HADOOP CLUSTER SETUP USING CLOUDERA MANAGER

By

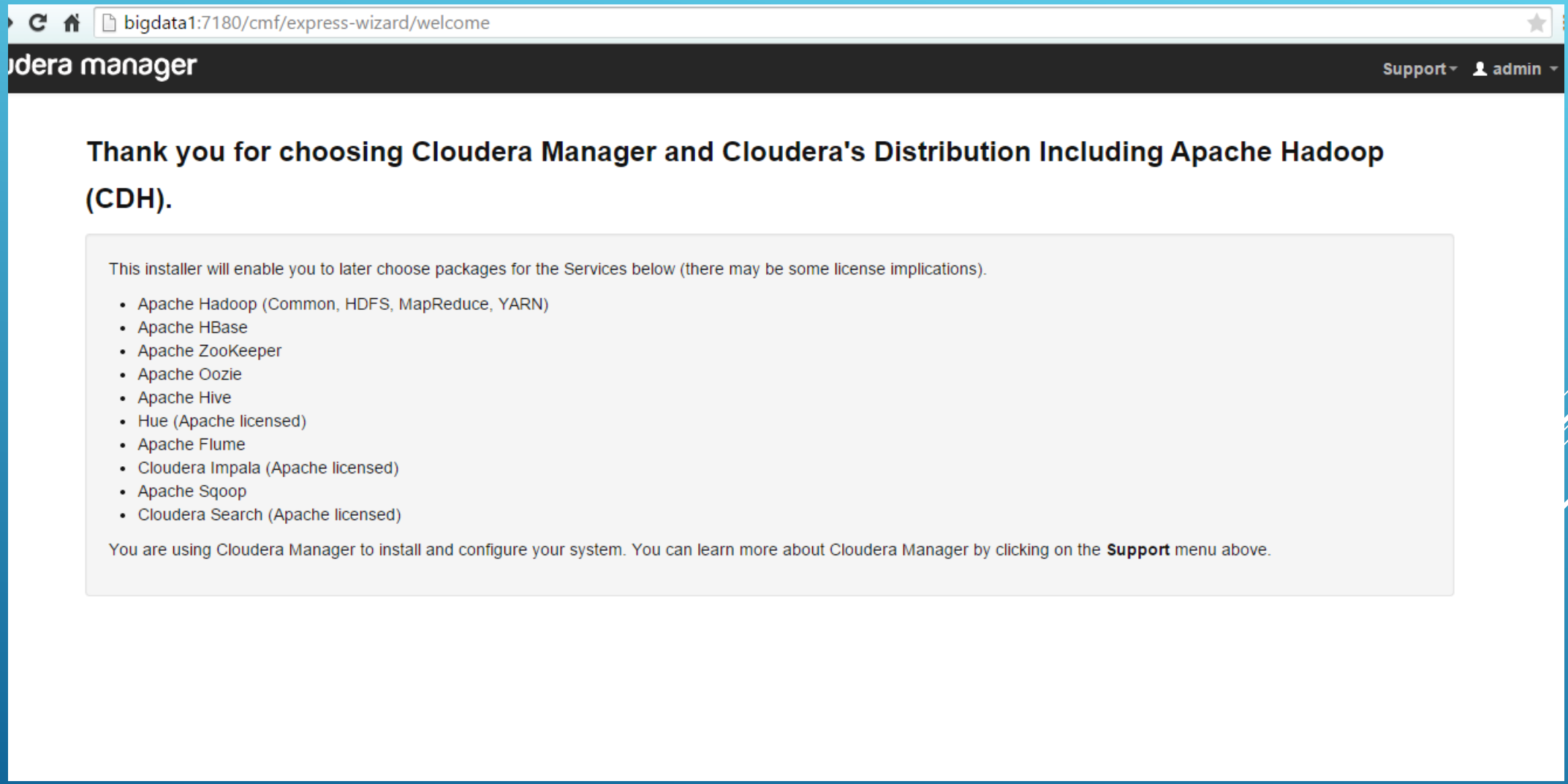
Sudheer Kondla

Solutions Architect

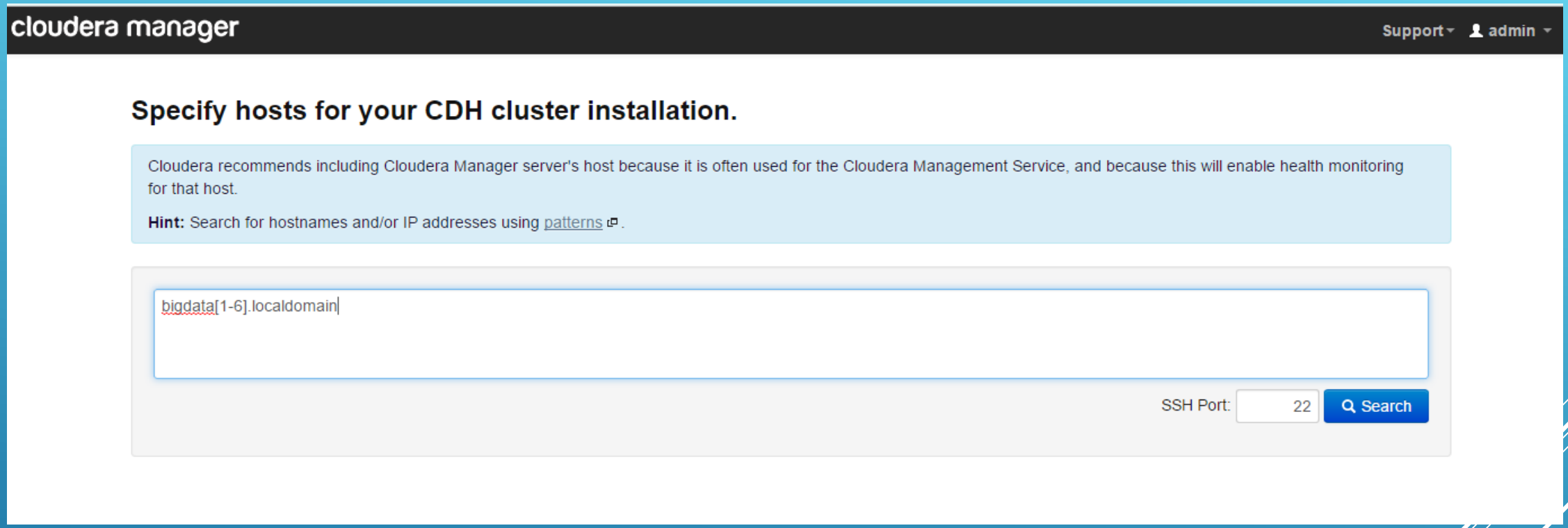
LOGIN TO CLOUDERA MANAGER

- 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

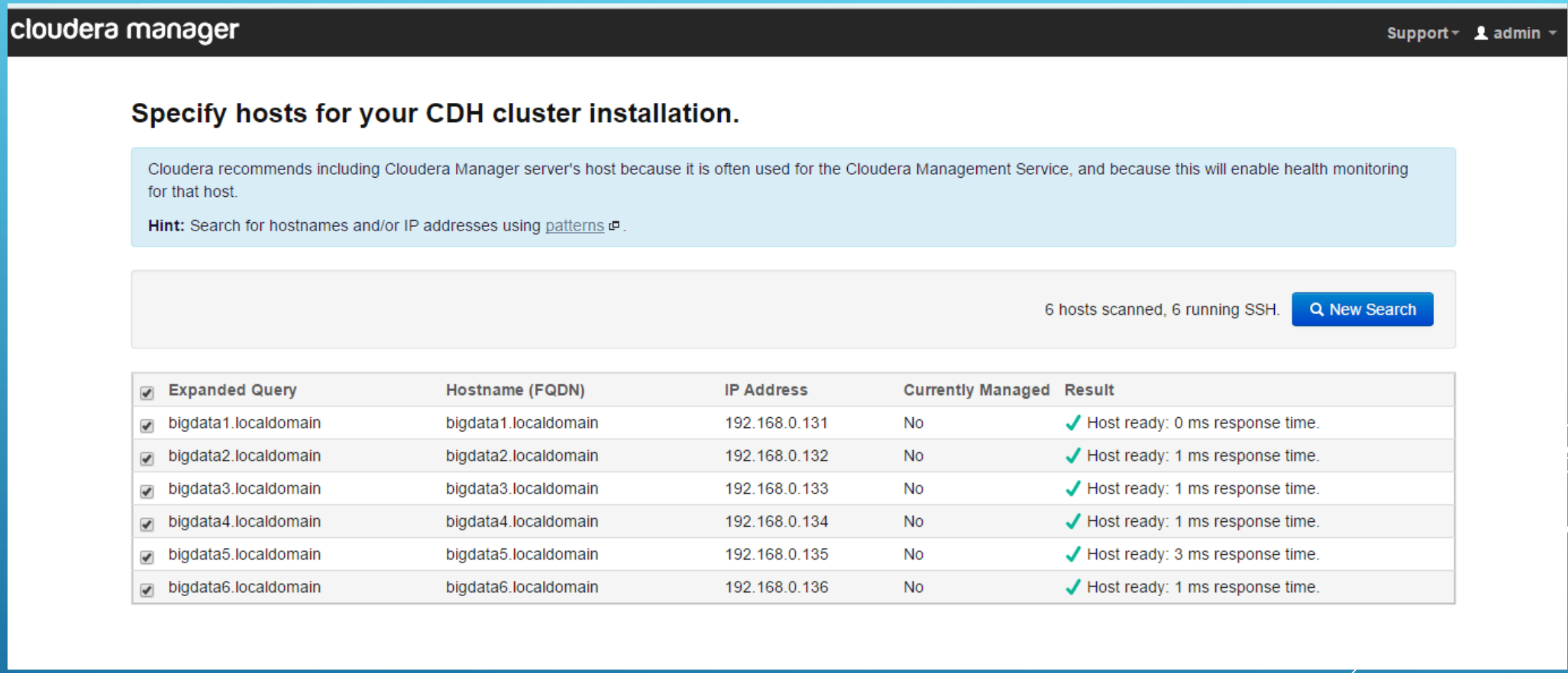
LOGIN TO CLOUDERA MANAGER



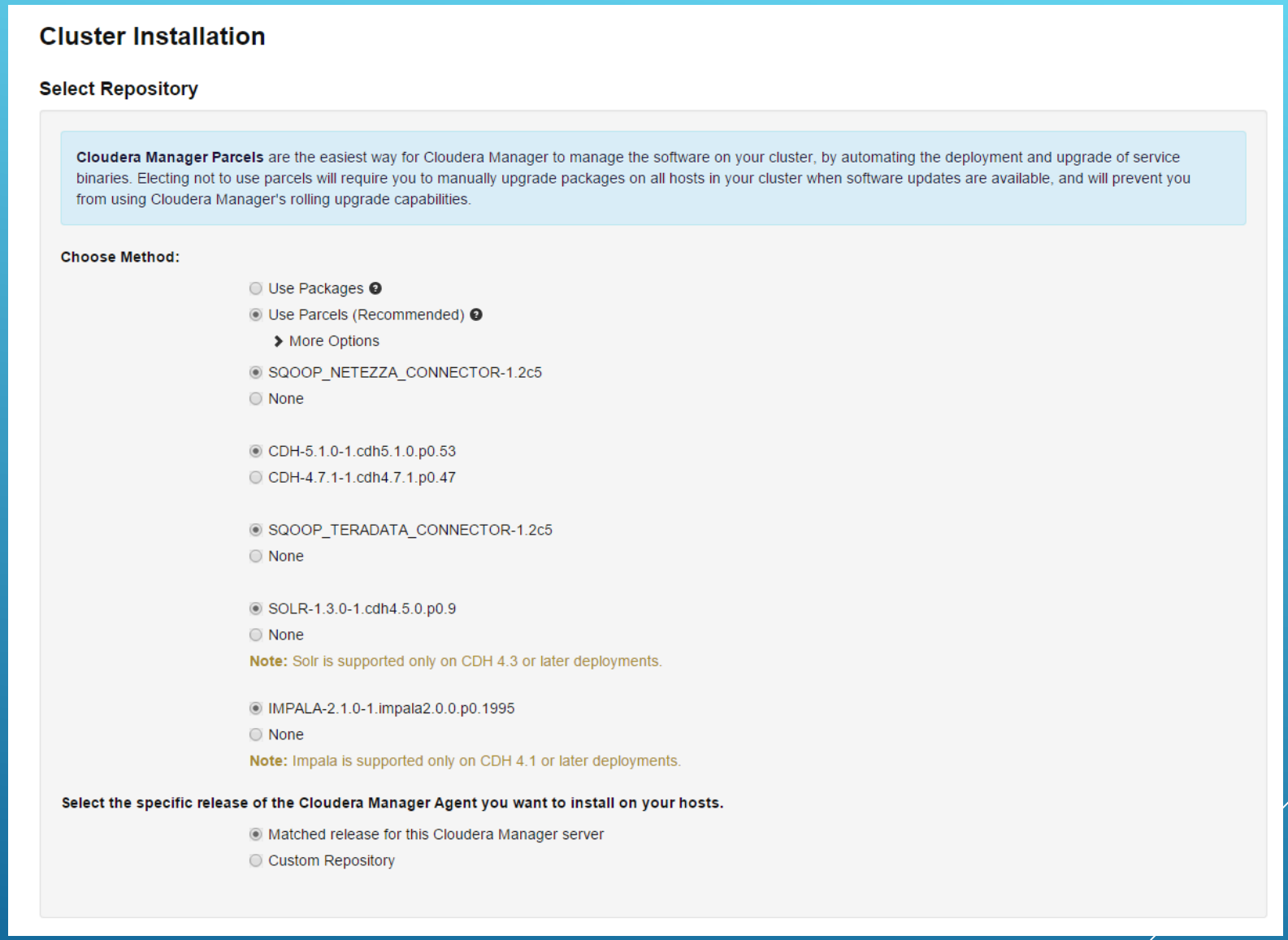
ENTER LIST OF CLUSTER NODES



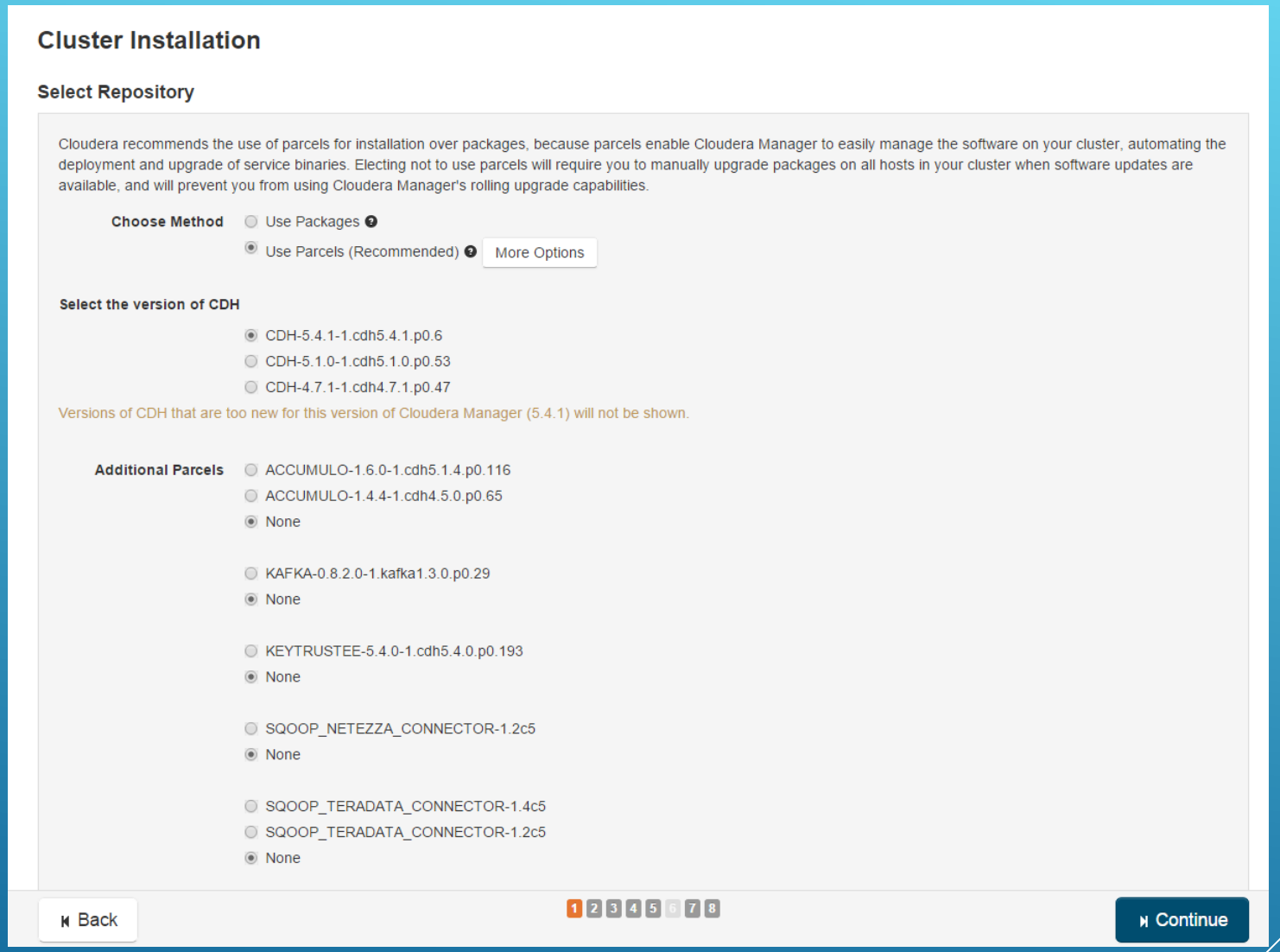
SEARCH CLUSTER NODES ON CLOUDERA MANAGER



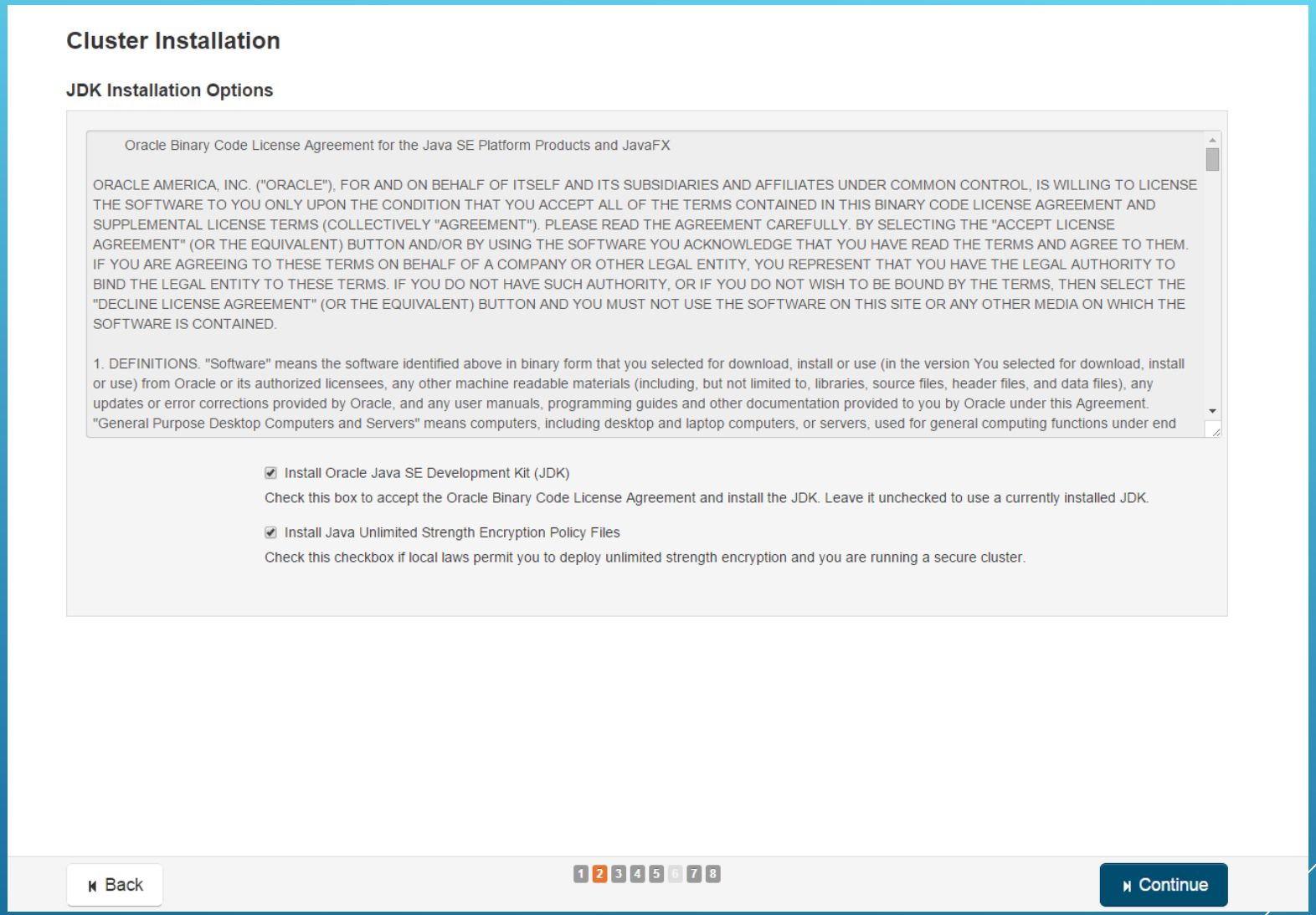
SPECIFY REPOSITORY INFORMATION CLOUDERA MANAGER



SPECIFY REPOSITORY INFORMATION CONTD..



SPECIFY JDK AND ENCRYPTION OPTIONS



SPECIFY SINGLE OR MULTI-USER MODE

Cluster Installation

Enable Single User Mode

Only supported for CDH 5.2 and above.

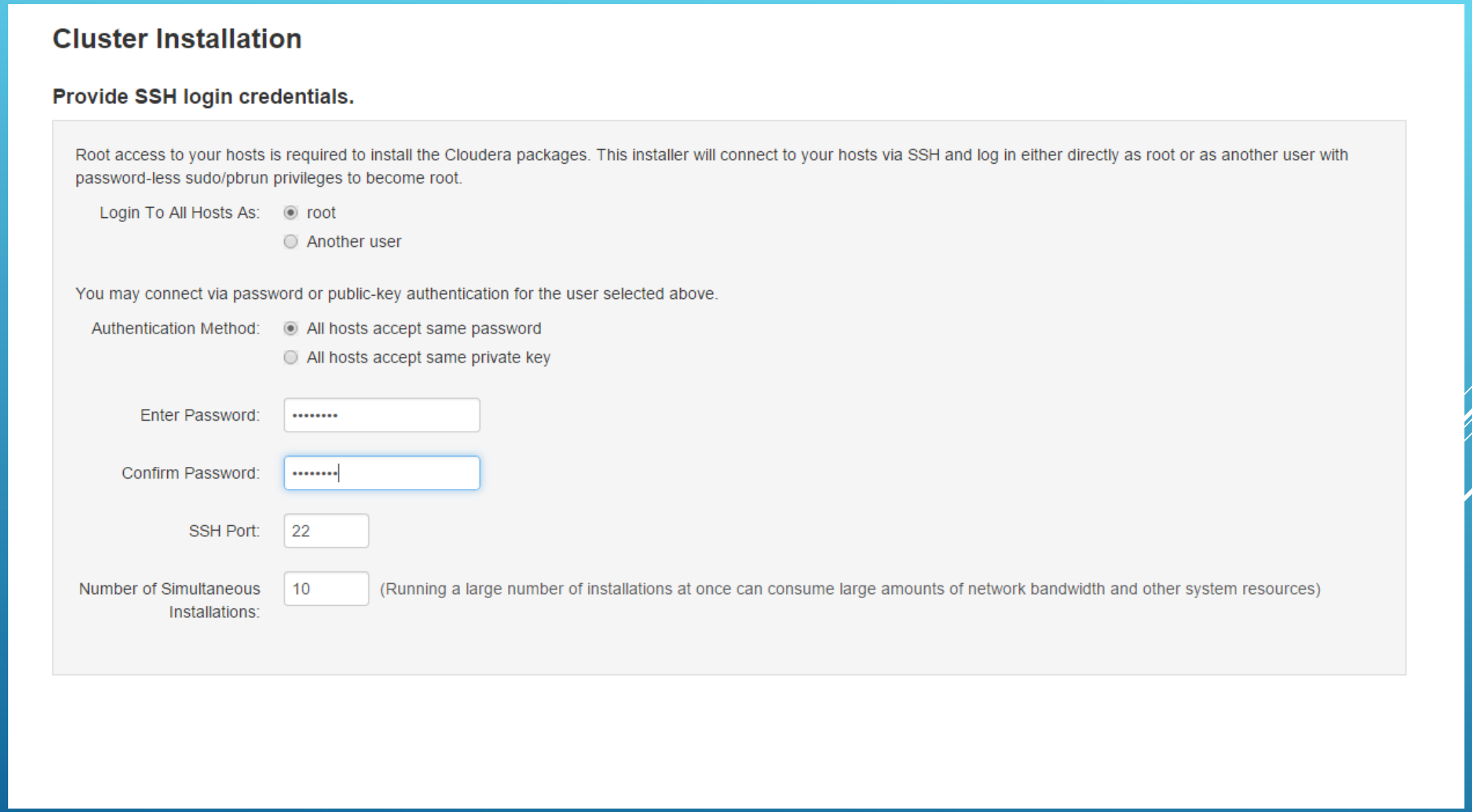
By default, service processes run as distinct users on the system. For example, HDFS DataNodes run as user "hdfs" and HBase RegionServers run as user "hbase." Enabling "single user mode" configures Cloudera Manager to run service processes as a single user, by default "cloudera-scm", thereby prioritizing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

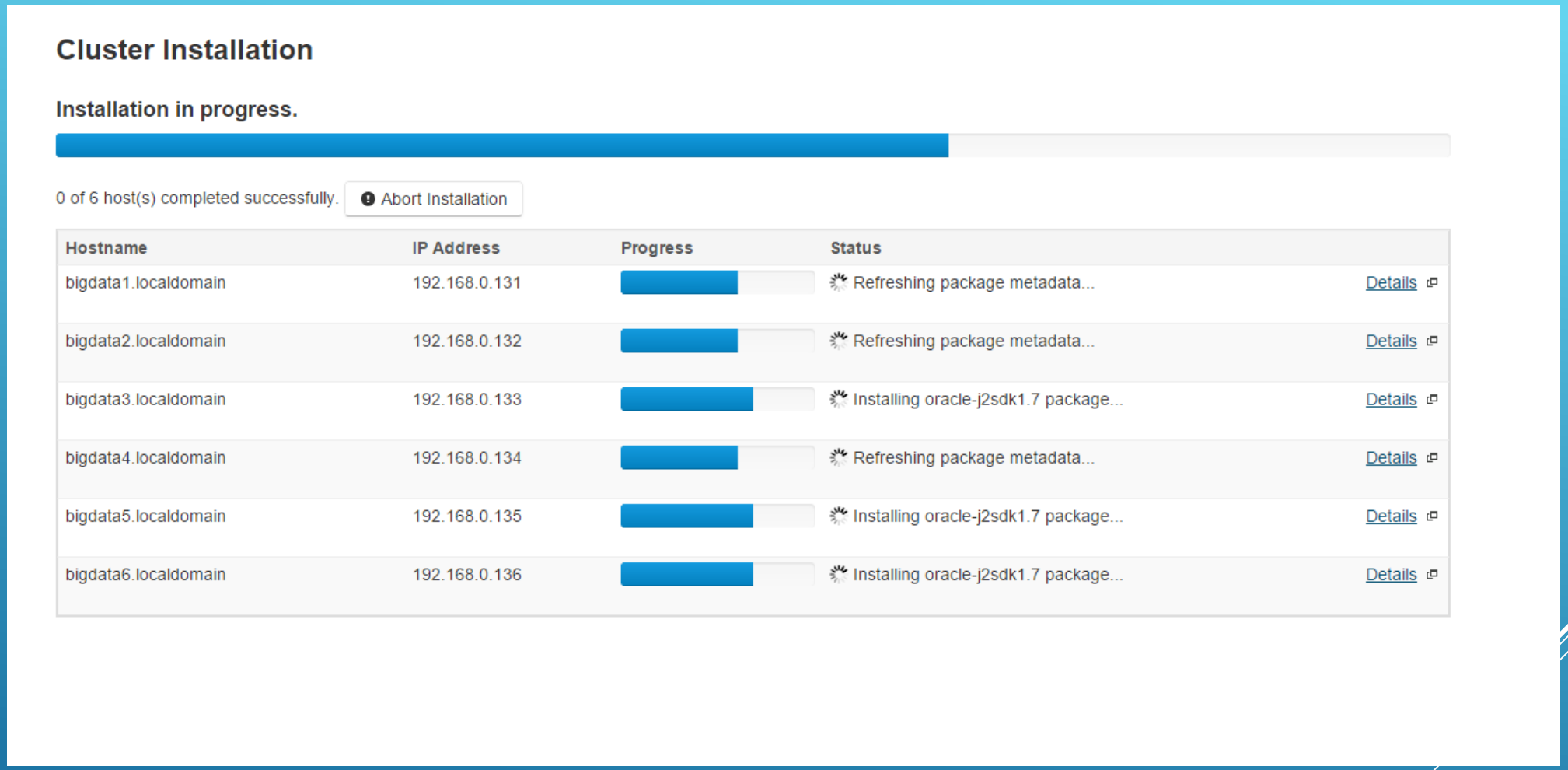
Switching back and forth between single user mode and regular mode is not supported.

Single User Mode ☐ ?

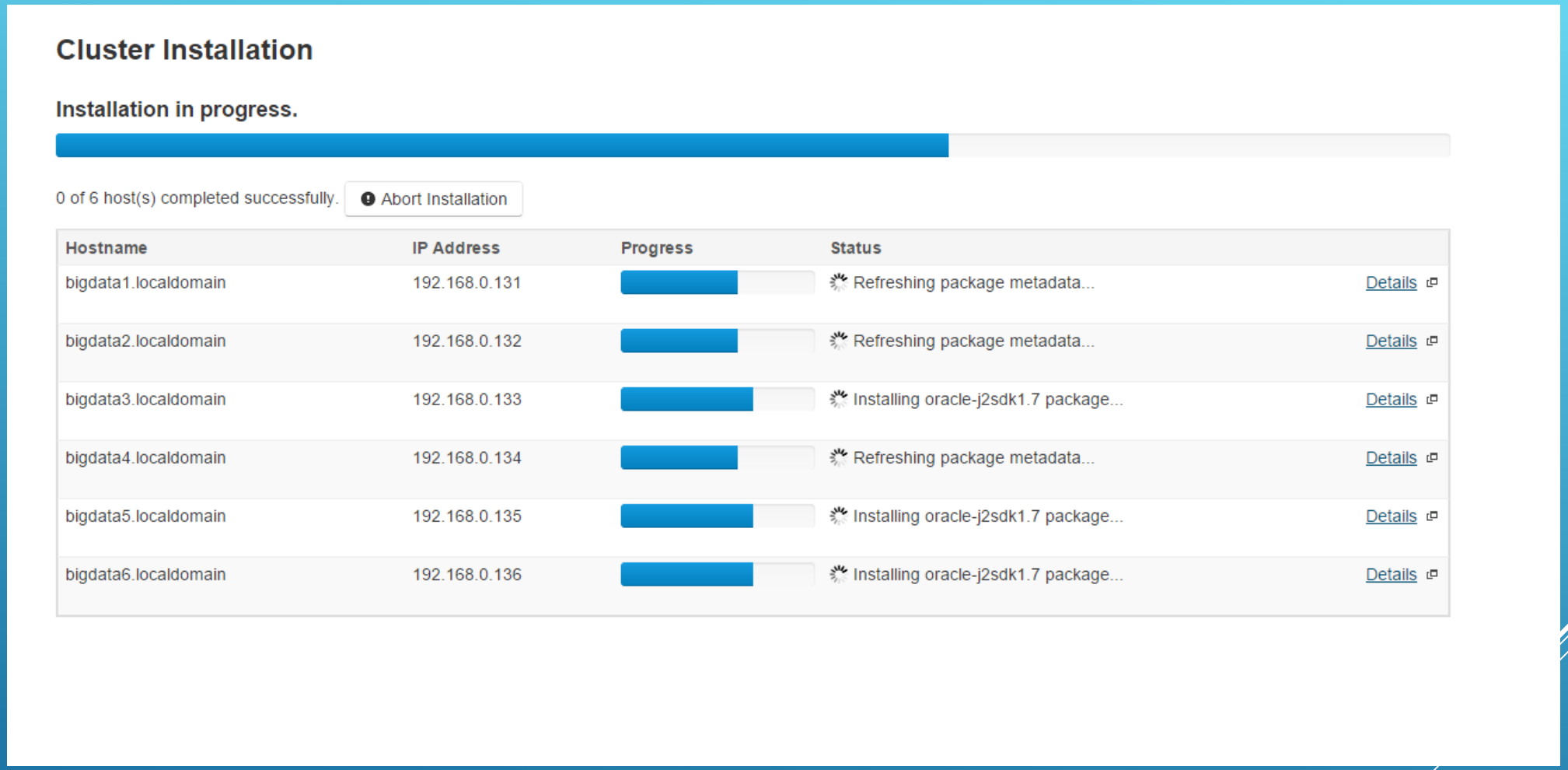
SPECIFY SSH LOGIN CREDENTIALS FOR THE USER



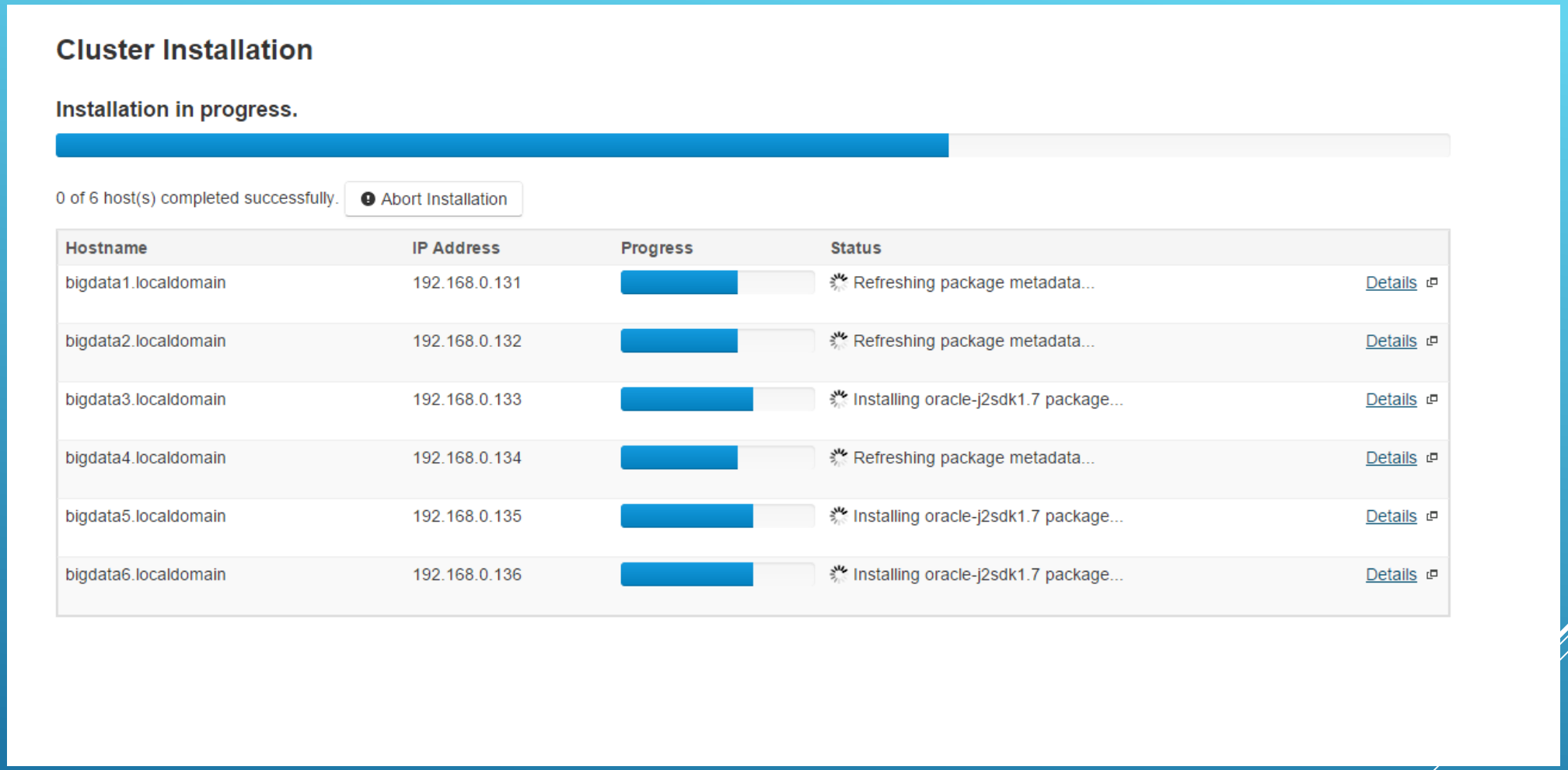
INSTALL SOFTWARE COMPONENTS ON CLUSTER NODES



INSTALL SOFTWARE COMPONENTS ON CLUSTER NODES



INSTALL SOFTWARE COMPONENTS ON CLUSTER NODES – CONTD..



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

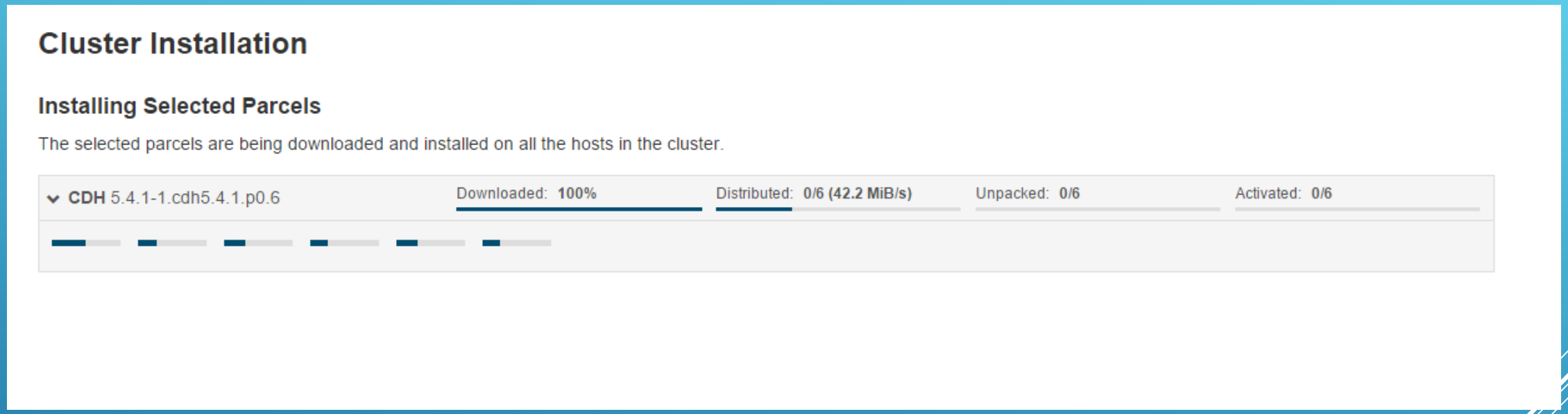
Cluster Installation

Installation completed successfully.

6 of 6 host(s) completed successfully.

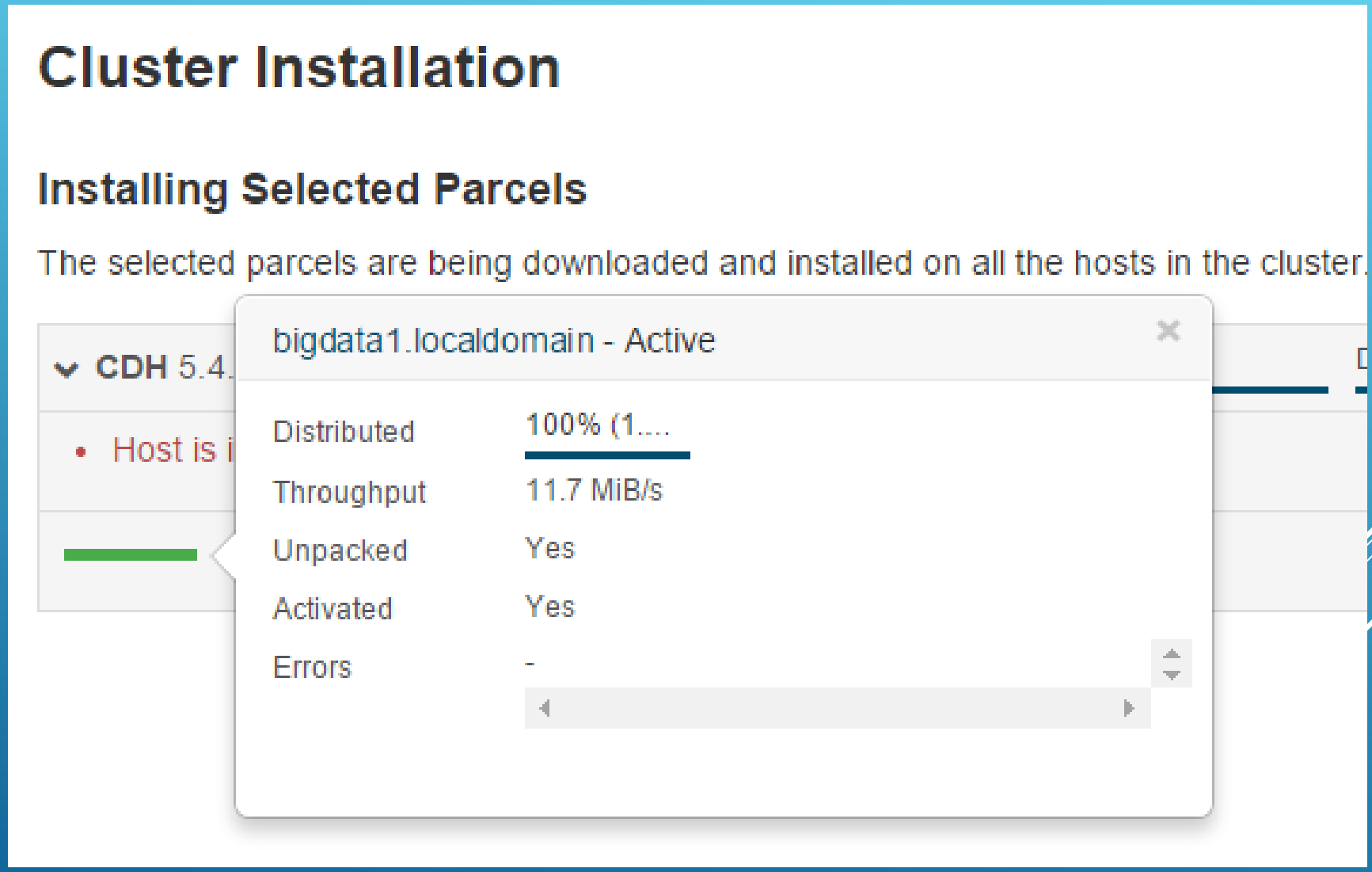
Hostname	IP Address	Progress	Status	
bigdata1.localdomain	192.168.0.131		✓ Installation completed successfully.	Details
bigdata2.localdomain	192.168.0.132		✓ Installation completed successfully.	Details
bigdata3.localdomain	192.168.0.133		✓ Installation completed successfully.	Details
bigdata4.localdomain	192.168.0.134		✓ Installation completed successfully.	Details
bigdata5.localdomain	192.168.0.135		✓ Installation completed successfully.	Details
bigdata6.localdomain	192.168.0.136		✓ Installation completed successfully.	Details

CLOUDERA CLUSTER INSTALLATION – INSTALL PARCELS



CLOUDERA CLUSTER INSTALLATION – INSTALL PARCELS

CLOUDERA CLUSTER INSTALLATION – INSTALL PARCELS



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.

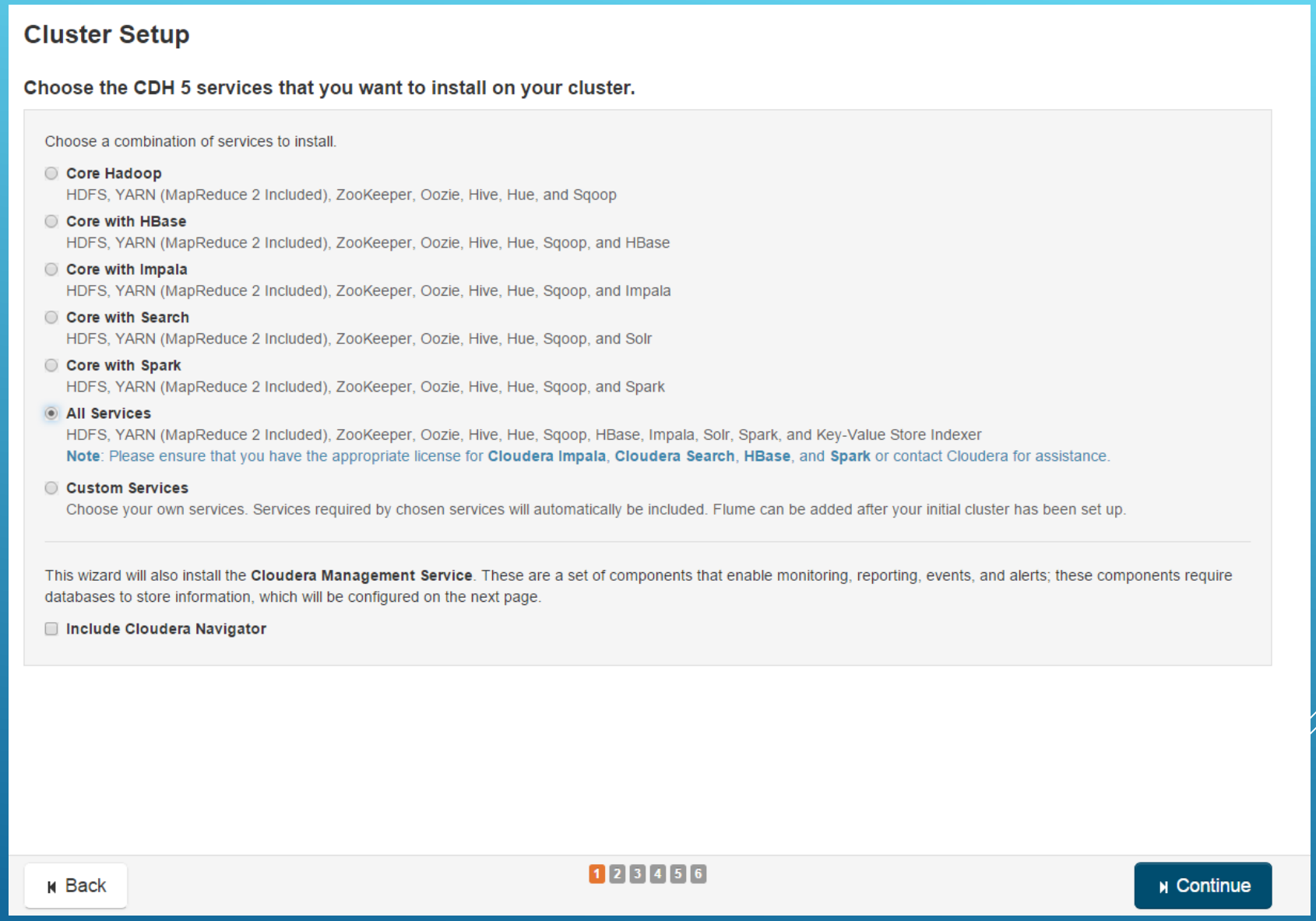
- [x] **Include Cloudera Navigator**
  - Please ensure that you have the appropriate license for **Cloudera Navigator** or contact Cloudera for assistance.
  - Some actions (for example, those performed by specific, internal users) are discarded by default. You can review and update these settings in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

CLOUDERA CLUSTER INSTALLATION – INSTALLED COMPONENT LIST

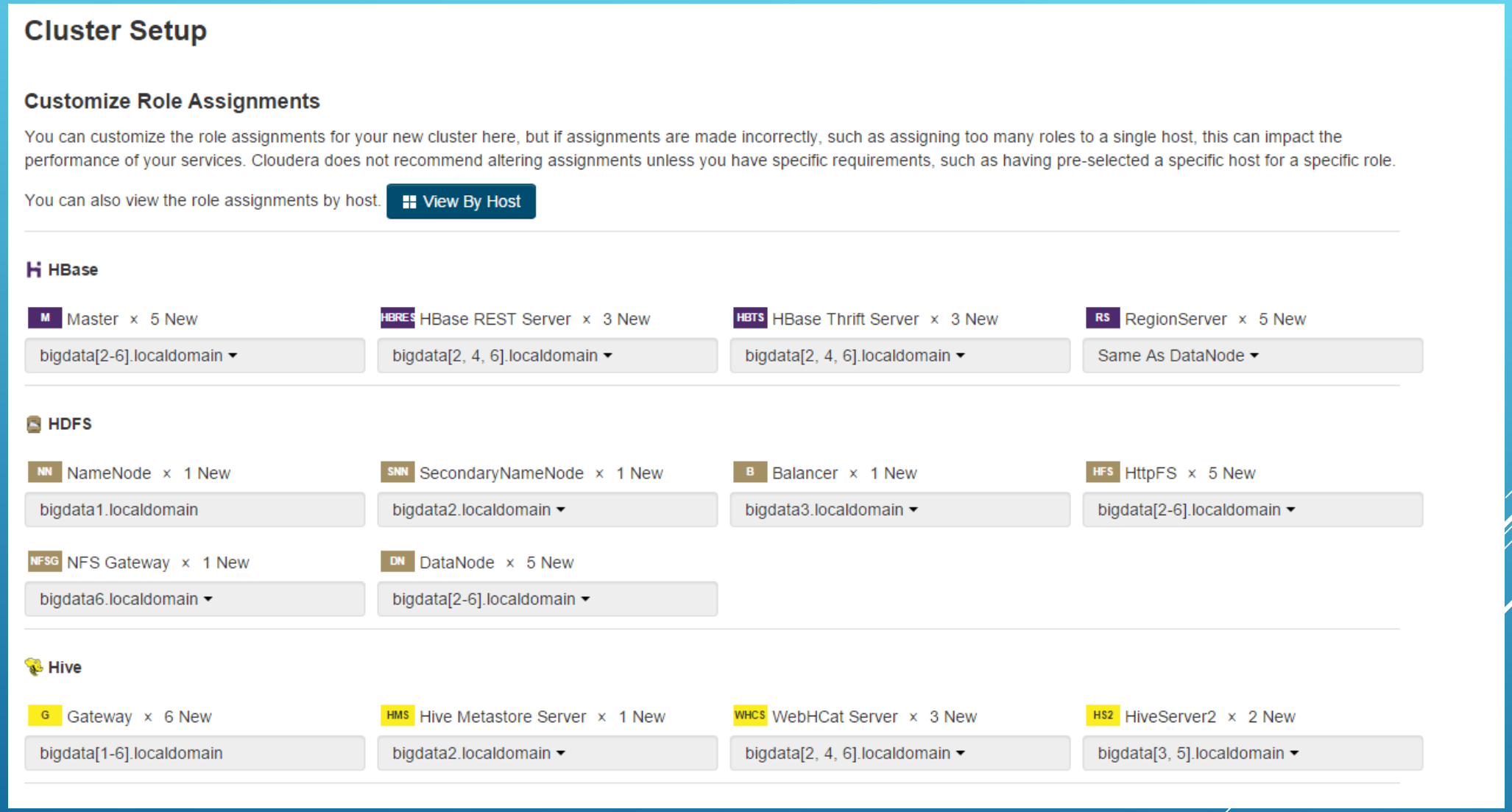
Version Summary

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

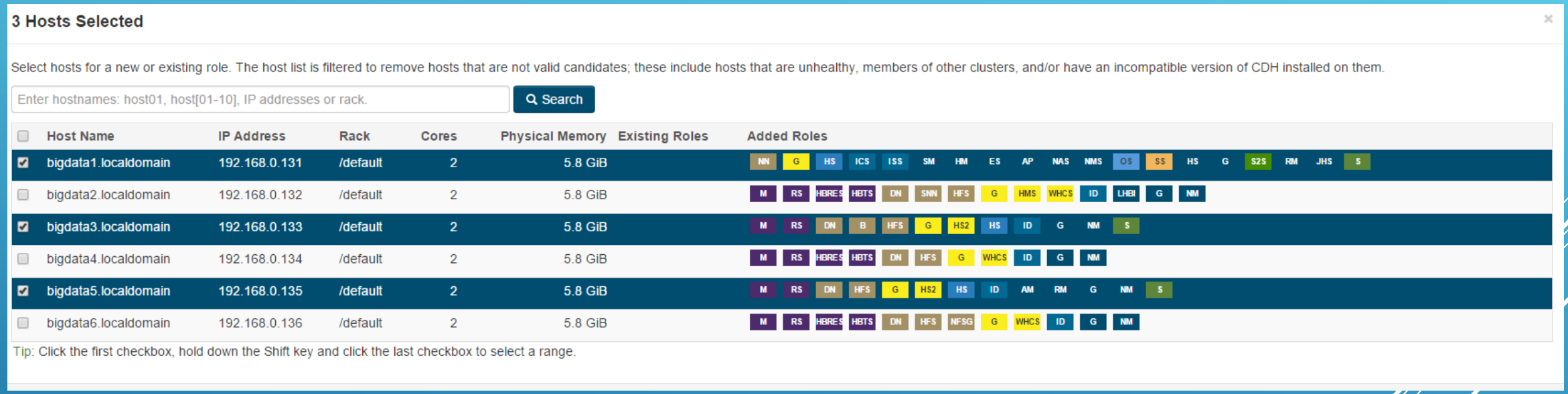
CLOUDERA CLUSTER INSTALLATION – COMPONENT LIST



CLOUDERA CLUSTER INSTALLATION – ROLE CUSTOMIZATION



CLOUDERA CLUSTER INSTALLATION – ROLE CUSTOMIZATION



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:	Database Type:	Database Name :	Username:	Password:
bigdata1.localdomain:7432	PostgreSQL	rman	rman	zqqo3LfWtf

Activity Monitor

✓ **Successful**

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

Database Host Name:	Database Type:	Database Name :	Username:	Password:
bigdata1.localdomain:7432	PostgreSQL	amon	amon	k1ut6qvtwL

Navigator Audit Server

✓ **Successful**

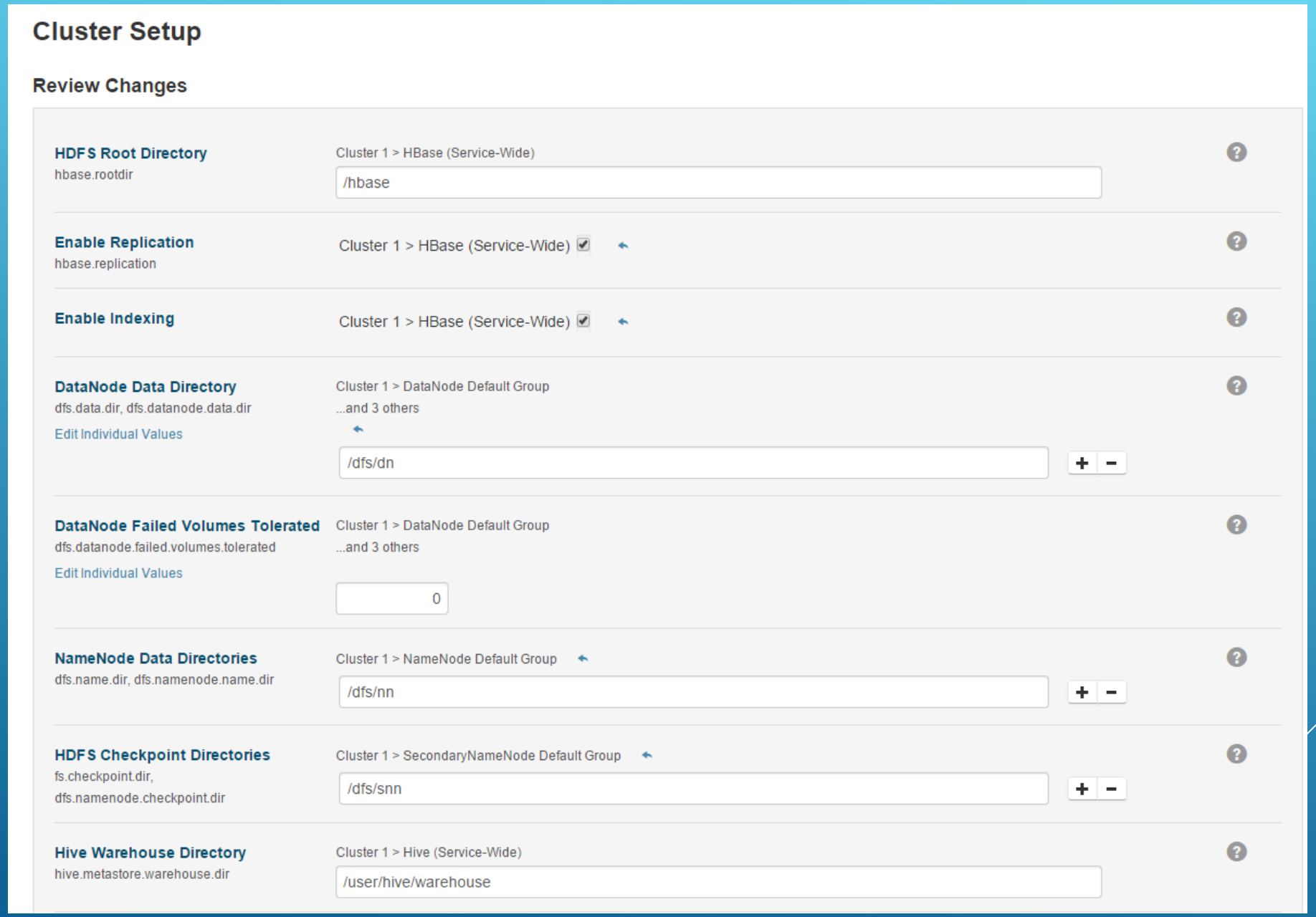
Currently assigned to run on **bigdata1.localdomain**.

Database Host Name:	Database Type:	Database Name :	Username:	Password:
bigdata1.localdomain:7432	PostgreSQL	nav	nav	hUX4nykqKF

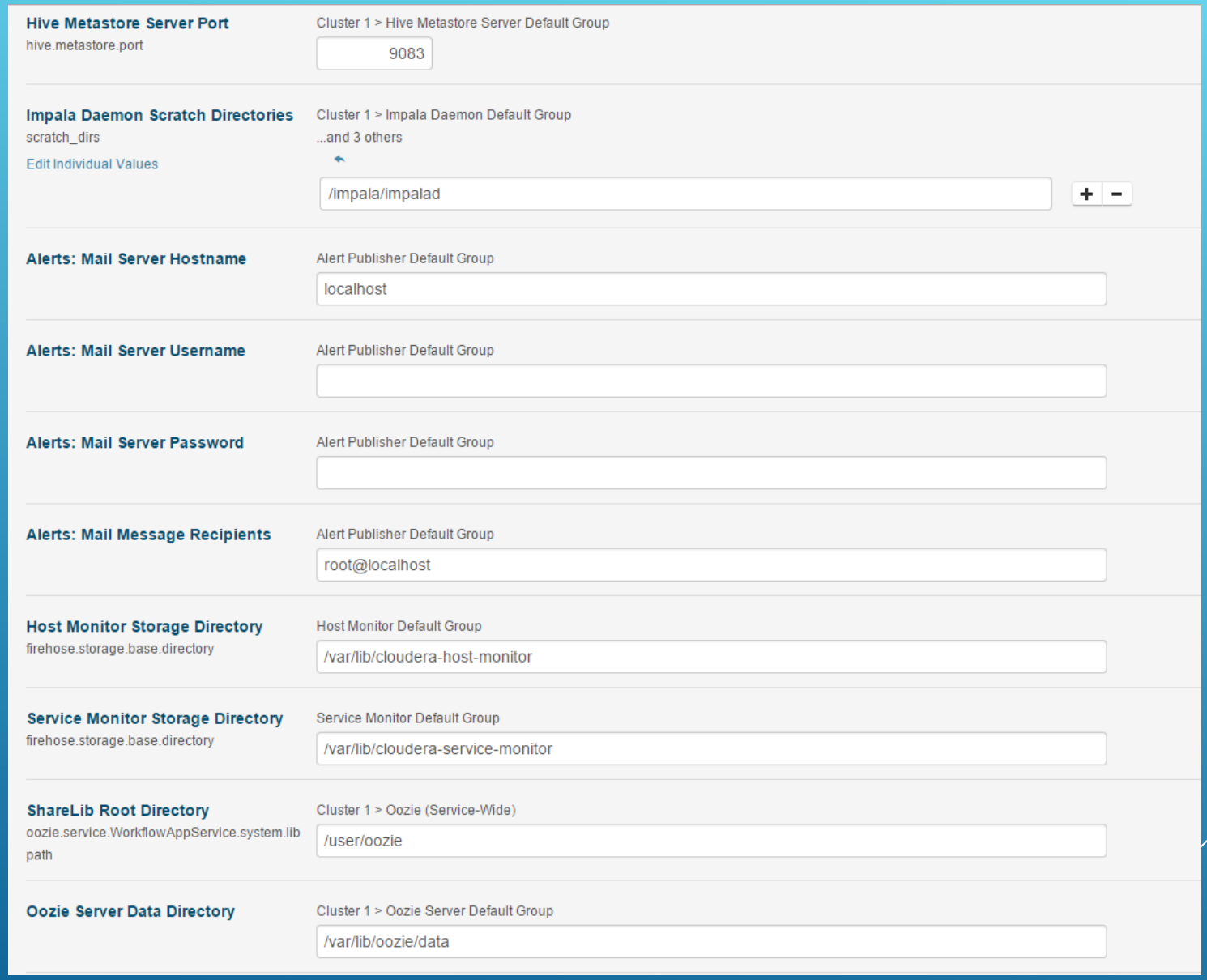
Navigator Metadata Server

✓ **Successful**

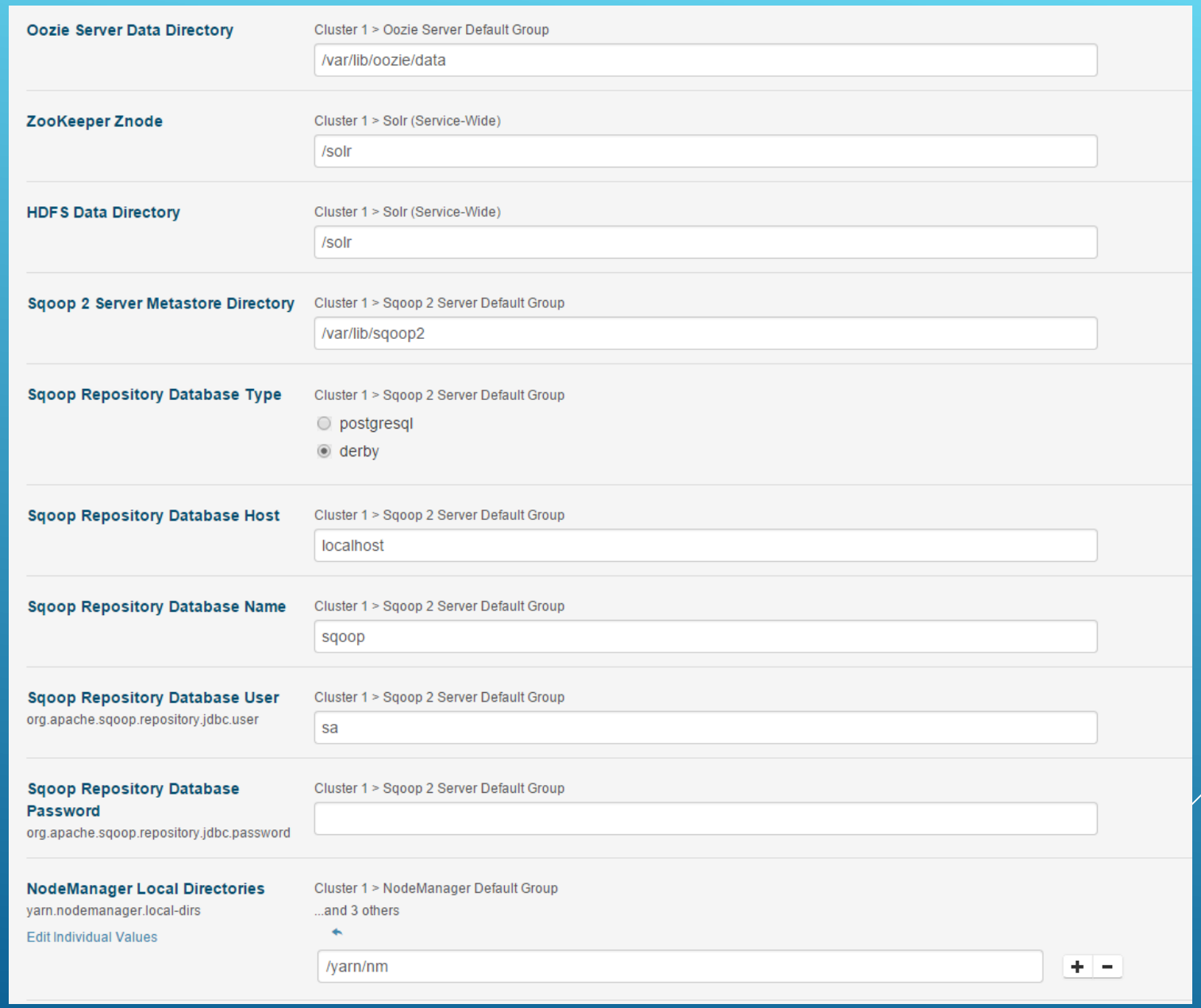
CLOUDERA CLUSTER INSTALLATION – CONFIGURATION CONTD..



CLOUDERA CLUSTER INSTALLATION – CONFIGURATION CONTD..



CLOUDERA CLUSTER INSTALLATION – CONFIGURATION CONTD..



CLOUDERA CLUSTER INSTALLATION – CLIENT CONFIGURATION

Command Details: Deploy Client Configuration

Command	Context	Status
✓ **Deploy Client Configuration**	Cluster 1	Finished

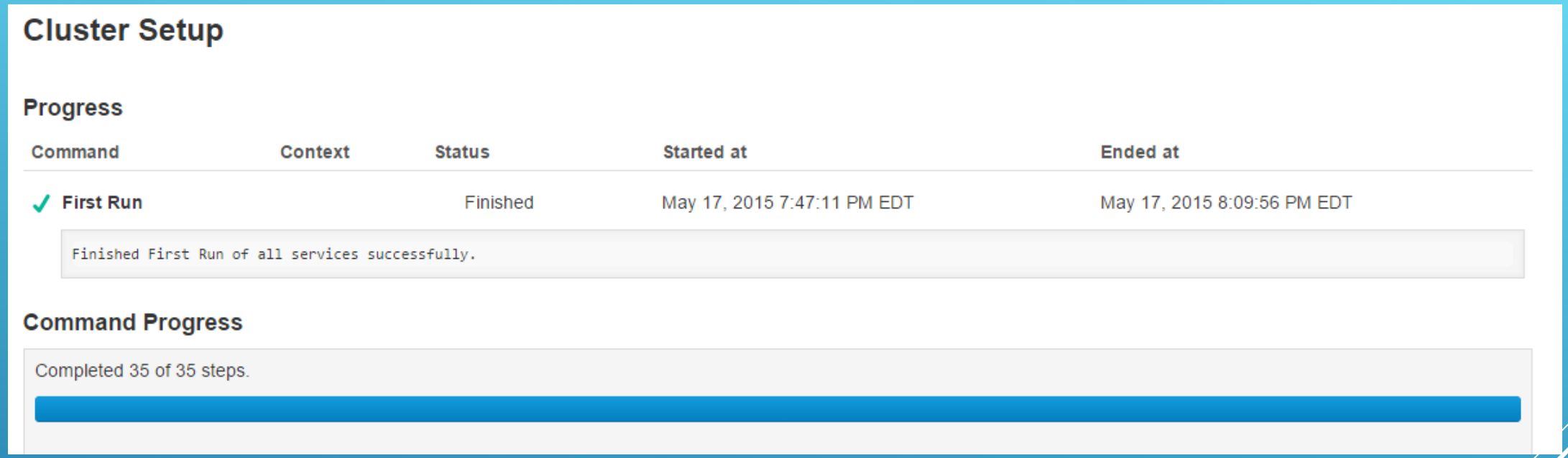
```
Successfully deployed all client configurations.
```

Command Progress

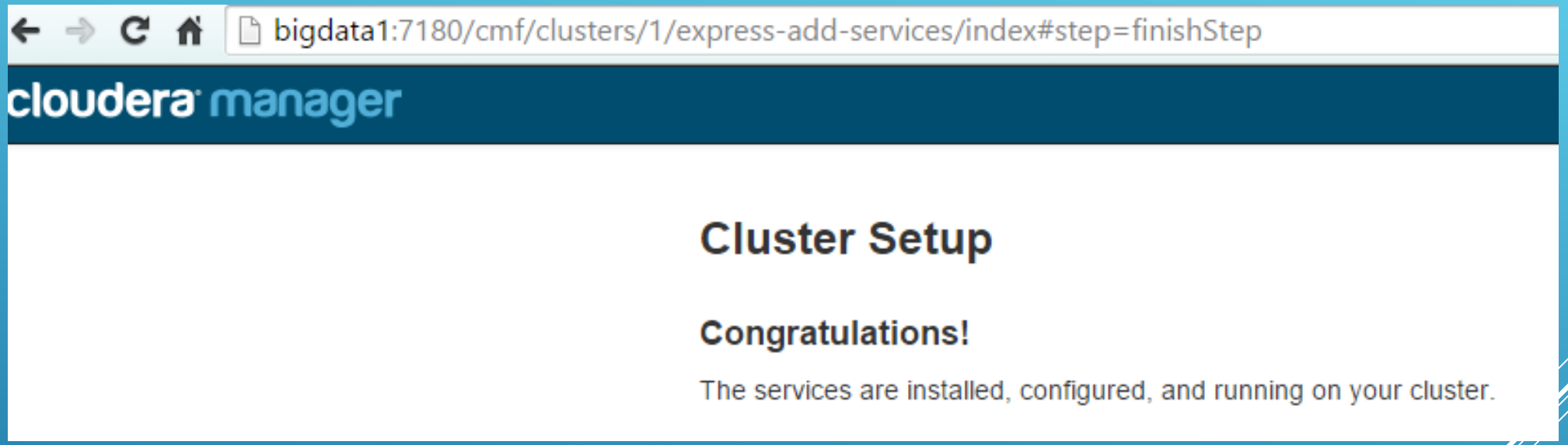
Completed 6 of 6 steps.

- ✓ 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

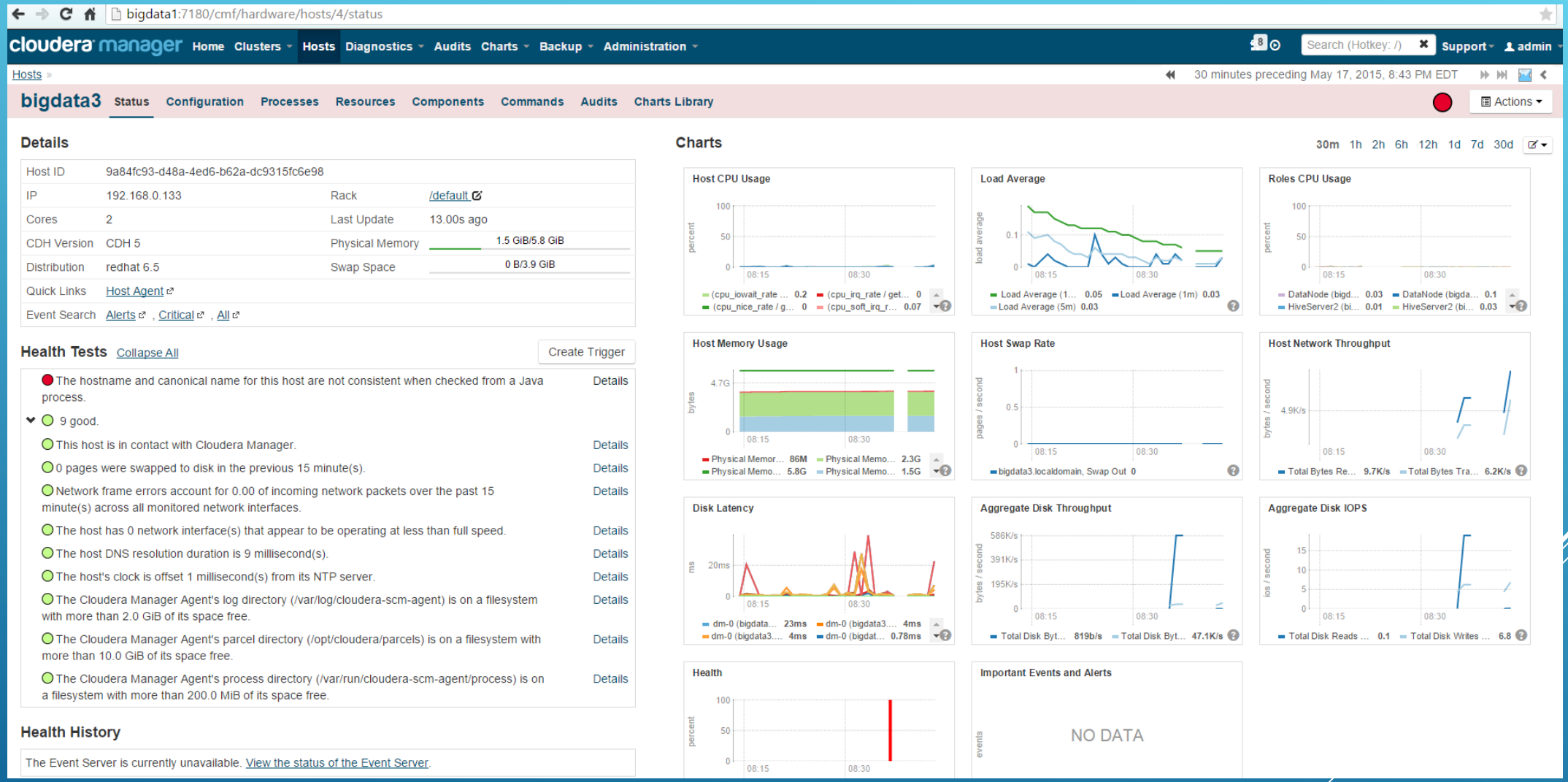
CLOUDERA CLUSTER INSTALLATION – CLIENT CONFIGURATION



CLOUDERA CLUSTER INSTALLATION COMPLETE



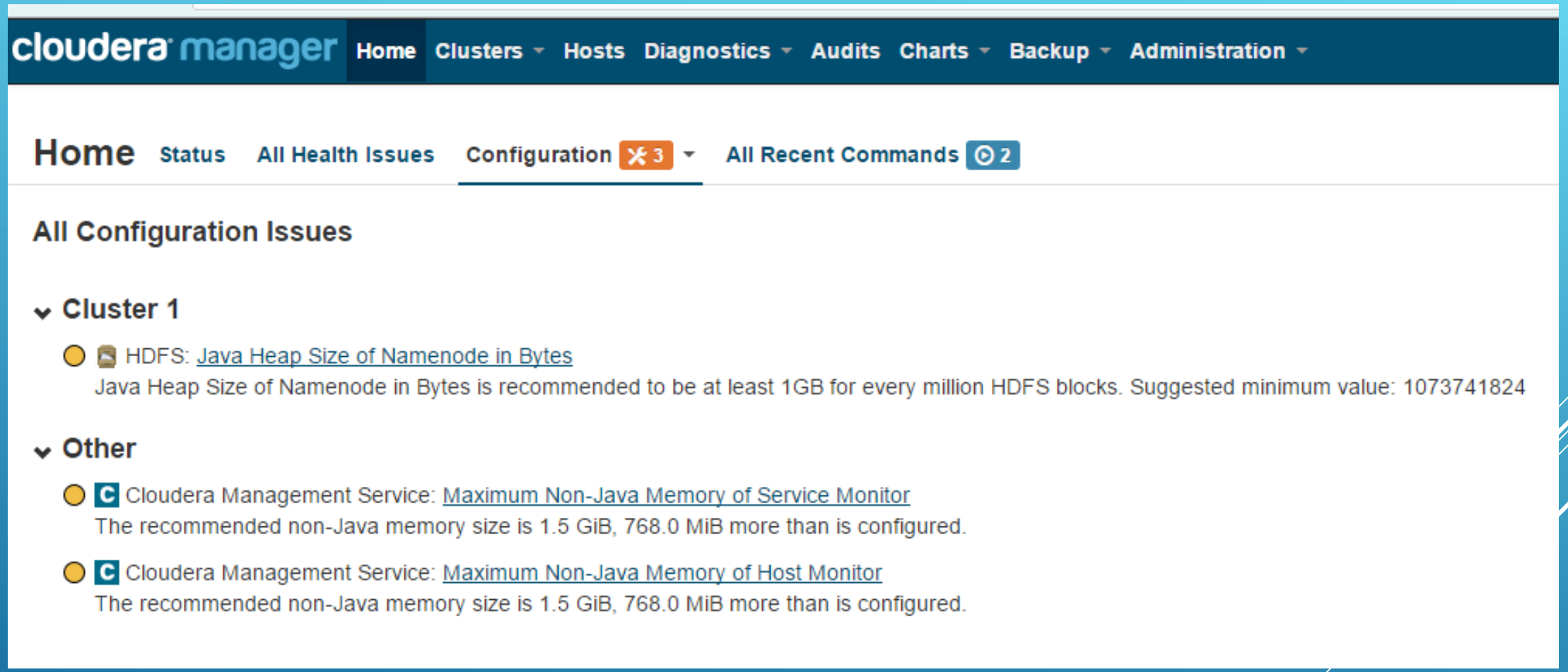
CLOUDERA MANAGER CONSOLE



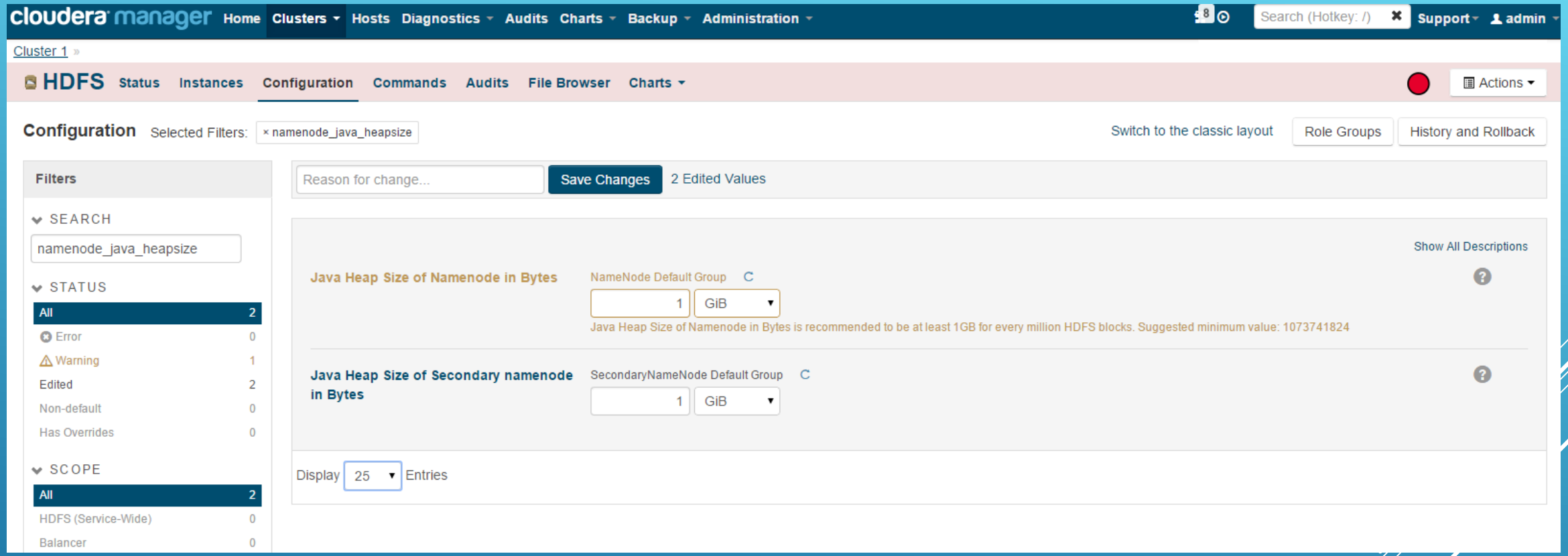
HADOOP COMPONENTS ON CLOUDERA HADOOP CLUSTER

```
[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
```

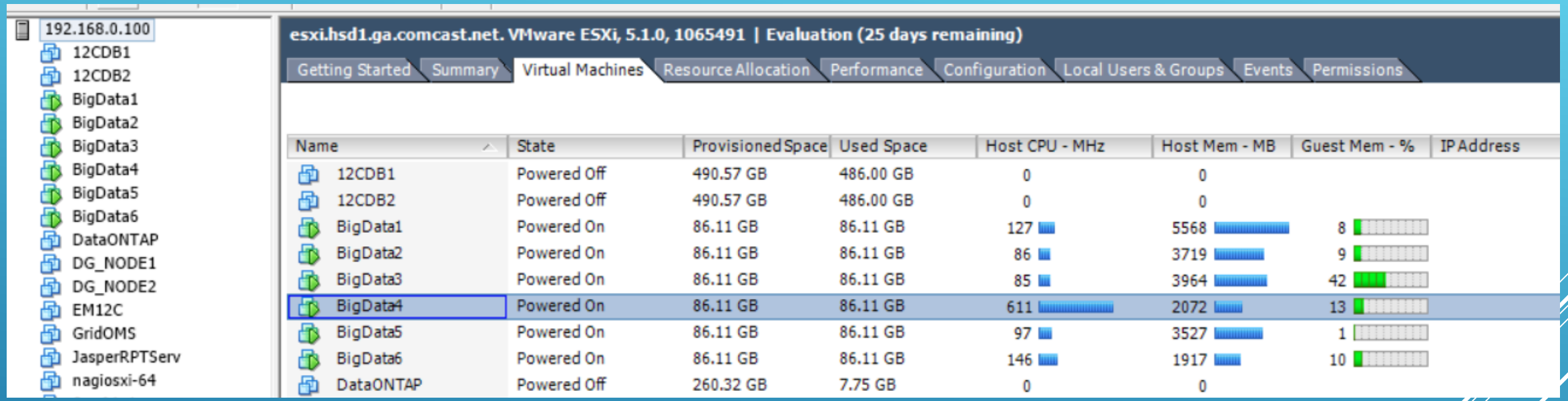
CLOUDERA HADOOP CLUSTER – POST DEPLOYMENT



CLOUDERA HADOOP CLUSTER – POST DEPLOYMENT CONTD



CLOUDERA HADOOP CLUSTER – VSPHERE CONSOLE



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	
BigData4	Powered On	86.11 GB	86.11 GB	611	2072	13	
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		