

# HADOOP CLUSTER SETUP USING CLOUDERA MANAGER

By

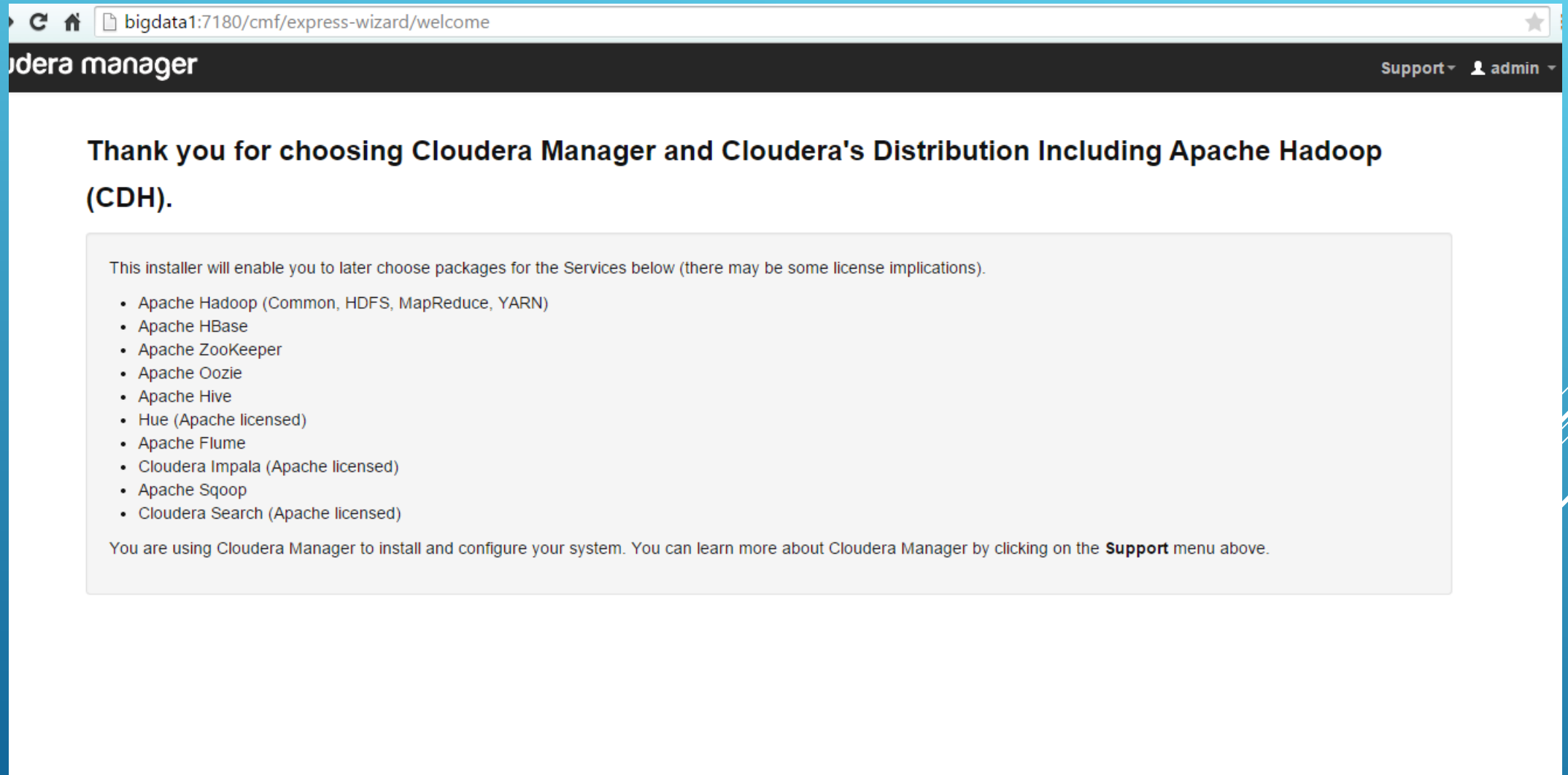
Sudheer Kondla

Solutions Architect

# LOGIN TO CLUSTERA MANAGER

- ▶ Step 1: Download and Run the Cloudera Manager Installer
- ▶ Step 2: Start the Cloudera Manager Admin Console
- ▶ Step 3: Use Cloudera Manager for Automated CDH Installation and Configuration
- ▶ Step 4: Change the Default Administrator Password
- ▶ Step 5: Test the Installation
- ▶ For example:
  - ▶ Once you downloaded cloudera-manager-installer.bin run the installation
  - ▶ `chmod u+x cloudera-manager-installer.bin`
  - ▶ `sudo setforce 0`
  - ▶ `sudo ./cloudera-manager-installer.bin`
  - ▶ when completed. Start the service
  - ▶ `sudo service clouder-scm-server restart` (or manually start Cloudera-manager if it not running)
  - ▶ Optionally add service to start automatically at boot
  - ▶ `chkconfig --list | grep -i Cloudera-scm-server`
  - ▶ `chkconfig clouder-scm-server on`
  - ▶ Now access Cloudera <http://servername:7180/>
  - ▶ Log into Cloudera Manager. The default credentials are: Username: admin Password: admin

# LOGIN TO CLOUDERA MANAGER



# ENTER LIST OF CLUSTER NODES

cloudera manager

Support admin

Specify hosts for your CDH cluster installation.

Cloudera recommends including Cloudera Manager server's host because it is often used for the Cloudera Management Service, and because this will enable health monitoring for that host.

**Hint:** Search for hostnames and/or IP addresses using [patterns](#).

bigdata[1-6].localdomain

SSH Port: 22 Search


# SEARCH CLUSTER NODES ON CLUSTERA MANAGER

cloudera manager

Support ▾  admin ▾

## Specify hosts for your CDH cluster installation.

Cloudera recommends including Cloudera Manager server's host because it is often used for the Cloudera Management Service, and because this will enable health monitoring for that host.

**Hint:** Search for hostnames and/or IP addresses using [patterns](#) .

6 hosts scanned, 6 running SSH.

 New Search

<input checked="" type="checkbox"/> Expanded Query	Hostname (FQDN)	IP Address	Currently Managed	Result
<input checked="" type="checkbox"/> bigdata1.localdomain	bigdata1.localdomain	192.168.0.131	No	✓ Host ready: 0 ms response time.
<input checked="" type="checkbox"/> bigdata2.localdomain	bigdata2.localdomain	192.168.0.132	No	✓ Host ready: 1 ms response time.
<input checked="" type="checkbox"/> bigdata3.localdomain	bigdata3.localdomain	192.168.0.133	No	✓ Host ready: 1 ms response time.
<input checked="" type="checkbox"/> bigdata4.localdomain	bigdata4.localdomain	192.168.0.134	No	✓ Host ready: 1 ms response time.
<input checked="" type="checkbox"/> bigdata5.localdomain	bigdata5.localdomain	192.168.0.135	No	✓ Host ready: 3 ms response time.
<input checked="" type="checkbox"/> bigdata6.localdomain	bigdata6.localdomain	192.168.0.136	No	✓ Host ready: 1 ms response time.

# SPECIFY REPOSITORY INFORMATION CLOUDERA MANAGER

## Cluster Installation

### Select Repository

**Cloudera Manager Parcels** are the easiest way for Cloudera Manager to manage the software on your cluster, by automating the deployment and upgrade of service binaries. Electing not to use parcels will require you to manually upgrade packages on all hosts in your cluster when software updates are available, and will prevent you from using Cloudera Manager's rolling upgrade capabilities.

#### Choose Method:

- ☐ Use Packages ⓘ
- ☒ Use Parcels (Recommended) ⓘ
  - More Options
- ☒ SQOOP\_NETEZZA\_CONNECTOR-1.2c5
- ☐ None
- ☒ CDH-5.1.0-1.cdh5.1.0.p0.53
- ☐ CDH-4.7.1-1.cdh4.7.1.p0.47

- ☒ SQOOP\_TERADATA\_CONNECTOR-1.2c5
- ☐ None

- ☒ SOLR-1.3.0-1.cdh4.5.0.p0.9
- ☐ None

**Note:** Solr is supported only on CDH 4.3 or later deployments.

- ☒ IMPALA-2.1.0-1.impala2.0.0.p0.1995
- ☐ None

**Note:** Impala is supported only on CDH 4.1 or later deployments.

Select the specific release of the Cloudera Manager Agent you want to install on your hosts.

- ☒ Matched release for this Cloudera Manager server
- ☐ Custom Repository

# SPECIFY REPOSITORY INFORMATION CONTD..

## Cluster Installation

### Select Repository

Cloudera recommends the use of parcels for installation over packages, because parcels enable Cloudera Manager to easily manage the software on your cluster, automating the deployment and upgrade of service binaries. Electing not to use parcels will require you to manually upgrade packages on all hosts in your cluster when software updates are available, and will prevent you from using Cloudera Manager's rolling upgrade capabilities.

#### Choose Method

- ☐ Use Packages ⓘ
- ☒ Use Parcels (Recommended) ⓘ [More Options](#)

#### Select the version of CDH

- ☒ CDH-5.4.1-1.cdh5.4.1.p0.6
- ☐ CDH-5.1.0-1.cdh5.1.0.p0.53
- ☐ CDH-4.7.1-1.cdh4.7.1.p0.47

Versions of CDH that are too new for this version of Cloudera Manager (5.4.1) will not be shown.

#### Additional Parcels

- ☐ ACCUMULO-1.6.0-1.cdh5.1.4.p0.116
- ☐ ACCUMULO-1.4.4-1.cdh4.5.0.p0.65
- ☒ None
- ☐ KAFKA-0.8.2.0-1.kafka1.3.0.p0.29
- ☒ None
- ☐ KEYTRUSTEE-5.4.0-1.cdh5.4.0.p0.193
- ☒ None
- ☐ SQOOP\_NETEZZA\_CONNECTOR-1.2c5
- ☒ None
- ☐ SQOOP\_TERADATA\_CONNECTOR-1.4c5
- ☐ SQOOP\_TERADATA\_CONNECTOR-1.2c5
- ☒ None

⏪ Back

1 2 3 4 5 6 7 8

⏩ Continue

# SPECIFY JDK AND ENCRYPTION OPTIONS

## Cluster Installation

### JDK Installation Options

Oracle Binary Code License Agreement for the Java SE Platform Products and JavaFX

ORACLE AMERICA, INC. ("ORACLE"), FOR AND ON BEHALF OF ITSELF AND ITS SUBSIDIARIES AND AFFILIATES UNDER COMMON CONTROL, IS WILLING TO LICENSE THE SOFTWARE TO YOU ONLY UPON THE CONDITION THAT YOU ACCEPT ALL OF THE TERMS CONTAINED IN THIS BINARY CODE LICENSE AGREEMENT AND SUPPLEMENTAL LICENSE TERMS (COLLECTIVELY "AGREEMENT"). PLEASE READ THE AGREEMENT CAREFULLY. BY SELECTING THE "ACCEPT LICENSE AGREEMENT" (OR THE EQUIVALENT) BUTTON AND/OR BY USING THE SOFTWARE YOU ACKNOWLEDGE THAT YOU HAVE READ THE TERMS AND AGREE TO THEM. IF YOU ARE AGREEING TO THESE TERMS ON BEHALF OF A COMPANY OR OTHER LEGAL ENTITY, YOU REPRESENT THAT YOU HAVE THE LEGAL AUTHORITY TO BIND THE LEGAL ENTITY TO THESE TERMS. IF YOU DO NOT HAVE SUCH AUTHORITY, OR IF YOU DO NOT WISH TO BE BOUND BY THE TERMS, THEN SELECT THE "DECLINE LICENSE AGREEMENT" (OR THE EQUIVALENT) BUTTON AND YOU MUST NOT USE THE SOFTWARE ON THIS SITE OR ANY OTHER MEDIA ON WHICH THE SOFTWARE IS CONTAINED.

1. DEFINITIONS. "Software" means the software identified above in binary form that you selected for download, install or use (in the version You selected for download, install or use) from Oracle or its authorized licensees, any other machine readable materials (including, but not limited to, libraries, source files, header files, and data files), any updates or error corrections provided by Oracle, and any user manuals, programming guides and other documentation provided to you by Oracle under this Agreement. "General Purpose Desktop Computers and Servers" means computers, including desktop and laptop computers, or servers, used for general computing functions under end

☒ Install Oracle Java SE Development Kit (JDK)

Check this box to accept the Oracle Binary Code License Agreement and install the JDK. Leave it unchecked to use a currently installed JDK.

☒ Install Java Unlimited Strength Encryption Policy Files

Check this checkbox if local laws permit you to deploy unlimited strength encryption and you are running a secure cluster.

◀ Back

1 2 3 4 5 6 7 8

▶ Continue



# SPECIFY SINGLE OR MULTI-USER MODE

## Cluster Installation

### Enable Single User Mode

**Only supported for CDH 5.2 and above.**

By default, service processes run as distinct users on the system. For example, HDFS DataNodes run as user "hdfs" and HBase RegionServers run as user "hbase." Enabling "single user mode" configures Cloudera Manager to run service processes as a single user, by default "cloudera-scm", thereby prioritizing isolation between managed services and the rest of the system over isolation between the managed services.

The **major benefit** of this option is that the Agent does not run as root. However, this mode complicates installation, which is described fully in the [documentation](#). Most notably, directories which in the regular mode are created automatically by the Agent, must be created manually on every host with appropriate permissions, and sudo (or equivalent) access must be set up for the configured user.

Switching back and forth between single user mode and regular mode is not supported.

Single User Mode



# SPECIFY SSH LOGIN CREDENTIALS FOR THE USER

## Cluster Installation

### Provide SSH login credentials.

Root access to your hosts is required to install the Cloudera packages. This installer will connect to your hosts via SSH and log in either directly as root or as another user with password-less sudo/pbrun privileges to become root.

Login To All Hosts As: ☒ root  
☐ Another user

You may connect via password or public-key authentication for the user selected above.

Authentication Method: ☒ All hosts accept same password  
☐ All hosts accept same private key

Enter Password:

Confirm Password:

SSH Port:

Number of Simultaneous Installations:  (Running a large number of installations at once can consume large amounts of network bandwidth and other system resources)







# INSTALL SOFTWARE COMPONENTS ON CLUSTER NODES

## Cluster Installation

Installation in progress.



0 of 6 host(s) completed successfully. [Abort Installation](#)

Hostname	IP Address	Progress	Status	
bigdata1.localdomain	192.168.0.131	<div><div></div></div>	⚙ Refreshing package metadata...	<a href="#">Details</a> 
bigdata2.localdomain	192.168.0.132	<div><div></div></div>	⚙ Refreshing package metadata...	<a href="#">Details</a> 
bigdata3.localdomain	192.168.0.133	<div><div></div></div>	⚙ Installing oracle-j2sdk1.7 package...	<a href="#">Details</a> 
bigdata4.localdomain	192.168.0.134	<div><div></div></div>	⚙ Refreshing package metadata...	<a href="#">Details</a> 
bigdata5.localdomain	192.168.0.135	<div><div></div></div>	⚙ Installing oracle-j2sdk1.7 package...	<a href="#">Details</a> 
bigdata6.localdomain	192.168.0.136	<div><div></div></div>	⚙ Installing oracle-j2sdk1.7 package...	<a href="#">Details</a> 







# INSTALL SOFTWARE COMPONENTS ON CLUSTER NODES

## Cluster Installation

Installation in progress.



0 of 6 host(s) completed successfully. [Abort Installation](#)

Hostname	IP Address	Progress	Status	
bigdata1.localdomain	192.168.0.131	<div><div></div></div>	⚙ Refreshing package metadata...	<a href="#">Details</a> 
bigdata2.localdomain	192.168.0.132	<div><div></div></div>	⚙ Refreshing package metadata...	<a href="#">Details</a> 
bigdata3.localdomain	192.168.0.133	<div><div></div></div>	⚙ Installing oracle-j2sdk1.7 package...	<a href="#">Details</a> 
bigdata4.localdomain	192.168.0.134	<div><div></div></div>	⚙ Refreshing package metadata...	<a href="#">Details</a> 
bigdata5.localdomain	192.168.0.135	<div><div></div></div>	⚙ Installing oracle-j2sdk1.7 package...	<a href="#">Details</a> 
bigdata6.localdomain	192.168.0.136	<div><div></div></div>	⚙ Installing oracle-j2sdk1.7 package...	<a href="#">Details</a> 







# INSTALL SOFTWARE COMPONENTS ON CLUSTER NODES – CONTD..

## Cluster Installation

Installation in progress.



0 of 6 host(s) completed successfully. [Abort Installation](#)

Hostname	IP Address	Progress	Status	
bigdata1.localdomain	192.168.0.131	<div><div></div></div>	⚙ Refreshing package metadata...	<a href="#">Details</a> 
bigdata2.localdomain	192.168.0.132	<div><div></div></div>	⚙ Refreshing package metadata...	<a href="#">Details</a> 
bigdata3.localdomain	192.168.0.133	<div><div></div></div>	⚙ Installing oracle-j2sdk1.7 package...	<a href="#">Details</a> 
bigdata4.localdomain	192.168.0.134	<div><div></div></div>	⚙ Refreshing package metadata...	<a href="#">Details</a> 
bigdata5.localdomain	192.168.0.135	<div><div></div></div>	⚙ Installing oracle-j2sdk1.7 package...	<a href="#">Details</a> 
bigdata6.localdomain	192.168.0.136	<div><div></div></div>	⚙ Installing oracle-j2sdk1.7 package...	<a href="#">Details</a> 

# REVIEW INSTALLATION LOGS

## Installation completed successfully. (Current Step)

Last Refreshed: May 17, 2015 5:01:00 PM EDT

```
'CDH_LLAMA_HOME': '/usr/lib/llama/', 'CDH_HTTPFS_HOME': '/usr/lib/hadoop-httpfs', 'CDH_HADOOP_HOME': '/usr/lib/hadoop', 'CDH_HIVE_HOME': '/usr/lib/hive', 'CDH_HCAT_HOME': '/usr/lib/hive-hcatalog',
'CDH_SENTRY_HOME': '/usr/lib/sentry', 'CDH_SPARK_HOME': '/usr/lib/spark', 'TOMCAT_HOME': '/usr/lib/bigtop-tomcat', 'CDH_FLUME_HOME': '/usr/lib/flume-ng'}
>>[17/May/2015 16:57:50 +0000] 9867 MainThread agent INFO To override these variables, use /etc/cloudera-scm-agent/config.ini. Environment variables for CDH locations are not used when CDH is installed from
parcels.
>>[17/May/2015 16:57:50 +0000] 9867 MainThread agent INFO Created /var/run/cloudera-scm-agent/process
>>[17/May/2015 16:57:50 +0000] 9867 MainThread agent INFO Chmod'ing /var/run/cloudera-scm-agent/process to 0751
>>[17/May/2015 16:57:50 +0000] 9867 MainThread agent INFO Created /var/run/cloudera-scm-agent/supervisor
>>[17/May/2015 16:57:50 +0000] 9867 MainThread agent INFO Chmod'ing /var/run/cloudera-scm-agent/supervisor to 0751
>>[17/May/2015 16:57:50 +0000] 9867 MainThread agent INFO Created /var/run/cloudera-scm-agent/supervisor/include
>>[17/May/2015 16:57:50 +0000] 9867 MainThread agent INFO Chmod'ing /var/run/cloudera-scm-agent/supervisor/include to 0751
>>[17/May/2015 16:57:50 +0000] 9867 MainThread agent ERROR Failed to connect to previous supervisor.
>>Traceback (most recent call last):
>> File "/usr/lib64/cmf/agent/src/cmf/agent.py", line 1522, in find_or_start_supervisor
>> self.configure_supervisor_clients()
>> File "/usr/lib64/cmf/agent/src/cmf/agent.py", line 1761, in configure_supervisor_clients
>> supervisor_options.realize(args=["-c", os.path.join(self.supervisor_dir, "supervisord.conf")])
>> File "/usr/lib64/cmf/agent/build/env/lib/python2.6/site-packages/supervisor-3.0-py2.6.egg/supervisor/options.py", line 1563, in realize
>> Options.realize(self, *arg, **kw)
>> File "/usr/lib64/cmf/agent/build/env/lib/python2.6/site-packages/supervisor-3.0-py2.6.egg/supervisor/options.py", line 310, in realize
>> self.process_config()
>> File "/usr/lib64/cmf/agent/build/env/lib/python2.6/site-packages/supervisor-3.0-py2.6.egg/supervisor/options.py", line 318, in process_config
>> self.process_config_file(do_usage)
>> File "/usr/lib64/cmf/agent/build/env/lib/python2.6/site-packages/supervisor-3.0-py2.6.egg/supervisor/options.py", line 353, in process_config_file
>> self.usage(str(msg))
>> File "/usr/lib64/cmf/agent/build/env/lib/python2.6/site-packages/supervisor-3.0-py2.6.egg/supervisor/options.py", line 141, in usage
>> self.exit(2)
>>SystemExit: 2
>>[17/May/2015 16:57:50 +0000] 9867 MainThread tmpfs INFO Successfully mounted tmpfs at /var/run/cloudera-scm-agent/process
END (0)
end of agent logs.
scm agent started
```

## Installation script completed successfully.

```
all done
closing logging file descriptor
```

Close













# CLOUDERA CLUSTER INSTALLATION COMPLETE

## Cluster Installation

Installation completed successfully.



6 of 6 host(s) completed successfully.

Hostname	IP Address	Progress	Status	
bigdata1.localdomain	192.168.0.131		✓ Installation completed successfully.	<a href="#">Details</a> 
bigdata2.localdomain	192.168.0.132		✓ Installation completed successfully.	<a href="#">Details</a> 
bigdata3.localdomain	192.168.0.133		✓ Installation completed successfully.	<a href="#">Details</a> 
bigdata4.localdomain	192.168.0.134		✓ Installation completed successfully.	<a href="#">Details</a> 
bigdata5.localdomain	192.168.0.135		✓ Installation completed successfully.	<a href="#">Details</a> 
bigdata6.localdomain	192.168.0.136		✓ Installation completed successfully.	<a href="#">Details</a> 

# CLUDERA CLUSTER INSTALLATION – INSTALL PARCELS

## Cluster Installation

### Installing Selected Parcels

The selected parcels are being downloaded and installed on all the hosts in the cluster.

▼ CDH 5.4.1-1.cd5.4.1.p0.6	Downloaded: 100%	Distributed: 0/6 (42.2 MiB/s)	Unpacked: 0/6	Activated: 0/6
<div><div></div></div>				



# CLOUDERA CLUSTER INSTALLATION – INSTALL PARCELS

## Cluster Installation

### Installing Selected Parcels

The selected parcels are being downloaded and installed on all the hosts in the cluster.

▼ CDH 5.4.1-1.cdh5.4.1.p0.6	Downloaded: 100%	Distributed: 6/6	Unpacked: 6/6	Activated: 6/6
<div><div></div></div>				

# CLOUDERA CLUSTER INSTALLATION – INSTALL PARCELS

## Cluster Installation

### Installing Selected Parcels

The selected parcels are being downloaded and installed on all the hosts in the cluster.

▼ CDH 5.4.4

• Host is i

bigdata1.localdomain - Active

Distributed	100% (1....
Throughput	11.7 MiB/s
Unpacked	Yes
Activated	Yes
Errors	-

# CLOUDERA CLUSTER INSTALLATION – INSPECT DEPENDENCIES

## Cluster Installation

Inspect hosts for correctness [↻ Run Again](#)

### Validations

- ✓ Inspector ran on all 6 hosts.
- ✓ The following failures were observed in checking hostnames...
- ✓ No errors were found while looking for conflicting init scripts.
- ✓ No errors were found while checking /etc/hosts.
- ✓ All hosts resolved localhost to 127.0.0.1.
- ✓ All hosts checked resolved each other's hostnames correctly and in a timely manner.
- ✓ Host clocks are approximately in sync (within ten minutes).
- ✓ Host time zones are consistent across the cluster.
- ✓ No users or groups are missing.
- ✓ No conflicts detected between packages and parcels.
- ✓ No kernel versions that are known to be bad are running.
- ✓ All hosts have /proc/sys/vm/swappiness set to 0.
- ✓ No performance concerns with Transparent Huge Pages settings.
- ✓ CDH 5 Hue Python version dependency is satisfied.
- ✓ 0 hosts are running CDH 4 and 6 hosts are running CDH5.
- ✓ All checked hosts in each cluster are running the same version of components.
- ✓ All managed hosts have consistent versions of Java.
- ✓ All checked Cloudera Management Daemons versions are consistent with the server.
- ✓ All checked Cloudera Management Agents versions are consistent with the server.

# CLOUDERA CLUSTER INSTALLATION – CLUSTER SETUP

## Cluster Setup

Choose the CDH 5 services that you want to install on your cluster.

Choose a combination of services to install.

- ☐ **Core Hadoop**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, and Sqoop
- ☐ **Core with HBase**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, Sqoop, and HBase
- ☐ **Core with Impala**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, Sqoop, and Impala
- ☐ **Core with Search**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, Sqoop, and Solr
- ☐ **Core with Spark**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, Sqoop, and Spark
- ☒ **All Services**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, Sqoop, HBase, Impala, Solr, Spark, and Key-Value Store Indexer  
**Note:** Please ensure that you have the appropriate license for **Cloudera Impala**, **Cloudera Search**, **HBase**, and **Spark** or contact Cloudera for assistance.
- ☐ **Custom Services**  
Choose your own services. Services required by chosen services will automatically be included. Flume can be added after your initial cluster has been set up.

This wizard will also install the **Cloudera Management Service**. These are a set of components that enable monitoring, reporting, events, and alerts; these components require databases to store information, which will be configured on the next page.

☒ **Include Cloudera Navigator**

- Please ensure that you have the appropriate license for **Cloudera Navigator** or contact Cloudera for assistance.
- Some actions (for example, those performed by specific, internal users) are discarded by default. You can review and update these settings in the Audit Event Filter for each audited service.

# CLOUDERA CLUSTER INSTALLATION – INSTALLED COMPONENT LIST

## Version Summary

### Cluster 1 — CDH 5

#### Hosts

bigdata[1-6].localdomain

Component	Version	Release	CDH Version
Bigtop-Tomcat (CDH 5 only)	0.7.0+cdh5.4.1+0	1.cdh5.4.1.p0.8	CDH 5
Crunch (CDH 5 only)	0.11.0+cdh5.4.1+46	1.cdh5.4.1.p0.8	CDH 5
Flume NG	1.5.0+cdh5.4.1+121	1.cdh5.4.1.p0.8	CDH 5
MapReduce 1	2.6.0+cdh5.4.1+565	1.cdh5.4.1.p0.8	CDH 5
Hadoop	2.6.0+cdh5.4.1+565	1.cdh5.4.1.p0.8	CDH 5
HDFS	2.6.0+cdh5.4.1+565	1.cdh5.4.1.p0.8	CDH 5
HttpFS	2.6.0+cdh5.4.1+565	1.cdh5.4.1.p0.8	CDH 5
hadoop-kms	2.6.0+cdh5.4.1+565	1.cdh5.4.1.p0.8	CDH 5
MapReduce 2	2.6.0+cdh5.4.1+565	1.cdh5.4.1.p0.8	CDH 5
YARN	2.6.0+cdh5.4.1+565	1.cdh5.4.1.p0.8	CDH 5
HBase	1.0.0+cdh5.4.1+141	1.cdh5.4.1.p0.8	CDH 5
Lily HBase Indexer	1.5+cdh5.4.1+50	1.cdh5.4.1.p0.8	CDH 5
Hive	1.1.0+cdh5.4.1+126	1.cdh5.4.1.p0.8	CDH 5
HCatalog	1.1.0+cdh5.4.1+126	1.cdh5.4.1.p0.8	CDH 5
Hue	3.7.0+cdh5.4.1+1173	1.cdh5.4.1.p0.8	CDH 5
Impala	2.2.0+cdh5.4.1+0	1.cdh5.4.1.p0.8	CDH 5
Kite (CDH 5 only)	1.0.0+cdh5.4.1+17	1.cdh5.4.1.p0.8	CDH 5
Llama (CDH 5 only)	1.0.0+cdh5.4.1+0	1.cdh5.4.1.p0.8	CDH 5
Mahout	0.9+cdh5.4.1+21	1.cdh5.4.1.p0.8	CDH 5
Oozie	4.1.0+cdh5.4.1+142	1.cdh5.4.1.p0.8	CDH 5
Parquet	1.5.0+cdh5.4.1+93	1.cdh5.4.1.p0.8	CDH 5
Pig	0.12.0+cdh5.4.1+56	1.cdh5.4.1.p0.8	CDH 5
sentry	1.4.0+cdh5.4.1+163	1.cdh5.4.1.p0.8	CDH 5
Solr	4.10.3+cdh5.4.1+251	1.cdh5.4.1.p0.8	CDH 5

# CLOUDERA CLUSTER INSTALLATION – COMPONENT LIST

## Cluster Setup

Choose the CDH 5 services that you want to install on your cluster.

Choose a combination of services to install.

- ☐ **Core Hadoop**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, and Sqoop
- ☐ **Core with HBase**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, Sqoop, and HBase
- ☐ **Core with Impala**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, Sqoop, and Impala
- ☐ **Core with Search**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, Sqoop, and Solr
- ☐ **Core with Spark**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, Sqoop, and Spark
- ☒ **All Services**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, Sqoop, HBase, Impala, Solr, Spark, and Key-Value Store Indexer  
**Note:** Please ensure that you have the appropriate license for **Cloudera Impala**, **Cloudera Search**, **HBase**, and **Spark** or contact Cloudera for assistance.
- ☐ **Custom Services**  
Choose your own services. Services required by chosen services will automatically be included. Flume can be added after your initial cluster has been set up.

This wizard will also install the **Cloudera Management Service**. These are a set of components that enable monitoring, reporting, events, and alerts; these components require databases to store information, which will be configured on the next page.

☐ **Include Cloudera Navigator**

⏪ Back

1 2 3 4 5 6

⏩ Continue

# CLOUDERA CLUSTER INSTALLATION – ROLE CUSTOMIZATION

## Cluster Setup

### Customize Role Assignments

You can customize the role assignments for your new cluster here, but if assignments are made incorrectly, such as assigning too many roles to a single host, this can impact the performance of your services. Cloudera does not recommend altering assignments unless you have specific requirements, such as having pre-selected a specific host for a specific role.

You can also view the role assignments by host. [View By Host](#)

#### HBase

<b>M</b> Master × 5 New bigdata[2-6].localdomain ▼	<b>HBRS</b> HBase REST Server × 3 New bigdata[2, 4, 6].localdomain ▼	<b>HBTS</b> HBase Thrift Server × 3 New bigdata[2, 4, 6].localdomain ▼	<b>RS</b> RegionServer × 5 New Same As DataNode ▼
---	---	---	--

#### HDFS

<b>NN</b> NameNode × 1 New bigdata1.localdomain	<b>SNN</b> SecondaryNameNode × 1 New bigdata2.localdomain ▼	<b>B</b> Balancer × 1 New bigdata3.localdomain ▼	<b>HFS</b> HttpFS × 5 New bigdata[2-6].localdomain ▼
<b>NFSG</b> NFS Gateway × 1 New bigdata6.localdomain ▼	<b>DN</b> DataNode × 5 New bigdata[2-6].localdomain ▼		

#### Hive

<b>G</b> Gateway × 6 New bigdata[1-6].localdomain	<b>HMS</b> Hive Metastore Server × 1 New bigdata2.localdomain ▼	<b>WHCS</b> WebHCat Server × 3 New bigdata[2, 4, 6].localdomain ▼	<b>HS2</b> HiveServer2 × 2 New bigdata[3, 5].localdomain ▼
--	--	--	---

# CLOUDERA CLUSTER INSTALLATION – ROLE CUSTOMIZATION

## 3 Hosts Selected

Select hosts for a new or existing role. The host list is filtered to remove hosts that are not valid candidates; these include hosts that are unhealthy, members of other clusters, and/or have an incompatible version of CDH installed on them.

<input type="checkbox"/>	Host Name	IP Address	Rack	Cores	Physical Memory	Existing Roles	Added Roles
<input checked="" type="checkbox"/>	bigdata1.localdomain	192.168.0.131	/default	2	5.8 GiB		NN G HS ICS ISS SM HM ES AP NAS NMS OS SS HS G SZS RM JHS S
<input type="checkbox"/>	bigdata2.localdomain	192.168.0.132	/default	2	5.8 GiB		M RS HBRES HBTS DN SNN HFS G HMS WHCS ID LHEI G NM
<input checked="" type="checkbox"/>	bigdata3.localdomain	192.168.0.133	/default	2	5.8 GiB		M RS DN B HFS G HS2 HS ID G NM S
<input type="checkbox"/>	bigdata4.localdomain	192.168.0.134	/default	2	5.8 GiB		M RS HBRES HBTS DN HFS G WHCS ID G NM
<input checked="" type="checkbox"/>	bigdata5.localdomain	192.168.0.135	/default	2	5.8 GiB		M RS DN HFS G HS2 HS ID AM RM G NM S
<input type="checkbox"/>	bigdata6.localdomain	192.168.0.136	/default	2	5.8 GiB		M RS HBRES HBTS DN HFS NFSG G WHCS ID G NM

Tip: Click the first checkbox, hold down the Shift key and click the last checkbox to select a range.



# CLOUDERA CLUSTER INSTALLATION – DATABASE SETUP

## Database Setup

Configure and test database connections. If using custom databases, create the databases first according to the **Installing and Configuring an External Database** section of the [Installation Guide](#).

- ☐ Use Custom Databases  
☒ Use Embedded Database

When using the embedded database, passwords are automatically generated. Please copy them down.

### Reports Manager

✓ Successful

Currently assigned to run on **bigdata5.localdomain**.

Database Host Name:

bigdata1.localdomain:7432

Database Type:

PostgreSQL ▼

Database Name :

rman

Username:

rman

Password:

zqqo3LfWtf

### Activity Monitor

✓ Successful

Currently assigned to run on **bigdata1.localdomain**.

Database Host Name:

bigdata1.localdomain:7432

Database Type:

PostgreSQL ▼

Database Name :

amon

Username:

amon

Password:

k1ut6qvtwL

### Navigator Audit Server

✓ Successful

Currently assigned to run on **bigdata1.localdomain**.

Database Host Name:

bigdata1.localdomain:7432

Database Type:

PostgreSQL ▼

Database Name :

nav

Username:

nav

Password:

hUX4nykqKF

### Navigator Metadata Server

✓ Successful

# CLOUDERA CLUSTER INSTALLATION – CONFIGURATION CONTD..

## Cluster Setup

### Review Changes

#### HDFS Root Directory

hbase.rootdir

Cluster 1 > HBase (Service-Wide)



/hbase

#### Enable Replication

hbase.replication

Cluster 1 > HBase (Service-Wide) ☒



#### Enable Indexing

Cluster 1 > HBase (Service-Wide) ☒



#### DataNode Data Directory

dfs.data.dir, dfs.datanode.data.dir

[Edit Individual Values](#)

Cluster 1 > DataNode Default Group

...and 3 others



/dfs/dn

#### DataNode Failed Volumes Tolerated

dfs.datanode.failed.volumes.tolerated

[Edit Individual Values](#)

Cluster 1 > DataNode Default Group

...and 3 others



0

#### NameNode Data Directories

dfs.name.dir, dfs.namenode.name.dir

Cluster 1 > NameNode Default Group



/dfs/nn

#### HDFS Checkpoint Directories

fs.checkpoint.dir,  
dfs.namenode.checkpoint.dir

Cluster 1 > SecondaryNameNode Default Group



/dfs/snn

#### Hive Warehouse Directory

hive.metastore.warehouse.dir

Cluster 1 > Hive (Service-Wide)



/user/hive/warehouse

# CLOUDERA CLUSTER INSTALLATION – CONFIGURATION CONTD..

<b>Hive Metastore Server Port</b> hive.metastore.port	Cluster 1 > Hive Metastore Server Default Group <input type="text" value="9083"/>
<b>Impala Daemon Scratch Directories</b> scratch_dirs <a href="#">Edit Individual Values</a>	Cluster 1 > Impala Daemon Default Group ...and 3 others <div><input type="text" value="/impala/impalad"/> <span>+</span> <span>-</span></div>
<b>Alerts: Mail Server Hostname</b>	Alert Publisher Default Group <input type="text" value="localhost"/>
<b>Alerts: Mail Server Username</b>	Alert Publisher Default Group <input type="text"/>
<b>Alerts: Mail Server Password</b>	Alert Publisher Default Group <input type="password"/>
<b>Alerts: Mail Message Recipients</b>	Alert Publisher Default Group <input type="text" value="root@localhost"/>
<b>Host Monitor Storage Directory</b> firehose.storage.base.directory	Host Monitor Default Group <input type="text" value="/var/lib/cloudera-host-monitor"/>
<b>Service Monitor Storage Directory</b> firehose.storage.base.directory	Service Monitor Default Group <input type="text" value="/var/lib/cloudera-service-monitor"/>
<b>ShareLib Root Directory</b> oozie.service.WorkflowAppService.system.lib path	Cluster 1 > Oozie (Service-Wide) <input type="text" value="/user/oozie"/>
<b>Oozie Server Data Directory</b>	Cluster 1 > Oozie Server Default Group <input type="text" value="/var/lib/oozie/data"/>

# CLOUDERA CLUSTER INSTALLATION – CONFIGURATION CONTD..

<b>Oozie Server Data Directory</b>	Cluster 1 > Oozie Server Default Group <input type="text" value="/var/lib/oozie/data"/>
<b>ZooKeeper Znode</b>	Cluster 1 > Solr (Service-Wide) <input type="text" value="/solr"/>
<b>HDFS Data Directory</b>	Cluster 1 > Solr (Service-Wide) <input type="text" value="/solr"/>
<b>Sqoop 2 Server Metastore Directory</b>	Cluster 1 > Sqoop 2 Server Default Group <input type="text" value="/var/lib/sqoop2"/>
<b>Sqoop Repository Database Type</b>	Cluster 1 > Sqoop 2 Server Default Group <input type="radio"/> postgresql <input checked="" type="radio"/> derby
<b>Sqoop Repository Database Host</b>	Cluster 1 > Sqoop 2 Server Default Group <input type="text" value="localhost"/>
<b>Sqoop Repository Database Name</b>	Cluster 1 > Sqoop 2 Server Default Group <input type="text" value="sqoop"/>
<b>Sqoop Repository Database User</b> org.apache.sqoop.repository.jdbc.user	Cluster 1 > Sqoop 2 Server Default Group <input type="text" value="sa"/>
<b>Sqoop Repository Database Password</b> org.apache.sqoop.repository.jdbc.password	Cluster 1 > Sqoop 2 Server Default Group <input type="text"/>
<b>NodeManager Local Directories</b> yarn.nodemanager.local-dirs <a href="#">Edit Individual Values</a>	Cluster 1 > NodeManager Default Group ...and 3 others <div><input type="text" value="/yarn/nm"/> <span>+</span> <span>-</span></div>

# CLOUDERA CLUSTER INSTALLATION – CLIENT CONFIGURATION

## Command Details: Deploy Client Configuration

Command	Context	Status
✓ <b>Deploy Client Configuration</b>	<a href="#">Cluster 1</a>	Finished

Successfully deployed all client configurations.

## Command Progress

Completed 6 of 6 steps.

- ✓ Execute command deployClientConfig on service HDFS  
Successfully deployed client configuration.  
[Details](#) ↗
- ✓ Execute command deployClientConfig on service Solr  
Successfully deployed client configuration.  
[Details](#) ↗
- ✓ Execute command deployClientConfig on service HBase  
Successfully deployed client configuration.  
[Details](#) ↗
- ✓ Execute command deployClientConfig on service YARN (MR2 Included)  
Successfully deployed client configuration.  
[Details](#) ↗
- ✓ Execute command deployClientConfig on service Spark  
Successfully deployed client configuration.  
[Details](#) ↗
- ✓ Execute command deployClientConfig on service Hive  
Successfully deployed client configuration.  
[Details](#) ↗

# CLOUDERA CLUSTER INSTALLATION – CLIENT CONFIGURATION

## Cluster Setup

### Progress

Command	Context	Status	Started at	Ended at
✓ First Run		Finished	May 17, 2015 7:47:11 PM EDT	May 17, 2015 8:09:56 PM EDT

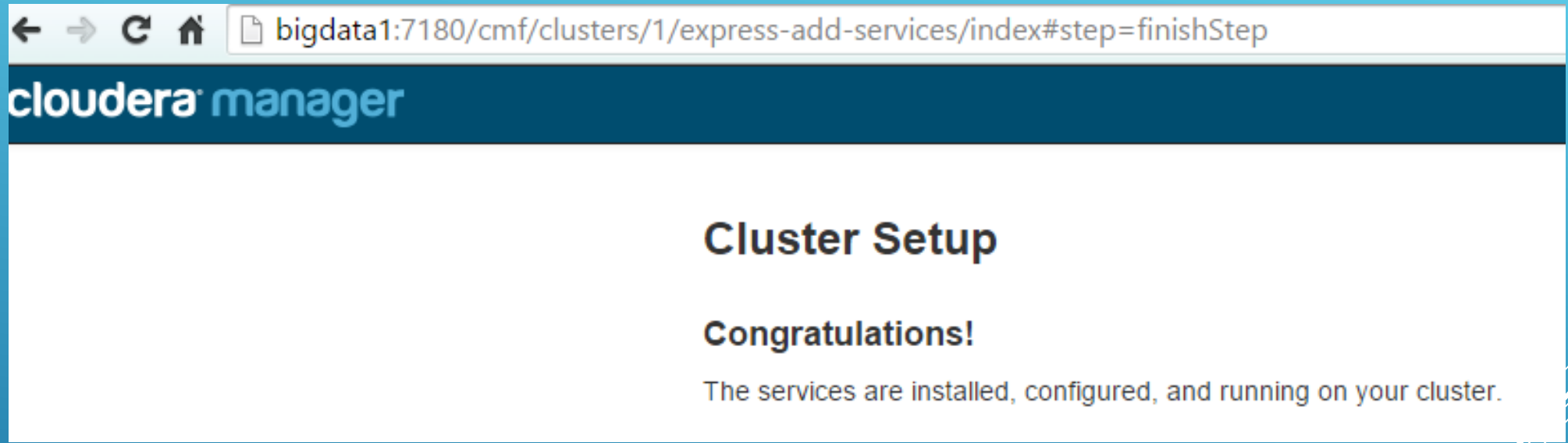
Finished First Run of all services successfully.

### Command Progress

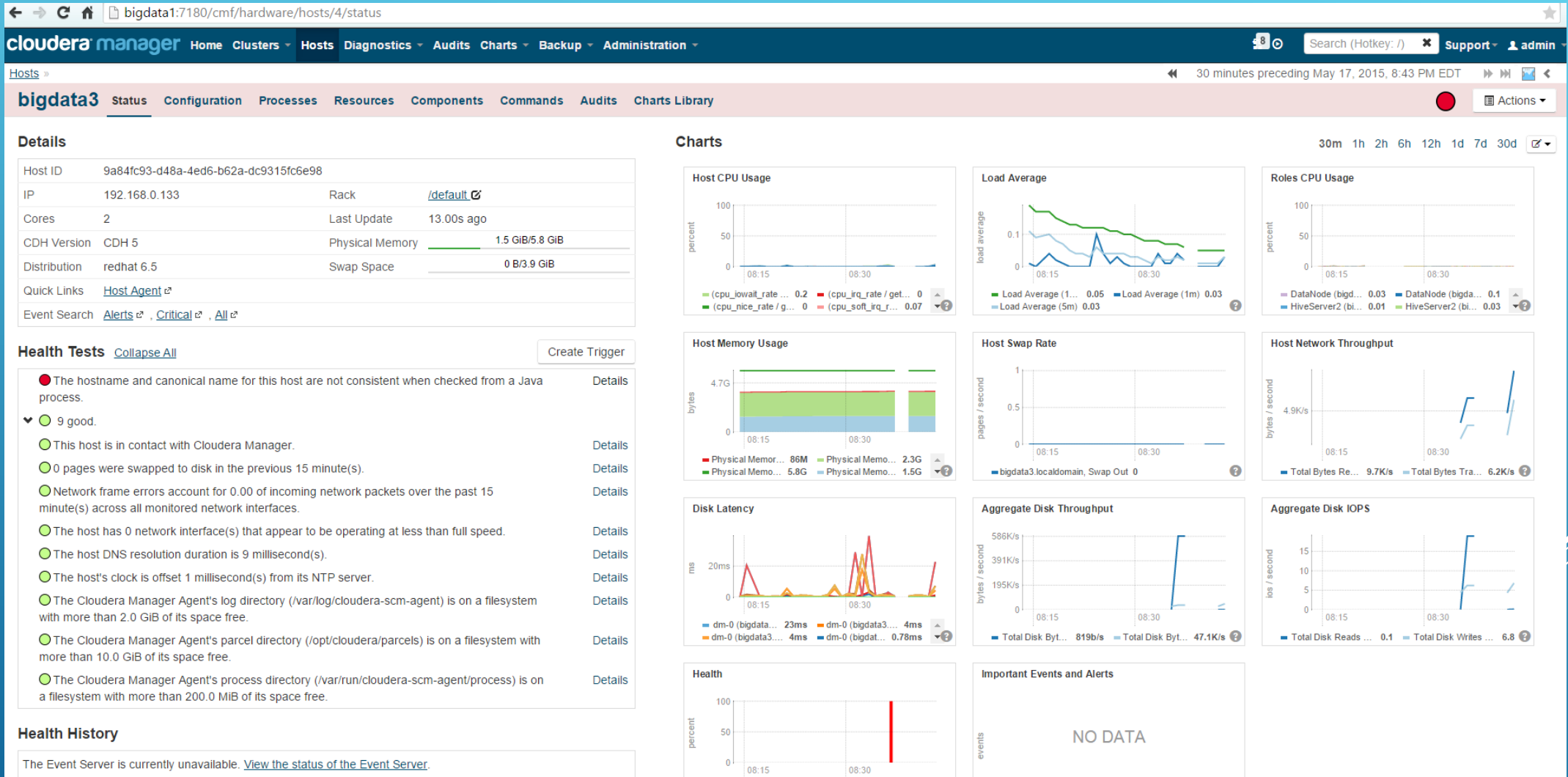
Completed 35 of 35 steps.



# CLOUDERA CLUSTER INSTALLATION COMPLETE



# CLUSTERA MANAGER CONSOLE

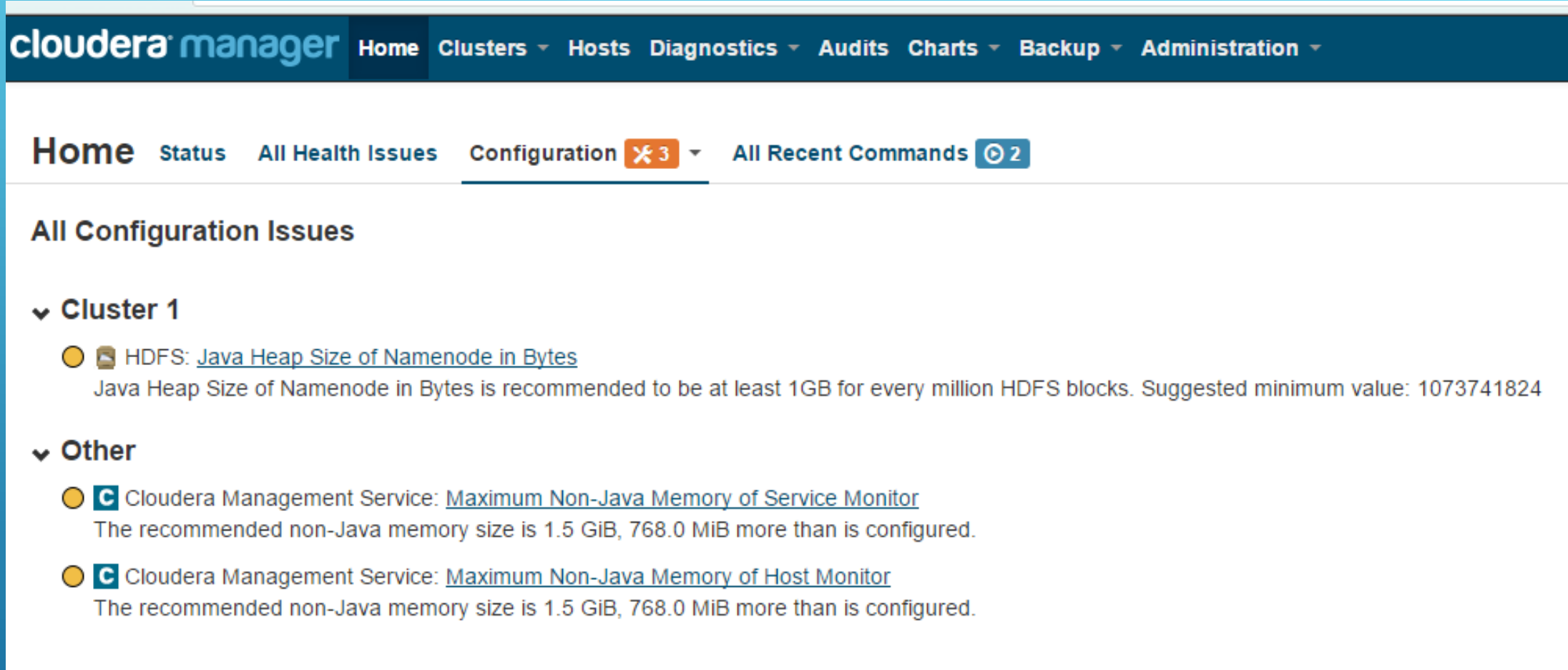




# HADOOP COMPONENTS ON CLUDERA HADOOP CLUSTER

```
7.0.0_60/bin/jps
[root@bigdata1 ~]# ./dcli -g mynodes /usr/java/jdk1.7.0_60/bin/jps
bigdata1: 9499 Jps
bigdata1: 7382 AlertPublisher
bigdata1: 30373 QuorumPeerMain
bigdata1: 3495 Bootstrap
bigdata1: 7426 Main
bigdata1: 1012 HistoryServer
bigdata1: 5321
bigdata1: 32447 JobHistoryServer
bigdata1: 4764 RunJar
bigdata1: 7481 Main
bigdata1: 7454 NavServer
bigdata1: 32634 ResourceManager
bigdata1: 9517 EventCatcherService
bigdata1: 7064 Bootstrap
bigdata1: 7540 Main
bigdata1: 30533 NameNode
bigdata1: 9302 NavigatorMain
bigdata1: 31637 Bootstrap
bigdata1: 20899 Main
bigdata2: 31283 RunJar
bigdata2: 28522 HMaster
bigdata2: 27879 SecondaryNameNode
bigdata2: 31718
bigdata2: 28569 RESTServer
bigdata2: 28451 ThriftServer
bigdata2: 6531 Jps
bigdata2: 27956 Bootstrap
bigdata2: 27865 DataNode
bigdata2: 29683 NodeManager
bigdata2: 28470 HRegionServer
bigdata3: 12561 QuorumPeerMain
bigdata3: 12682 Bootstrap
bigdata3: 12793 DataNode
bigdata3: 13085 HMaster
bigdata3: 15696
bigdata3: 15214 RunJar
bigdata3: 13977 NodeManager
bigdata3: 21710 Jps
bigdata3: 13066 HRegionServer
bigdata4: 13458 NodeManager
bigdata4: 15111
bigdata4: 12338 HMaster
bigdata4: 11908 DataNode
bigdata4: 12493 HRegionServer
bigdata4: 12434 ThriftServer
bigdata4: 12380 RESTServer
bigdata4: 11933 Bootstrap
bigdata4: 21080 Jps
bigdata4: 14713 RunJar
bigdata5: 13406 QuorumPeerMain
bigdata5: 16203 RunJar
bigdata5: 14947 NodeManager
bigdata5: 13916 HMaster
bigdata5: 13952 HRegionServer
bigdata5: 16637
```

# CLOUDERA HADOOP CLUSTER – POST DEPLOYMENT





The screenshot shows the Cloudera Manager web interface. The top navigation bar includes 'cloudera manager' and links for 'Home', 'Clusters', 'Hosts', 'Diagnostics', 'Audits', 'Charts', 'Backup', and 'Administration'. Below this, a secondary navigation bar shows 'Home', 'Status', 'All Health Issues', 'Configuration' (with a red icon and '3'), and 'All Recent Commands' (with a blue icon and '2'). The main content area is titled 'All Configuration Issues'. Under the 'Cluster 1' section, there is a yellow warning icon and a link to 'HDFS: Java Heap Size of Namenode in Bytes', with a note that the value should be at least 1GB for every million HDFS blocks. Under the 'Other' section, there are two items, each with a blue 'C' icon and a yellow warning icon, both pointing to 'Maximum Non-Java Memory of Service Monitor' and 'Maximum Non-Java Memory of Host Monitor' respectively, with a note that the recommended non-Java memory size is 1.5 GiB, 768.0 MiB more than is configured.

**cloudera manager** Home Clusters Hosts Diagnostics Audits Charts Backup Administration





Home Status All Health Issues **Configuration** 3 All Recent Commands 2

## All Configuration Issues

### Cluster 1

-   HDFS: [Java Heap Size of Namenode in Bytes](#)  
Java Heap Size of Namenode in Bytes is recommended to be at least 1GB for every million HDFS blocks. Suggested minimum value: 1073741824

### Other

-   Cloudera Management Service: [Maximum Non-Java Memory of Service Monitor](#)  
The recommended non-Java memory size is 1.5 GiB, 768.0 MiB more than is configured.
-   Cloudera Management Service: [Maximum Non-Java Memory of Host Monitor](#)  
The recommended non-Java memory size is 1.5 GiB, 768.0 MiB more than is configured.

# CLOUDERA HADOOP CLUSTER – POST DEPLOYMENT CONTD

**cloudera manager** Home Clusters Hosts Diagnostics Audits Charts Backup Administration

8 Search (Hotkey: /) Support admin

Cluster 1 »

**HDFS** Status Instances Configuration Commands Audits File Browser Charts Actions

Configuration Selected Filters: namenode\_java\_heapsize Switch to the classic layout Role Groups History and Rollback

Filters

SEARCH

namenode\_java\_heapsize

STATUS

All2

Error0

Warning1

Edited2

Non-default0

Has Overrides0

SCOPE

All2

HDFS (Service-Wide)0

Balancer0

Reason for change...

Save Changes

2 Edited Values

Java Heap Size of Namenode in Bytes

NameNode Default Group C

1 GiB

Java Heap Size of Namenode in Bytes is recommended to be at least 1GB for every million HDFS blocks. Suggested minimum value: 1073741824

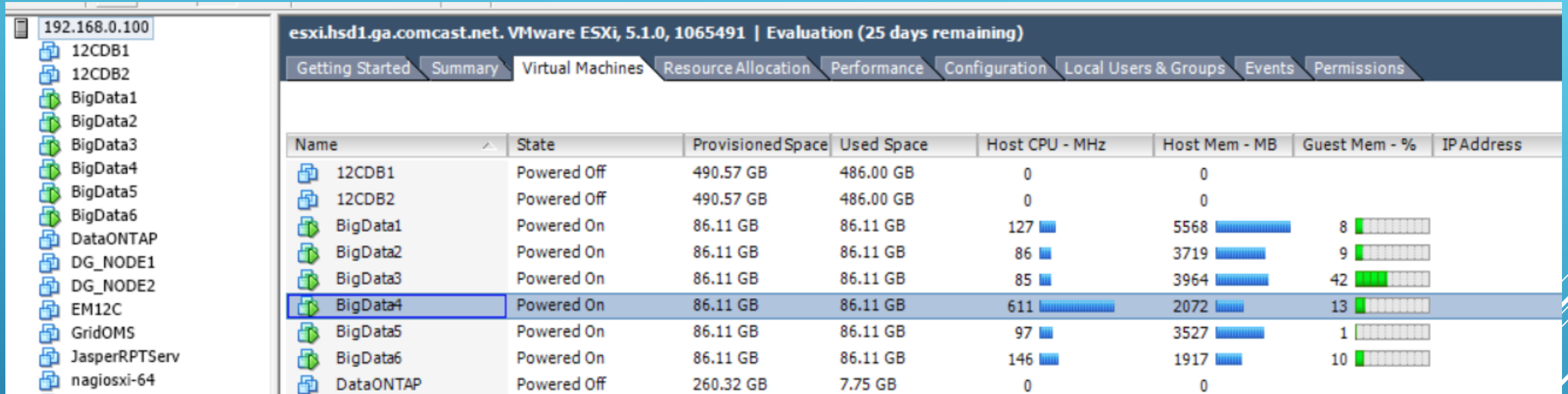
Java Heap Size of Secondary namenode in Bytes

SecondaryNameNode Default Group C

1 GiB

Display 25 Entries

# CLOUDERA HADOOP CLUSTER – VSPHERE CONSOLE



192.168.0.100

esxi.hsd1.ga.comcast.net. VMware ESXi, 5.1.0, 1065491 | Evaluation (25 days remaining)

Getting Started Summary **Virtual Machines** Resource Allocation Performance Configuration Local Users & Groups Events Permissions

Name	State	Provisioned Space	Used Space	Host CPU - MHz	Host Mem - MB	Guest Mem - %	IP Address
12CDB1	Powered Off	490.57 GB	486.00 GB	0	0		
12CDB2	Powered Off	490.57 GB	486.00 GB	0	0		
BigData1	Powered On	86.11 GB	86.11 GB	127	5568	8	
BigData2	Powered On	86.11 GB	86.11 GB	86	3719	9	
BigData3	Powered On	86.11 GB	86.11 GB	85	3964	42	
<b>BigData4</b>	<b>Powered On</b>	<b>86.11 GB</b>	<b>86.11 GB</b>	<b>611</b>	<b>2072</b>	<b>13</b>	
BigData5	Powered On	86.11 GB	86.11 GB	97	3527	1	
BigData6	Powered On	86.11 GB	86.11 GB	146	1917	10	
DataONTAP	Powered Off	260.32 GB	7.75 GB	0	0		