1.5. Optional: Configure Local Repositories	10
1.5.1. Obtaining the Repositories	11
1.5.2. Setting Up a Local Repository	13
2. Installing Ambari Server	19
2.1. Set Up the Bits	19
2.1.1. RHEL/CentOS/Oracle Linux 5.x	20
2.1.2. RHEL/CentOS/Oracle Linux 6.x	20
2.1.3. SLES 11	20
2.2. Set Up the Server	21
2.2.1. Setup Options	22
2.3. Start the Ambari Server	22
3. Installing, Configuring, and Deploying a HDP Cluster	24
3.1. Log In to Apache Ambari	24
3.2. Name Your Cluster	24
3.3. Select Stack	25
3.4. Install Options	26
3.5. Confirm Hosts	27
3.6. Choose Services	27
3.7. Assign Masters	28
3.8. Assign Slaves and Clients	28
3.9. Customize Services	29
3.10. Review	29
3.11. Install, Start and Test	29
3.12. Complete	30

List of Tables

1.1. Setting up Ambari Repository - No Internet Access	11
1.2. Setting up Ambari Repository - Temporary Internet Access	11
1.3. HDP 2.1 tarballs:	11
1.4. HDP 2.0 tarballs:	12
1.5. HDP 1.3 tarballs:	12
1.6. HDP 2.1 repository files:	12
1.7. HDP 2.0 repository files:	13
1.8. HDP 1.3 repository files:	13
1.9. Untar Locations for a Local Repository - No Internet Access	15
1.10. URLs for a Local Repository - No Internet Access	15
1.11. URLs for the New Repository	17
1.12. Base URLs for a Local Repository	17
2.1. Download the Ambari repository	19
3.1. Operating Systems mapped to each OS Family	26

1. Getting Ready

This section describes the information and materials you need to get ready to install Hadoop using Apache Ambari. **Apache Ambari** provides an end-to-end management and monitoring solution for Apache Hadoop. With Ambari, you can deploy and operate a Hadoop Stack using a Web UI and REST API to manage configuration changes and monitor services for all the 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 1.2.0	HDP 1.2.1.	HDP 1.3.0	HDP 1.3.2	HDP 1.3.3	HDP 2.0 ^a	HDP 2.1 ^b
1.6.1				X	X	X	X
1.6.0				X	X	X	X
1.5.1				X	X	X	X
1.5.0				X	X	X	
1.4.4.23				X	X	X	
1.4.3.38				X	X	X	
1.4.2.104				X	X	X	
1.4.1.61				X	X	X	
1.4.1.25				X		X	
1.2.5.17		X	X	X			
1.2.4.9	X	X	X				
1.2.3.7	X	X	X				
1.2.3.6	X	X					
1.2.2.5	X	X					
1.2.2.4	X	X					
1.2.2.3	X						
1.2.1.2	X						
1.2.0.1	X						

^aAmbari 1.6x does not install Flume and Hue services for the HDP 2.0 Stack.

^bAmbari 1.6x does not install Accumulo, Flume, Hue, Knox, or Solr services for the HDP 2.1 Stack.

For more information about:

- Installing Accumulo, Flume, Hue, Knox, and Solr services, see [Installing HDP Manually](#).

- HDP 2.0.6 stack (or later) patch releases, see HDP release notes, available at [HDP Documentation](#).
- Deploying Ambari and the HDP Stack, see [Deploying, Configuring, and Upgrading HDP](#).

1.2. Meet Minimum System Requirements

To run Hadoop, your system must meet minimum requirements.

- [Hardware Recommendations](#)
- [Operating Systems Requirements](#)
- [Browser Requirements](#)
- [Software Requirements](#)
- [JDK Requirements](#)
- [Database Requirements](#)
- [File System Partitioning Recommendations](#)

1.2.1. Hardware Recommendations

There is no single hardware requirement set for installing Hadoop.

For more information on the parameters that may affect your installation, see [Hardware Recommendations For Apache Hadoop](#).

1.2.2. Operating Systems Requirements

The following operating systems are supported:

- Red Hat Enterprise Linux (RHEL) v5.x or 6.x (64-bit)
- CentOS v5.x or 6.x (64-bit)
- Oracle Linux v5.x or 6.x (64-bit)
- SUSE Linux Enterprise Server (SLES) 11, SP1 or SP3 (64-bit)



Note

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.



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 [Optional: Configure the Local Repositories](#)

1.2.3. Browser Requirements

The Ambari Install Wizard runs as a browser-based Web app. You must have a machine capable of running a graphical browser to use this tool. The supported browsers are:

- Windows (Vista, 7)
 - Internet Explorer 9.0 and higher (for Vista + Windows 7)
 - Firefox latest stable release
 - Safari latest stable release
 - Google Chrome latest stable release
- Mac OS X (10.6 or later)
 - Firefox latest stable release
 - Safari latest stable release
 - Google Chrome latest stable release
- Linux (RHEL, CentOS, SLES, Oracle Linux)
 - Firefox latest stable release
 - Google Chrome latest stable release

1.2.4. Software Requirements

On each of your hosts:

- yum and rpm (RHEL/CentOS/Oracle Linux)
- zypper (SLES)
- scp, curl, and wget
- python (2.6 or later)



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.

1.2.5. JDK Requirements

The following Java runtime environments are supported:

- Oracle JDK 1.7_45 64-bit (default)
- Oracle JDK 1.6_31 64-bit



Note

Deprecated as of Ambari 1.5.1

- OpenJDK 7 64-bit (not supported on SLES)

1.2.6. Database Requirements

Hive/HCatalog, Oozie, and Ambari all require their own internal databases.

- Hive/HCatalog: By default uses an Ambari-installed MySQL 5.x instance. With appropriate preparation, you can also use an existing PostgreSQL 9.x, MySQL 5.x, or Oracle 11g r2 instance. See [Using Non-Default Databases-Hive](#) for more information on using existing instances.
- Oozie: By default uses an Ambari-installed Derby instance. With appropriate preparation, you can also use an existing PostgreSQL 9.x, MySQL 5.x, or Oracle 11g r2 instance. See [Using Non-Default Databases-Oozie](#) for more information on using existing instances.
- Ambari: By default uses an Ambari-installed PostgreSQL 8.x instance. With appropriate preparation, you can also use an existing PostgreSQL 9.x, MySQL 5.x, or Oracle 11g r2 instance. See [Using Non-Default Databases-Ambari](#) for more information on using existing instances.

1.2.7. File System Partitioning Recommendations

For information on setting up file system partitions on master and slave nodes in a HDP cluster, see [File System Partitioning Recommendations](#).

1.2.8. Recommended 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:

```
"ulimit -Sn" and "ulimit -Hn"
```

1.3. Collect Information

To deploy your Hadoop installation, you need to collect the following information:

- The fully qualified domain name (FQDN) for each host in your system, and which components you want to set up on which host. The Ambari install wizard *does*

not support using IP addresses. You can use `hostname -f` to check for the FQDN if you do not know it.



Note

While it is possible to deploy all Hadoop components on a single host, this is appropriate only for initial evaluation. In general you should use at least three hosts: one master host and two slaves.

- 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



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:

- [Check Existing Installs](#)
- [Set up Password-less SSH](#)
- [Set up Users and Groups](#)
- [Enable NTP on the Cluster](#)
- [Check DNS](#)
- [Configure iptables](#)
- [Disable SELinux, PackageKit and Check umask Value](#)
- [Disable Transparent Huge Pages \(THP\)](#)

- [Make /tmp Executable](#)

1.4.1. Check Existing Installs

Ambari automatically installs the correct versions of the files that are necessary for Ambari and Hadoop to run. Versions other than the ones that Ambari uses can cause problems in running the installer, so remove any existing installs that do not match the following lists.

	RHEL/CentOS/Oracle Linux v5	RHEL/CentOS/Oracle Linux v6	SLES 11
Ambari Server	<ul style="list-style-type: none"> • 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 	<ul style="list-style-type: none"> • postgresql 8.4.13-1.el6_3 • postgresql-libs 8.4.13-1.el6_3 • postgresql-server 8.4.13-1.el6_3 	<ul style="list-style-type: none"> • postgresql 8.3.5-1 • postgresql-server 8.3.5-1 • postgresql-libs 8.3.5-1
Ambari Agent ^a	<ul style="list-style-type: none"> • libffi 3.0.5-1.el5 • python26 2.6.8-2.el5 • python26-libs 2.6.8-2.el5 	None	None
Nagios Server ^b	<ul style="list-style-type: none"> • nagios 3.5.0-99 • nagios-devel 3.5.0-99 • nagios-www 3.5.0-99 • nagios-plugins 1.4.9-1 	<ul style="list-style-type: none"> • nagios 3.5.0-99 • nagios-devel 3.5.0-99 • nagios-www 3.5.0-99 • nagios-plugins 1.4.9-1 	<ul style="list-style-type: none"> • 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 ^c	<ul style="list-style-type: none"> • 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 	<ul style="list-style-type: none"> • 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 	<ul style="list-style-type: none"> • 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 ^d	<ul style="list-style-type: none"> • ganglia-gmond 3.5.0-99 • libganglia 3.5.0-99 	<ul style="list-style-type: none"> • ganglia-gmond 3.5.0-99 • libganglia 3.5.0-99 	<ul style="list-style-type: none"> • ganglia-gmond 3.5.0-99 • libganglia 3.5.0-99

^aInstalled on each host in your cluster. Communicates with the Ambari Server to execute commands.

^bThe host that runs the Nagios server

^cThe host that runs the Ganglia Server

^dInstalled on each host in the cluster. Sends metrics data to the Ganglia Collector.

1.4.2. Set Up Password-less SSH

To have Ambari Server automatically install Ambari Agents in all your cluster hosts, you must set up password-less SSH connections between the main installation (Ambari Server) host and all other machines. The Ambari Server host acts as the client and uses the key-pair to access the other hosts in the cluster to install the Ambari Agent.



Note

You can choose to install the Agents on each cluster host manually. In this case you do not need to setup SSH. See [Installing Ambari Agents Manually](#) for more information.

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.

```
ssh root@{remote.target.host}
```

You may see this warning. This happens on your first connection and is normal.

```
Are you sure you want to continue connecting (yes/no)?
```

6. 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.

1.4.3. Set up Users and Groups

The Ambari installer automatically creates default user and group accounts for you. Ambari preserves any existing user and group accounts, and uses these accounts when configuring Hadoop services. User and group creation applies to user/group accounts on the local operating system and to LDAP/AD accounts.

For more information about customizing user and group accounts, see one of the following topics:

- [Customizing Services for HDP 1.x Stack](#)
- [Customizing Services for HDP 2.x Stack](#)

1.4.4. 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 Ambari Web must be able to synchronize with each other.

1.4.5. Check DNS

All hosts in your system must be configured for DNS and Reverse DNS.

If you are unable to configure DNS and Reverse DNS, you must edit the hosts file on every host in your cluster to contain the address of each of your hosts and to set the Fully Qualified Domain Name hostname of each of those hosts. The following instructions cover basic hostname network setup for generic Linux hosts. Different versions and flavors of Linux might require slightly different commands. Please refer to your specific operating system documentation for the specific details for your system.

1.4.5.1. 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
```



Note

Do **not** remove the following two lines from your host file, or various programs that require network functionality may fail.

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

1.4.5.2. 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 name you just set.

1.4.5.3. Edit the Network Configuration File

1. Using a text editor, open the network configuration file on every host. This file is used to 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
```

1.4.6. 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`.

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

You can restart `iptables` after setup is complete.

If the security protocols at your installation do not allow you to disable `iptables`, you can proceed with them on, as long as all of the relevant ports are open and available.

During the Ambari Server setup process, Ambari checks to see if `iptables` is running. If it is, a warning prints to remind you to check that the necessary ports are open and available. The **Host Confirm** step of the Cluster Install Wizard will also issue a warning for each host that has `iptables` running.



Important

If you leave `iptables` enabled and do not set up the necessary ports, the cluster installation will fail.

1.4.7. Disable SELinux and PackageKit and check the UMASK Value

The following steps serve as a security checklist:

- You must temporarily 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.

- On an installation host running RHEL/CentOS with PackageKit installed, open `/etc/yum/pluginconf.d/refresh-packagekit.conf` using a text editor and make this change:

```
enabled=0
```



Note

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

- 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
```

```
umask 022
```

1.4.8. Disable Transparent Huge Pages (THP)

When installing Ambari on CentOS6.x using the Cluster Installer Wizard at the Host Checks step, one or more host checks may fail if you have not disabled Transparent Huge Pages on all hosts. Host Checks warns you when a failure occurs. To disable THP:

1. Add the following command to your `/etc/rc.local` file:

- RHEL6

```
if test -f /sys/kernel/mm/redhat_transparent_hugepage/defrag; then echo never > /sys/kernel/mm/redhat_transparent_hugepage/defrag fi
```

- SLES

```
if test -f /sys/kernel/mm/transparent_hugepage/defrag; then echo never > /sys/kernel/mm/transparent_hugepage/defrag fi
```

2. To confirm, reboot the host. Then, run the following command:

```
$ cat /sys/kernel/mm/transparent_hugepage/enabled always madvise [never]
```

1.4.9. Make /tmp Executable

If your hosts use `/tmp` as a mount point, confirm that the `/tmp` mount allows execute. In other words, verify that the `/tmp` mount is not set to `noexec`.

1.5. Optional: Configure Local Repositories

If your cluster is behind a firewall 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](#)

1.5.1. Obtaining the Repositories

This section describes how to obtain:

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

1.5.1.1. Ambari Repositories

If you do not have Internet access for setting up the Ambari repository, use the following tarballs:

Table 1.1. Setting up Ambari Repository - No Internet Access

Cluster OS	Ambari Repository Tarballs
RHEL/ CentOS/Oracle Linux 5.x	wget http://public-repo-1.hortonworks.com/ambari/centos5/ambari-1.6.1-
RHEL/CentOS/Oracle Linux 6.x	wget http://public-repo-1.hortonworks.com/ambari/centos6/ambari-1.6.1-
SLES 11	wget http://public-repo-1.hortonworks.com/ambari/suse11/ambari-1.6.1-

If you have temporary Internet access for setting up the Ambari repository, use the following repository configuration files:

Table 1.2. Setting up Ambari Repository - Temporary Internet Access

Cluster OS	Ambari Repository Configuration Files
RHEL/CentOS/ Oracle Linux 5.x	wget http://public-repo-1.hortonworks.com/ambari/centos5/1.x/updates/1.6.1/ambari.repo -
RHEL/CentOS/ Oracle Linux 6.x	wget http://public-repo-1.hortonworks.com/ambari/centos6/1.x/updates/1.6.1/ambari.repo -
SLES 11	wget http://public-repo-1.hortonworks.com/ambari/suse11/1.x/updates/1.6.1/ambari.repo -O

1.5.1.2. HDP Stack Repositories

If you do not have Internet access to set up the Stack repositories, use the following tarballs based on the HDP Stack version you plan to install:

Table 1.3. HDP 2.1 tarballs:

Cluster OS	HDP Repository Tarballs
RHEL/ CentOS/Oracle Linux 5.x	wget http://public-repo-1.hortonworks.com/HDP/centos5/HDP-2.1.5.0-centos5-rpm.
	wget http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.17/repos/centos5/HDE
RHEL/CentOS/Oracle Linux 6.x	wget http://public-repo-1.hortonworks.com/HDP/centos6/HDP-2.1.5.0-centos6-rpm.
	wget http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.17/repos/centos6/HDE

Cluster OS	HDP Repository Tarballs
SLES 11 SP1	wget http://public-repo-1.hortonworks.com/HDP/sles11sp1/HDP-2.1.5.0-sles11sp1- wget http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.19/repos/sles11sp1/H
SLES 11 SP3	wget http://public-repo-1.hortonworks.com/HDP/suse11sp3/HDP-2.1.5.0-suse11sp3- wget http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.19/repos/suse11sp3/H

Table 1.4. HDP 2.0 tarballs:

Cluster OS	HDP Repository Tarballs
RHEL/ CentOS/ Oracle Linux 5.x	wget http://public-repo-1.hortonworks.com/HDP/centos5/HDP-2.0.12.0-centos5- rpm.tar.gz wget 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.x	wget http://public-repo-1.hortonworks.com/HDP/centos6/HDP-2.0.12.0-centos6- rpm.tar.gz wget 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 http://public-repo-1.hortonworks.com/HDP/suse11/HDP-2.0.12.0-suse11- rpm.tar.gz wget http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.17/repos/suse11/ HDP-UTILS-1.1.0.17-suse11.tar.gz

Table 1.5. HDP 1.3 tarballs:

Cluster OS	HDP Repository Tarballs
RHEL/ CentOS/ Oracle Linux 5.x	wget http://public-repo-1.hortonworks.com/HDP/centos5/HDP-1.3.7.0-centos5- rpm.tar.gz wget http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.16/repos/centos5/ HDP-UTILS-1.1.0.16-centos5.tar.gz
RHEL/ CentOS/ Oracle Linux 6.x	wget http://public-repo-1.hortonworks.com/HDP/centos6/HDP-1.3.7.0-centos6- rpm.tar.gz wget 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 http://public-repo-1.hortonworks.com/HDP/suse11/HDP-1.3.7.0-suse11- rpm.tar.gz wget http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.16/repos/suse11/ HDP-UTILS-1.1.0.16-suse11.tar.gz

If you have temporary Internet access for setting up the Stack repositories, use the following repository configuration files based on the HDP Stack version you plan to install:

Table 1.6. HDP 2.1 repository files:

Cluster OS	HDP Repository Configuration Files
RHEL/ CentOS/ Oracle Linux 5.x	wget 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.x	wget http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.1.5.0/ hdp.repo -O /etc/yum.repos.d/hdp.repo

Cluster OS	HDP Repository Configuration Files
SLES 11 SP1	wget http://public-repo-1.hortonworks.com/HDP/sles11sp1/2.x/updates/2.1.5.0/hdp.repo -O /etc/zypp/repos.d/hdp.repo
SLES 11 SP3	wget http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.1.5.0/hdp.repo -O /etc/zypp/repos.d/hdp.repo

Table 1.7. HDP 2.0 repository files:

Cluster OS	HDP Repository Configuration Files
RHEL/CentOS/Oracle Linux 5.x	wget 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.x	wget 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 http://public-repo-1.hortonworks.com/HDP/suse11/2.x/updates/2.0.12.0/hdp.repo -O /etc/zypp/repos.d/hdp.repo

Table 1.8. HDP 1.3 repository files:

Cluster OS	HDP Repository Configuration Files
RHEL/CentOS/Oracle Linux 5.x	wget http://public-repo-1.hortonworks.com/HDP/centos5/1.x/updates/1.3.7.0/hdp.repo -O /etc/yum/repos.d/hdp.repo
RHEL/CentOS/Oracle Linux 6.x	wget 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 http://public-repo-1.hortonworks.com/HDP/suse11/1.x/updates/1.3.7.0/hdp.repo -O /etc/zypp/repos.d/hdp.repo

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](#)

1.5.2.1. Getting Started Setting Up a Local Repository

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

- Select a mirror server 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) or zypper (SLES)
- **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 the firewall settings (if any) 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
```

- 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.

1.5.2.2. 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
```

- b. Untar the repository tarballs to the following locations:

Table 1.9. 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.

Table 1.10. URLs for a Local Repository - No Internet Access

Repository	URL
Ambari Base URL	http://{web-server}/ambari/{os}/1.x/updates/1.6.1
HDP Base URL	http://{web-server}/hdp/HDP/{os}/2.x/updates/{latest}
HDP-UTILS Base URL	http://{web-server}/hdp/HDP-UTILS-{version}/repos/{os}



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.
 - a. Install the plug-in.

- **For RHEL and CentOS 5**

```
yum install yum-priorities
```

- **For RHEL and CentOS 6**

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

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

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

1.5.2.3. Setting up a Local Repository Having 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 the repositories are available.

For RHEL/CentOS/Oracle Linux:

```
yum repolist
```

For SLES:

```
zypper repos
```

3. Browse to the web server directory.

For RHEL/CentOS/Oracle Linux:

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

For SLES:

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

4. Synchronize the repository contents to your mirror server.

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

```
mkdir -p ambari/{$os}
cd ambari/{$os}
reposync -r Updates-ambari-1.6.1
```

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

```
mkdir -p hdp/{$os}
cd hdp/{$os}
reposync -r HDP-{$latest}
reposync -r HDP-UTILS-{$version}
```

5. Generate the repository metadata.

- For Ambari:

```
createrepo {web-server-directory}/ambari/{$os}/Updates-ambari-1.6.1
```

- For HDP Stack Repositories:

```
createrepo {web-server-directory}/hdp/{$os}/HDP-{$latest}
createrepo {web-server-directory}/hdp/{$os}/HDP-UTILS-{$version}
```

6. Confirm you can browse to the newly created repository.

Table 1.11. URLs for the New Repository

Repository	URL
Ambari Base URL	<code>http://{web-server}/ambari/{os}/Updates-ambari-1.6.1</code>
HDP Base URL	<code>http://{web-server}/hdp/{os}/HDP-{\$latest}</code>
HDP-UTILS Base URL	<code>http://{web-server}/hdp/{os}/HDP-UTILS-{\$version}</code>



Important

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

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

a. Install the plug-in.

- **RHEL and CentOS 5**

```
yum install yum-priorities
```

- **RHEL and CentOS 6**

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

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

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

1.5.2.4. 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/1.6.1/ambari.repo
```

- From the public repository:

```
http://public-repo-1.hortonworks.com/ambari/{os}/1.x/updates/1.6.1/
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 having No Internet Access](#), or step 5 in [Setting Up a Local Repository having Temporary Internet Access](#), if necessary.

Table 1.12. Base URLs for a Local Repository

Repository	URL
Ambari Base URL	<code>http://{web-server}/ambari/{os}/1.x/updates/1.6.1</code>

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-1.6.1]
name=ambari-1.6.1 - 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
```

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

a. **RHEL/CentOS/Oracle Linux**

```
/etc/yum.repos.d
```

SLES

```
/etc/zypp/repos.d
```

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

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

4. Proceed to [Running the Ambari Installer](#) to install and setup Ambari Server.

2. Installing Ambari Server

Ambari manages installing and deploying Hadoop. This section describes how to install Apache Ambari.

- [Set Up the Bits](#)
- [Set Up the Server](#)
- [Start the Ambari Server](#)

2.1. Set Up the Bits

1. Log into the machine that is to serve as the Ambari Server as `root`. You may login and `sudo` as `su` if this is what your environment requires. This machine is the main installation host.
2. Download the Ambari repository file and copy it to your `repos.d`.

Table 2.1. Download the Ambari repository

Platform	Access
RHEL, CentOS, and Oracle Linux 5	<pre>wget http://public-repo-1.hortonworks.com/ambari/centos5/1.x/updates/1.6.1/ambari.repo</pre> <pre>cp ambari.repo /etc/yum.repos.d</pre>
RHEL, CentOS and Oracle Linux 6	<pre>wget http://public-repo-1.hortonworks.com/ambari/centos6/1.x/updates/1.6.1/ambari.repo</pre> <pre>cp ambari.repo /etc/yum.repos.d</pre>
SLES 11	<pre>wget http://public-repo-1.hortonworks.com/ambari/suse11/1.x/updates/1.6.1/ambari.repo</pre> <pre>cp ambari.repo /etc/zypp/repos.d</pre>



Important

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



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 [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.

When you have the software, continue your installation based on your base platform.

2.1.1. RHEL/CentOS/Oracle Linux 5.x

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

```
yum repolist
```

You should see the Ambari and HDP utilities repositories in the list. The version values vary depending on the installation.

repo id	repo name
AMBARI-1.x	Ambari 1.x

2. Install the Ambari bits using yum. This also installs PostgreSQL.

```
yum install ambari-server
```

2.1.2. RHEL/CentOS/Oracle Linux 6.x

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

```
yum repolist
```

You should see the Ambari and HDP utilities repositories in the list. The version values vary depending the installation.

repo id	repo name
AMBARI-1.x	Ambari 1.x

2. Install the Ambari bits using yum. This also installs PostgreSQL.

```
yum install ambari-server
```

2.1.3. SLES 11

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

```
zypper repos
```

You should see the Ambari and HDP utilities in the list. The version values vary depending the installation.

#	Alias	Name
1	AMBARI.dev-1.x	Ambari 1.x

2. Install the Ambari bits using zypper. This also installs PostgreSQL.

```
zypper install ambari-server
```


2.2. Set Up the Server

The `ambari-server` command manages the setup process. Run the following command and respond to the prompts:

```
ambari-server setup
```

1. If you have *not* temporarily disabled SELinux, you may get a warning. Enter `y` to continue.
2. By default, Ambari Server runs under `root`. If you want to create a different user to run the Ambari Server instead, or to assign a previously created user, select `y` at **Customize user account for ambari-server daemon** then provide a user name at the prompt.
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.



Note

By default, Ambari Server setup downloads and installs Oracle JDK 1.7. If you plan to use a different version of the JDK, see [Setup Options](#) for more information.

5. Agree to the Oracle JDK license when asked. You must accept this license to be able to download the necessary JDK from Oracle. The JDK is installed during the deploy phase.
6. At **Enter advanced database configuration**:
 - To use the default PostgreSQL database, named `ambari`, with the default user name and password (`ambari/bigdata`), enter `n`.



Important

If you are using an existing PostgreSQL, MySQL, or Oracle database instance, you must prepare the instance using the steps detailed in [Using Non-Default Databases-Ambari](#) **before** running the installer.

- 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 required by 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 required by 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 required by the prompts, including host name, port, database name, user name, and password.

7. Setup completes.



Note

If your host accesses the Internet through a proxy server, you must configure Ambari Server to use this proxy server. See [Configure Ambari Server for Internet Proxy](#) for more information.

2.2.1. Setup Options

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

Option	Description
-j -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, Setup downloads the Oracle JDK 1.7 binary to <code>/var/lib/ambari-server/resources</code> on the Ambari Server and installs the JDK to <code>/usr/jdk64</code> . 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. This path must be valid on all hosts. For example: <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 and to make that JAR available to Ambari Server for distribution to cluster hosts during configuration. Use this option with the <code>-jdbc-db</code> option to specify the database type.
-jdbc-db	Specifies the database type. Valid values are: [postgres mysql oracle] Use this option with the <code>-jdbc-driver</code> option to specify the location of the JDBC driver JAR file.
-s -silent	Setup runs silently. Accepts all default prompt values.
-v -verbose	Prints verbose info and warning messages to the console during Setup.
-g -debug	Start Ambari Server in debug mode

2.3. Start the Ambari Server

- To start the Ambari Server:

```
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/HCatalog or for Oozie, you must complete the preparations described in [Using Non-Default Databases](#) **before** installing your Hadoop cluster.

Next Steps

[Installing, Deploying and Configuring a HDP Cluster](#)

3. Installing, Configuring, and Deploying a HDP Cluster

This section describes how to use the Ambari install wizard running in your browser to install, configure, and deploy your Hortonworks Data Platform (HDP) 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](#)

3.1. Log In to Apache Ambari

After starting the Ambari service, access Ambari using a web browser.

1. Point your browser to `http://{your.ambari.server}:8080`.
2. Log in to the Ambari Server using the default user name/password: admin/admin. You can change these credentials later.

3.2. Name Your Cluster

For a new cluster, the Ambari install wizard displays a Welcome page in which you define a cluster name.

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**.

3.3. 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 2.1 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 2.1
- HDP 2.0
- HDP 1.3.3
- HDP 1.3.2

[Advanced Repository Options](#)

[← Back](#) [Next →](#)

Under Advanced Repository Options, you can select the Base URL of a repository from which Stack software packages download. Ambari sets the following, default repository URLs, depending on the Internet connectivity available to the Ambari server host:

- 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 3, or HDP-2.1.3. If you want to install HDP 2.1 Stack patch release 2, or HDP-2.1.2 instead, obtain the repository Base URL from the HDP Stack documentation.
- If you are using local repositories, see [Optional: Configure Ambari for Local Repositories](#) for the Base URLs and enter them here to use your local repositories instead of the 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"/> redhat5	HDP-2.1	<input type="text" value="http://public-repo-1.hortonworks.com/HDP/centos5/2.x/u"/>
	HDP-UTILS-1.1.0.17	<input type="text" value="http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.1"/>
<input checked="" type="checkbox"/> redhat6	HDP-2.1	<input type="text" value="http://public-repo-1.hortonworks.com/HDP/centos6/2.x/u"/>
	HDP-UTILS-1.1.0.17	<input type="text" value="http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.1"/>
<input checked="" type="checkbox"/> suse11	HDP-2.1	<input type="text" value="http://public-repo-1.hortonworks.com/HDP/suse11/2.x/uj"/>
	HDP-UTILS-1.1.0.17	<input type="text" value="http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.1"/>

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.

Table 3.1. Operating Systems mapped to each OS Family

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

3.4. Install Options

In order to build up the cluster, the install wizard needs to know 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**. See [Installing Ambari Agents Manually](#) for more information.
4. Choose **Register and Confirm** to continue.

3.5. 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**.



Important

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**.

3.6. Choose Services

HDP comprises many services. You must install the HDFS and ZooKeeper 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.



Note

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**.

3.7. 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**.

3.8. Assign Slaves and Clients

The Ambari install 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.



Note

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**.

3.9. 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.



Note

In **HDFS Services Configs General**, 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 except for those related to databases in the Hive and the Oozie tabs, and two related properties in the Nagios tab. These four are marked in red and are the only ones you *must* set yourself. Click the name of the group in each tab to expand and collapse the display.

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

3.10. 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**.

3.11. Install, Start and Test

The progress of the install is shown on the screen. Each component is installed and started and a simple test is run on the component. You are given an overall status on the process in the progress bar at the top of the screen and a host by host status in the main section.

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**.

3.12. Complete

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