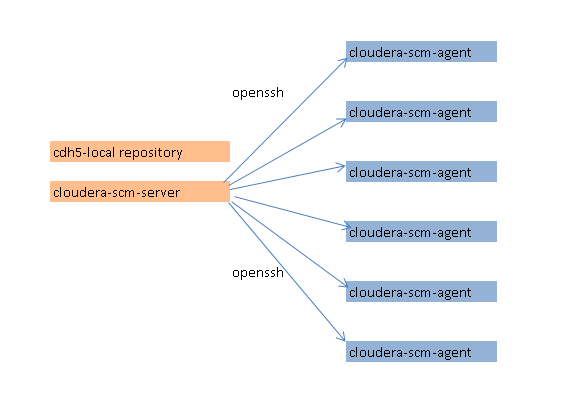
# CDH5 HADOOP INSTALLATION MANUAL

If you follow up [official installation document](http://www.cloudera.com/documentation/cdh/5-1-x/CDH5-Installation-Guide/CDH5-Installation-Guide.html) to install CDH5, you will experience a lot of issue in Citi. This manual is aim to guide you to install CDH5 more smoothly in Citi Cloud environment.

1. First of all, please apply for 4 or more virtual machines. One machine(scm node) will be used to install Cloudera Manager Server and other machine(agent node) will be used install Hadoop components and Cloudera agents.



1. Setup SSH PUBLIC KEY from the scm node to all scm agent node([document](https://catecollaboration.citigroup.net/domains/platstor/osunix/stdsrelateddocs/How%20to%20Generate%20User%20Keys%20and%20Set%20Up%20Public%20Key%20Authentication.pdf))).

On scm node:

/opt/tectia/util/generate\_keys root

This command will generate public key and private key.

cd /etc/ssh2/keys/root



id\_rsa\_2048\_a is a private key

id\_rsa\_2048\_a.pub a public key.

Please make a copy of public key, for example

cp id\_rsa\_2048\_a.pub vm-8630-6e63.pub

Then copy the public key to **all agent nodes** and save it to /etc/ssh2/keys/root

**Attention, Please convert Tetic ssh key to openssh2 key so that CDH can recognize it, the open ssh key will be used in the future.**

ssh-keygen --key-format openssh2 --import-private-key id\_rsa\_2048\_a id\_rsa\_2048\_a.openssh

On agent node:

/opt/tectia/util/generate\_keys root

Execute the command too and then edit **authorization** file to allow scm node to login.

cat /etc/ssh2/keys/root/authorization

# SSH Tectia authorized public key in UserConfigDirectory

Key id\_rsa\_2048\_a.pub

Options command="eval $SSH\_ORIGINAL\_COMMAND"

Key vm-8630-6e63.pub

Options command="eval $SSH\_ORIGINAL\_COMMAND",allow-from="vm-8630-6e63.nam.nsroot.net",allow-from="10.116.49.45"

# OpenSSH authorized public key in UserConfigDirectory

Key authorized\_keys2

Verify ssh setup, run ssh to make sure no password prompt:

# ssh vm-5a1d-9fa8 'hostname;date'

You are authorized to use this System for approved business purposes only.

Use for any other purpose is prohibited. All transactional records, reports,

email, software and other data generated by or residing upon this System,

to the extent permitted by local law, are the property of Citigroup Inc.

or one of its subsidiaries or their affiliates

(individually or collectively ' Citigroup ') and may be used by Citigroup

for any purpose authorized and permissible in your country of work.

Activities on this System are monitored to the extent permitted by local law.

vm-5a1d-9fa8

Thu Mar 17 06:13:24 EDT 2016

[Dev root @ vm-8630-6e63 ~]

1. On scm node

export http\_proxy=http://webproxy.ssmb.com:8080

export https\_proxy=http://webproxy.ssmb.com:8080

curl -o /etc/yum.repos.d/cloudera-manager.repo http://archive.cloudera.com/cm5/redhat/6/x86\_64/cm/cloudera-manager.repo

yum install -y oracle-j2sdk1.7

yum install -y cloudera-manager-daemons cloudera-manager-server

yum install -y cloudera-manager-server-db-2

service cloudera-scm-server-db start

/usr/share/cmf/bin/initialize\_embedded\_db.sh: line 360: cd: /home/cloudera-scm: No such file or directory

The files belonging to this database system will be owned by user "cloudera-scm".

This user must also own the server process.

The database cluster will be initialized with locale en\_US.UTF8.

The default text search configuration will be set to "english".

fixing permissions on existing directory /var/lib/cloudera-scm-server-db/data ... ok

creating subdirectories ... ok

selecting default max\_connections ... 100

selecting default shared\_buffers ... 32MB

creating configuration files ... ok

creating template1 database in /var/lib/cloudera-scm-server-db/data/base/1 ... ok

initializing pg\_authid ... ok

setting password ... ok

initializing dependencies ... ok

creating system views ... ok

loading system objects' descriptions ... ok

creating conversions ... ok

creating dictionaries ... ok

setting privileges on built-in objects ... ok

creating information schema ... ok

vacuuming database template1 ... ok

copying template1 to template0 ... ok

copying template1 to postgres ... ok

Success. You can now start the database server using:

postgres -D /var/lib/cloudera-scm-server-db/data

or

pg\_ctl -D /var/lib/cloudera-scm-server-db/data -l logfile start

Adding configs

Creating SCM configuration file: /etc/cloudera-scm-server/db.properties

waiting for server to start.... done

server started

CREATE ROLE

CREATE DATABASE

Created db properties file /etc/cloudera-scm-server/db.properties

Creating DB amon for role ACTIVITYMONITOR

CREATE ROLE

CREATE DATABASE

Created DB for role ACTIVITYMONITOR

Creating DB rman for role REPORTSMANAGER

CREATE ROLE

CREATE DATABASE

Created DB for role REPORTSMANAGER

Creating DB nav for role NAVIGATOR

CREATE ROLE

CREATE DATABASE

Created DB for role NAVIGATOR

Creating DB navms for role NAVIGATORMETASERVER

CREATE ROLE

CREATE DATABASE

Created DB for role NAVIGATORMETASERVER

Enabled remote connections

waiting for server to shut down.... done

server stopped

DB initialization done.

waiting for server to start.... done

server started

1. Install agent software in all agent nodes

Please run following commands in all agent nodes

export http\_proxy=http://webproxy.ssmb.com:8080

export https\_proxy=http://webproxy.ssmb.com:8080

curl -o /etc/yum.repos.d/cloudera-manager.repo http://archive.cloudera.com/cm5/redhat/6/x86\_64/cm/cloudera-manager.repo

yum install -y oracle-j2sdk1.7

yum install -y wget

yum install -y cloudera-manager-agent cloudera-manager-daemons

1. Install Hadoop components in all agent nodes

Please run following commands in all agent nodes

cat <<EOF >/etc/yum.repos.d/local-cdh5.repo

[cloudera-cdh5]

name=Cloudera's Distribution for Hadoop, Version 5 <local>

baseurl=http://vm-8630-6e63.nam.nsroot.net/cloudera-cdh5

gpgcheck = 0

EOF

yum clean all

yum install -y avro-tools crunch flume-ng hadoop-hdfs-fuse hadoop-hdfs-nfs3 hadoop-httpfs hadoop-kms hbase-solr hive-hbase hive-webhcat hue-beeswax hue-hbase hue-impala hue-pig hue-plugins hue-rdbms hue-search hue-spark hue-sqoop hue-zookeeper impala impala-shell kite llama mahout oozie pig pig-udf-datafu search sentry solr-mapreduce spark-python sqoop sqoop2 whirr

1. Fix file permission in all agent nodes

1: update /etc/cloudera-scm-agent/config.ini

Notes: port **9000** is used by memsql ops, so we need to change port number.

## It should not normally be necessary to modify these.

# Port that the CM agent should listen on.

listening\_port=49000

# Port that supervisord should listen on.

# NB: This only takes effect if supervisord is restarted.

supervisord\_port=49001

service cloudera-scm-agent stop

2: kill supervisor

ps -ef | grep -v grep | grep supervisor | awk '{ print $2 }' | xargs -I{} kill -9 {}

3: change file permissoin

chown -R cloudera-scm:cloudera-scm /var/run/cloudera-scm-agent

chown -R cloudera-scm:cloudera-scm /var/log/cloudera-scm-agent

sed -i '/NOPASSWD:/a%cloudera-scm ALL=(ALL) NOPASSWD: ALL' /etc/sudoers

grep NOPASSWD /etc/sudoers

sed -i 's/^Defaults requiretty/#Defaults requiretty/g' /etc/sudoers

grep 'Defaults requiretty' /etc/sudoers

ls -ltr /var/log | grep -i 1

chown -R cloudera-scm:cloudera-scm /var/log/zookeeper

chown -R cloudera-scm:cloudera-scm /var/lib/zookeeper

rm -f /var/lib/zookeeper/\*

chown -R cloudera-scm:cloudera-scm /var/log/hadoop-hdfs

ls -ld /var/log/hadoop-hdfs

#fix namenode

mkdir -p /dfs

chown -R cloudera-scm:cloudera-scm /dfs

chmod -R 755 /dfs

#data node

mkdir -p /var/run/hdfs-sockets

chown -R cloudera-scm:cloudera-scm /var/run/hdfs-sockets

chmod -R 755 /var/run/hdfs-sockets

#yarn

mkdir -p /var/log/hadoop-yarn

chown -R cloudera-scm:cloudera-scm /var/log/hadoop-yarn

chmod -R 755 /var/log/hadoop-yarn

mkdir -p /var/lib/hadoop-yarn

chown -R cloudera-scm:cloudera-scm /var/lib/hadoop-yarn

chmod -R 755 /var/lib/hadoop-yarn

# hue

mkdir -p /var/log/hue /var/lib/hue/

chown -R cloudera-scm:cloudera-scm /var/log/hue /var/lib/hue/

chmod -R 755 /var/log/hue /var/lib/hue/

# oozie

mkdir -p /var/log/oozie /var/lib/oozie/ /etc/oozie

chown -R cloudera-scm:cloudera-scm /var/log/oozie /var/lib/oozie/ /etc/oozie

chmod -R 755 /var/log/oozie /var/lib/oozie/ /etc/oozie

#

mkdir -p /var/log/cloudera-scm-firehose /var/lib/cloudera-scm-firehose /var/log/cloudera-scm-eventserver /var/lib/cloudera-scm-eventserver /var/lib/cloudera-host-monitor /var/log/cloudera-host-monitor /var/lib/cloudera-service-monitor /var/log/cloudera-service-monitor

chown -R cloudera-scm:cloudera-scm /var/log/cloudera-scm-firehose /var/lib/cloudera-scm-firehose /var/log/cloudera-scm-eventserver /var/lib/cloudera-scm-eventserver /var/lib/cloudera-host-monitor /var/log/cloudera-host-monitor /var/lib/cloudera-service-monitor /var/log/cloudera-service-monitor

chmod -R 755 /var/log/cloudera-scm-firehose /var/lib/cloudera-scm-firehose /var/log/cloudera-scm-eventserver /var/lib/cloudera-scm-eventserver /var/lib/cloudera-host-monitor /var/log/cloudera-host-monitor /var/lib/cloudera-service-monitor /var/log/cloudera-service-monitor

1. Start scm and agent

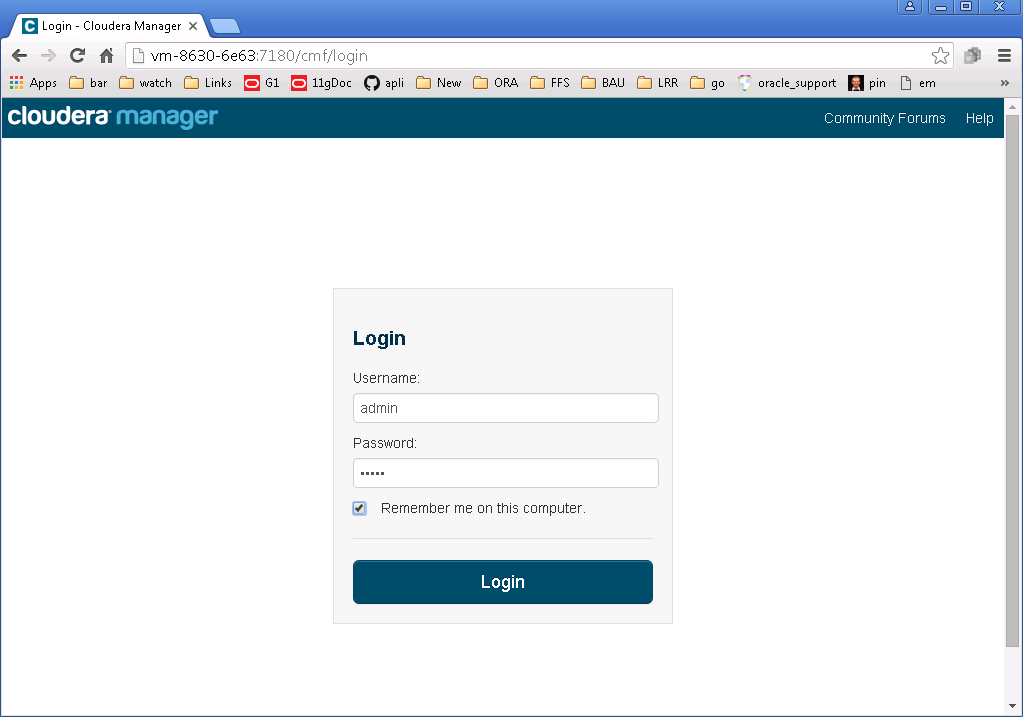
On scm node

service cloudera-scm-server start

On all agent node

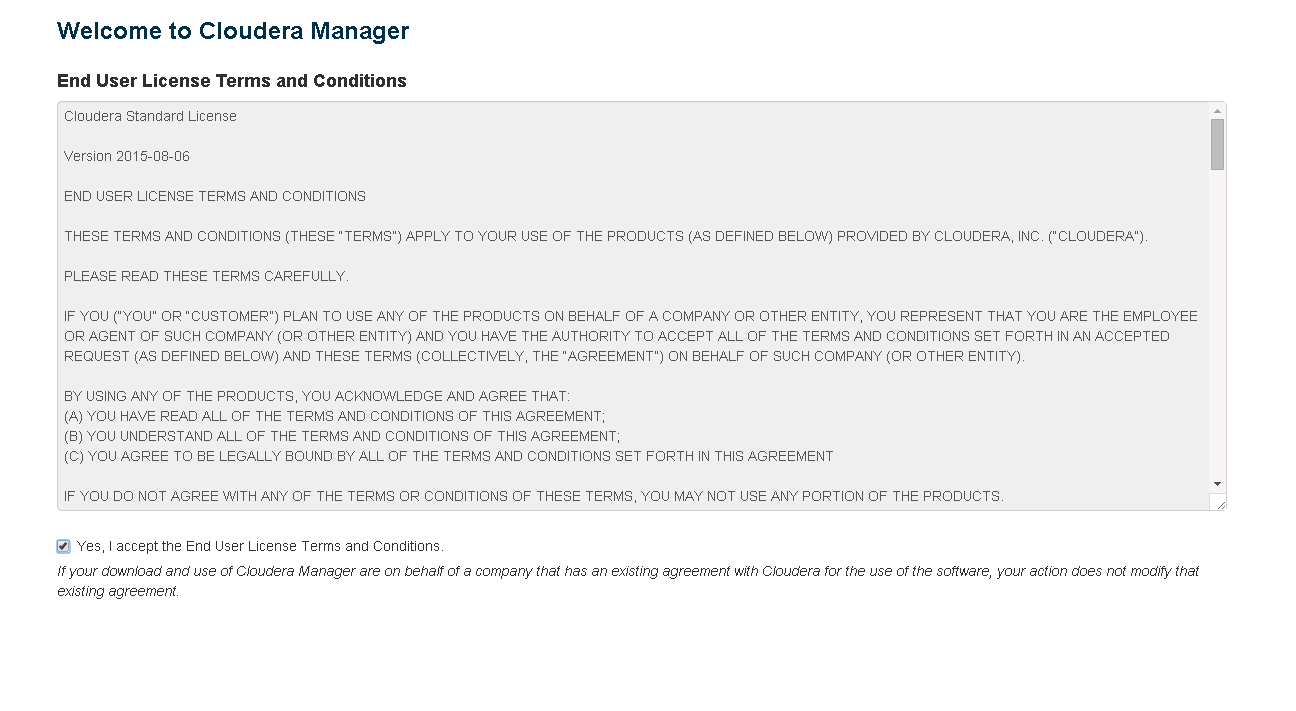
service cloudera-scm-agent start

1. Login admin portal and create cluster

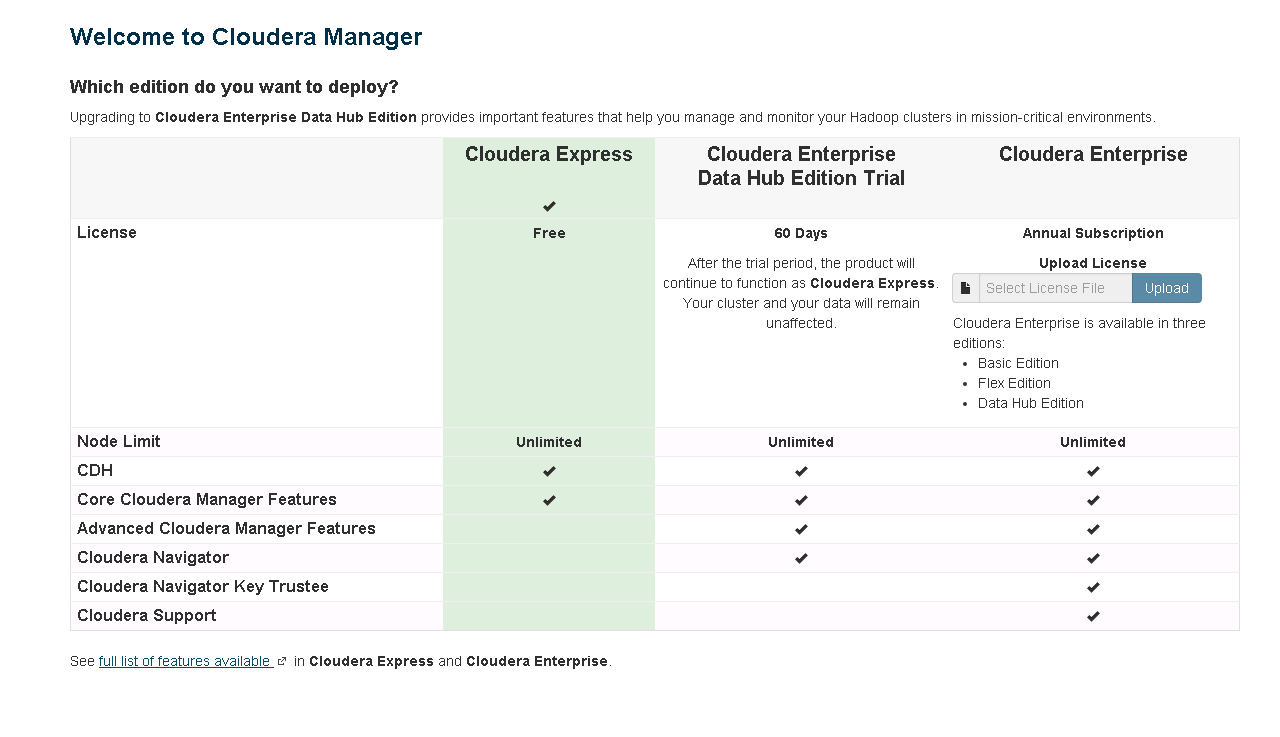


The default port of admin website is **7180**, the default username is **admin** and password is **admin**

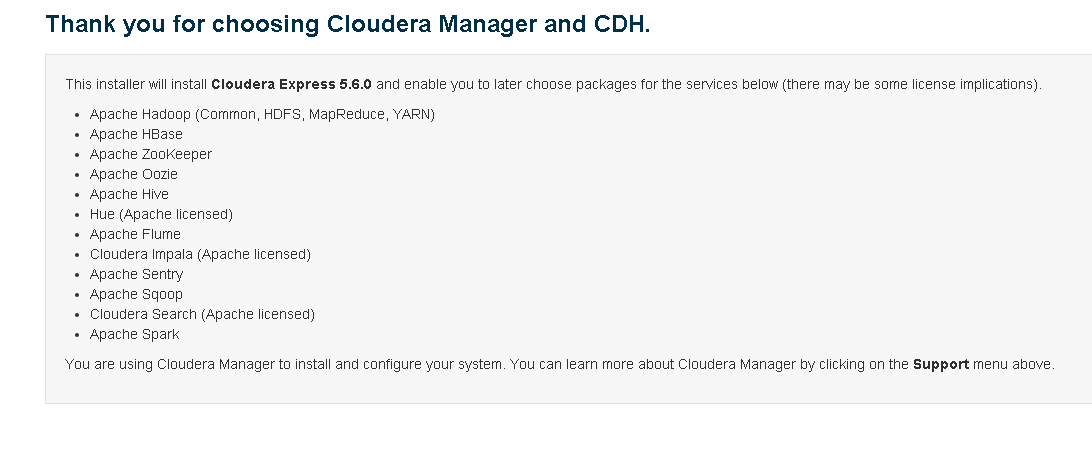
Click Login button to login



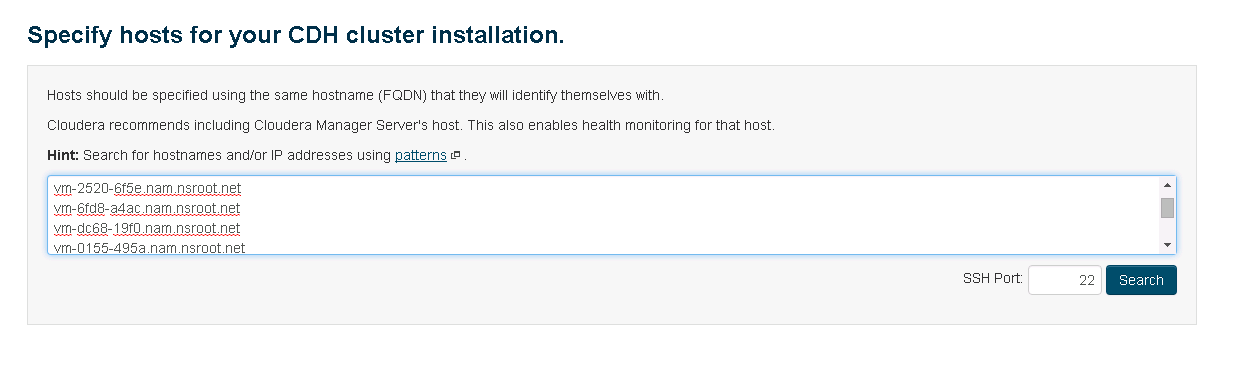
Please accept the user license terms.



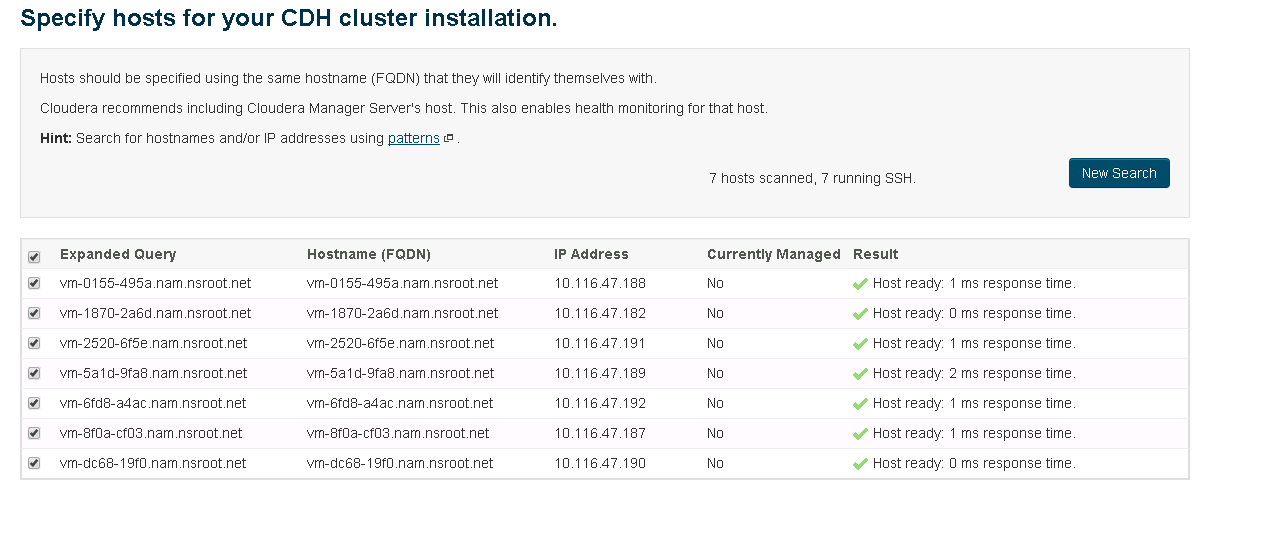
Please choose Cloudera express or upload license file.



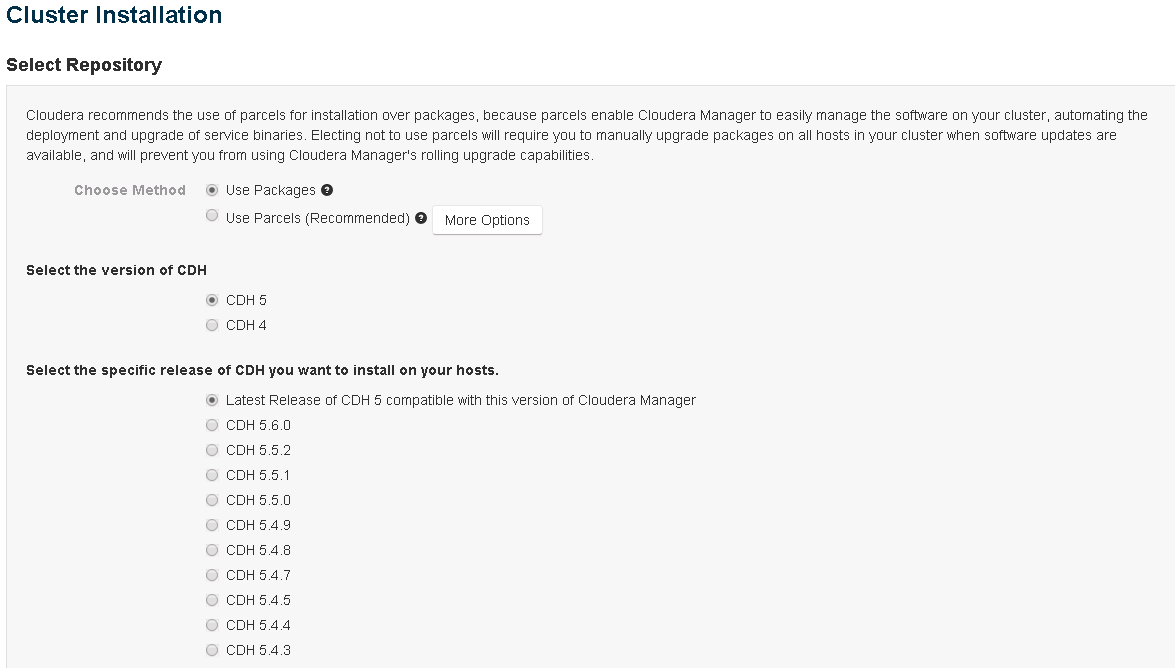
Please click continue button to continue.



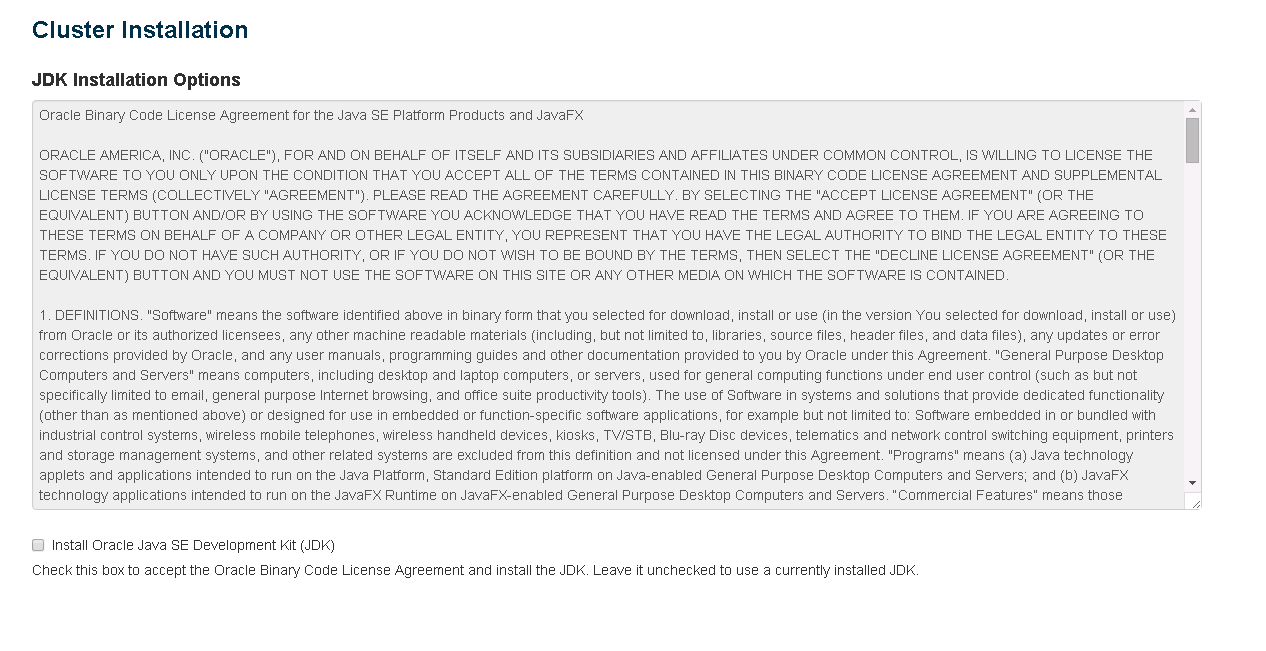
Please add hosts name of all agent nodes and click search.



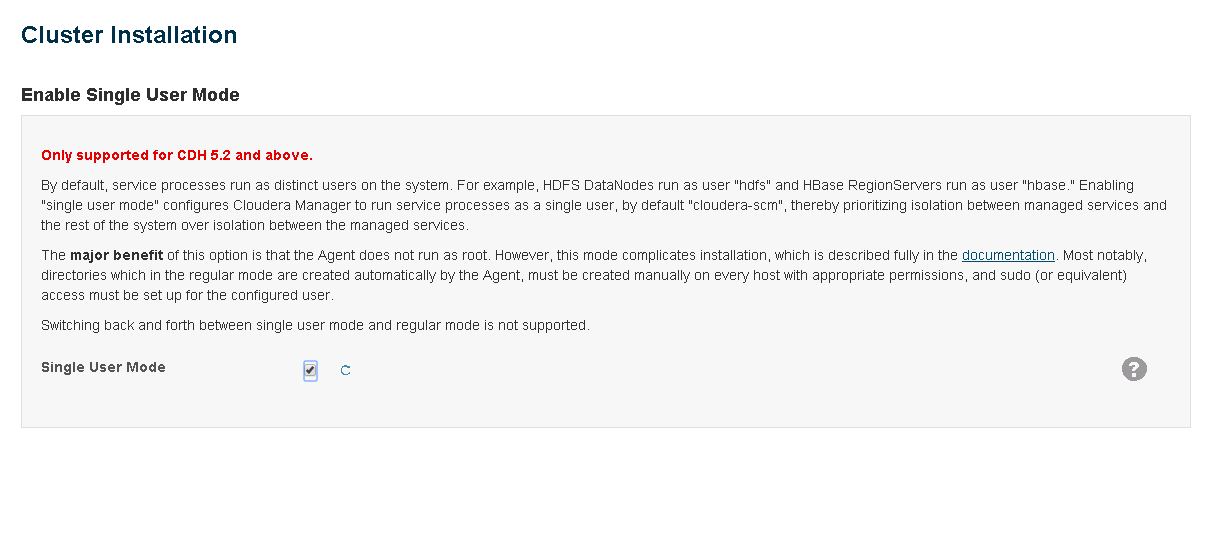
Please click continue button to continue.



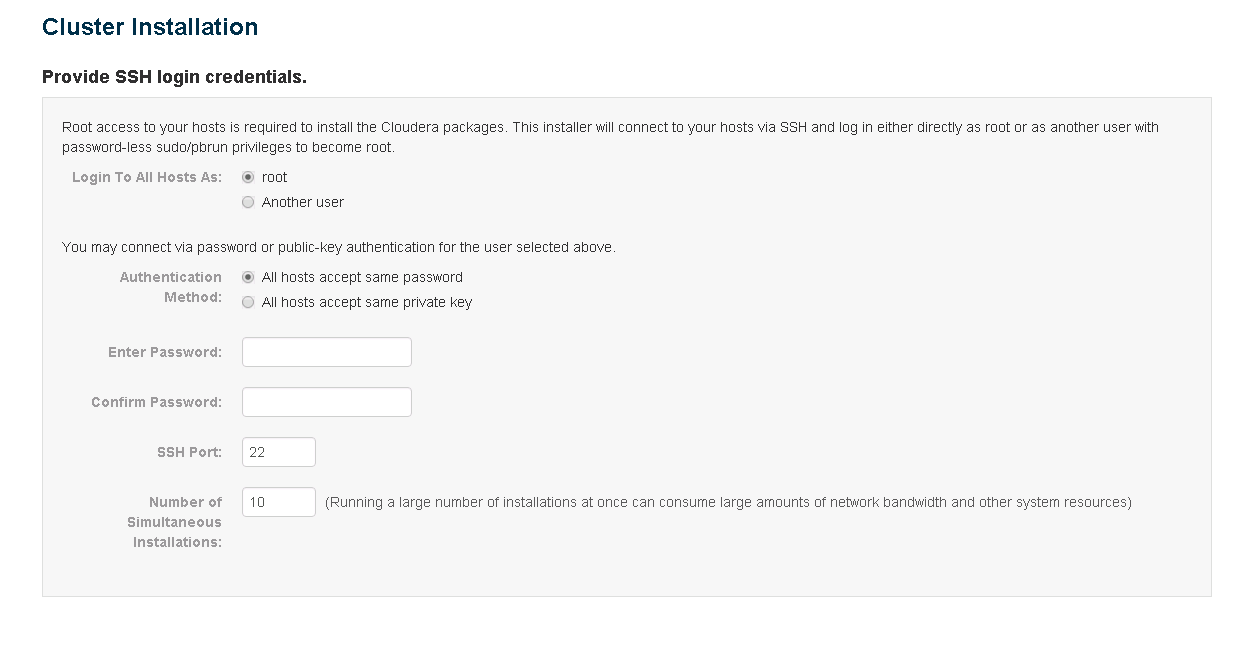
Please choose: “Packages + CDH5 + CDH5.6 “ and click continue button to continue.



Please check: “JDK installation “and click continue button to continue

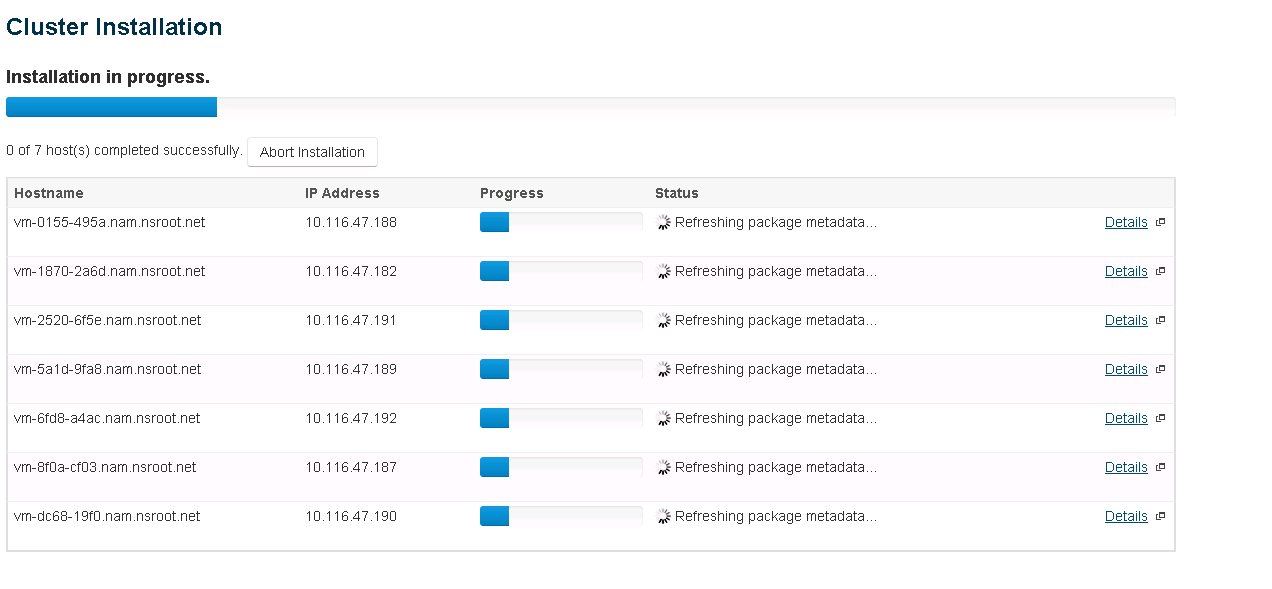


Please check “Single User Mode “and click continue button to continue

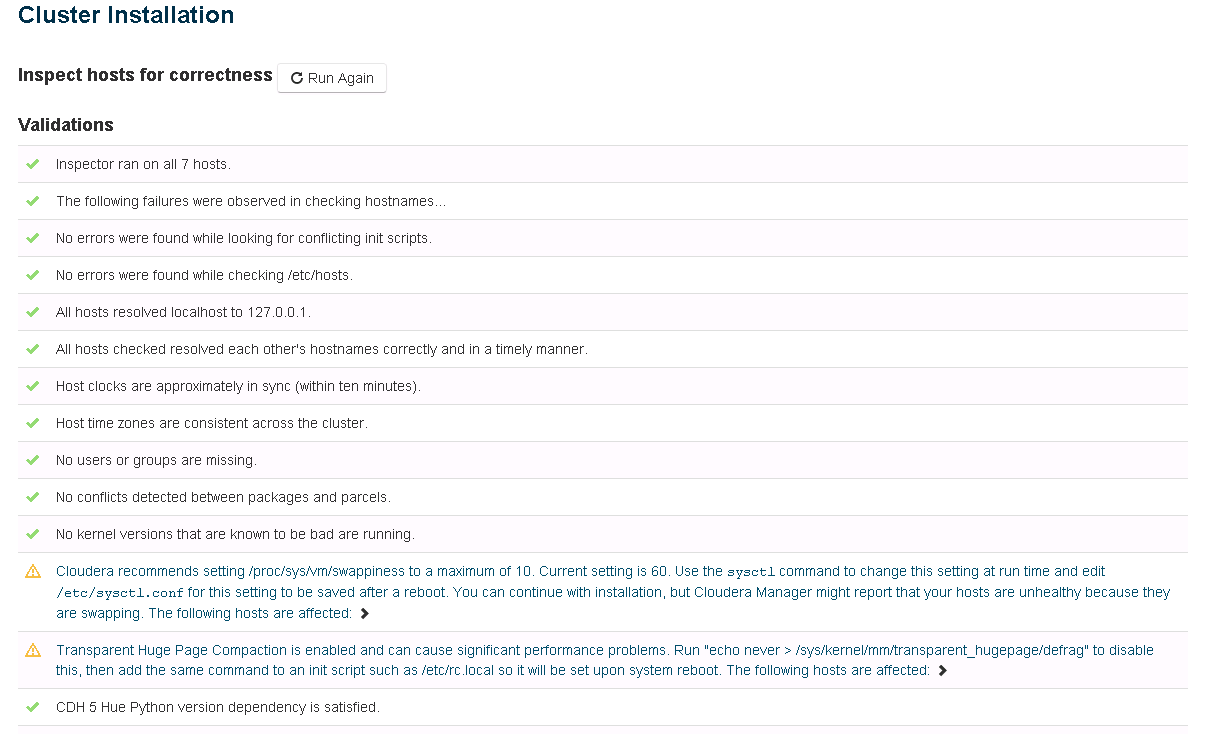


Please choose:

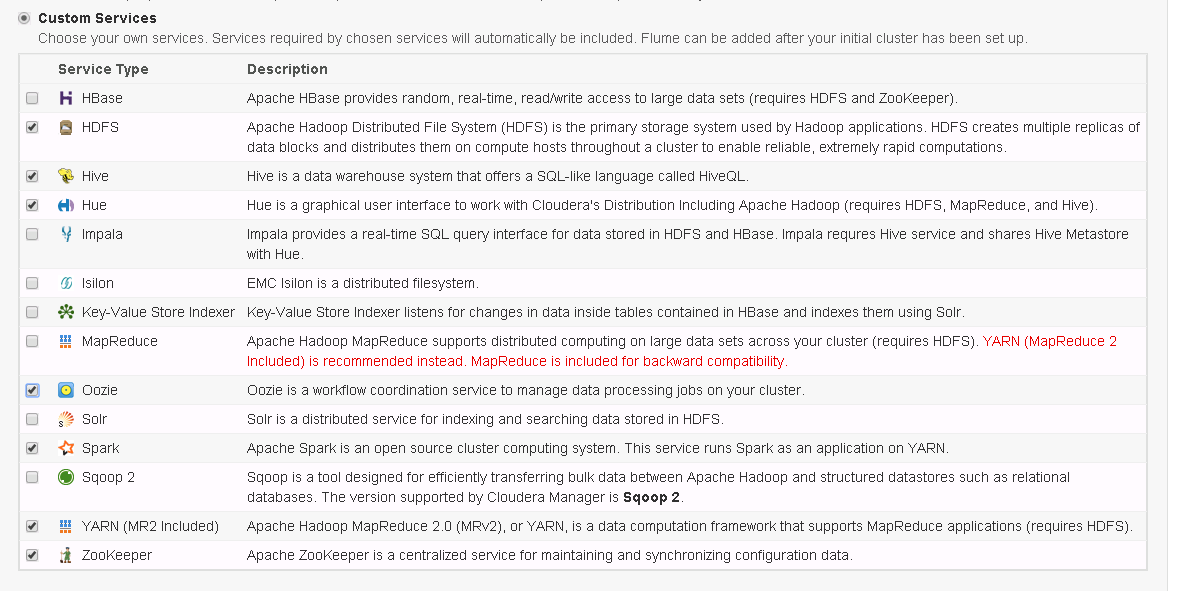
* Root
* All hosts accept same private key
* Use generated openssh private key.



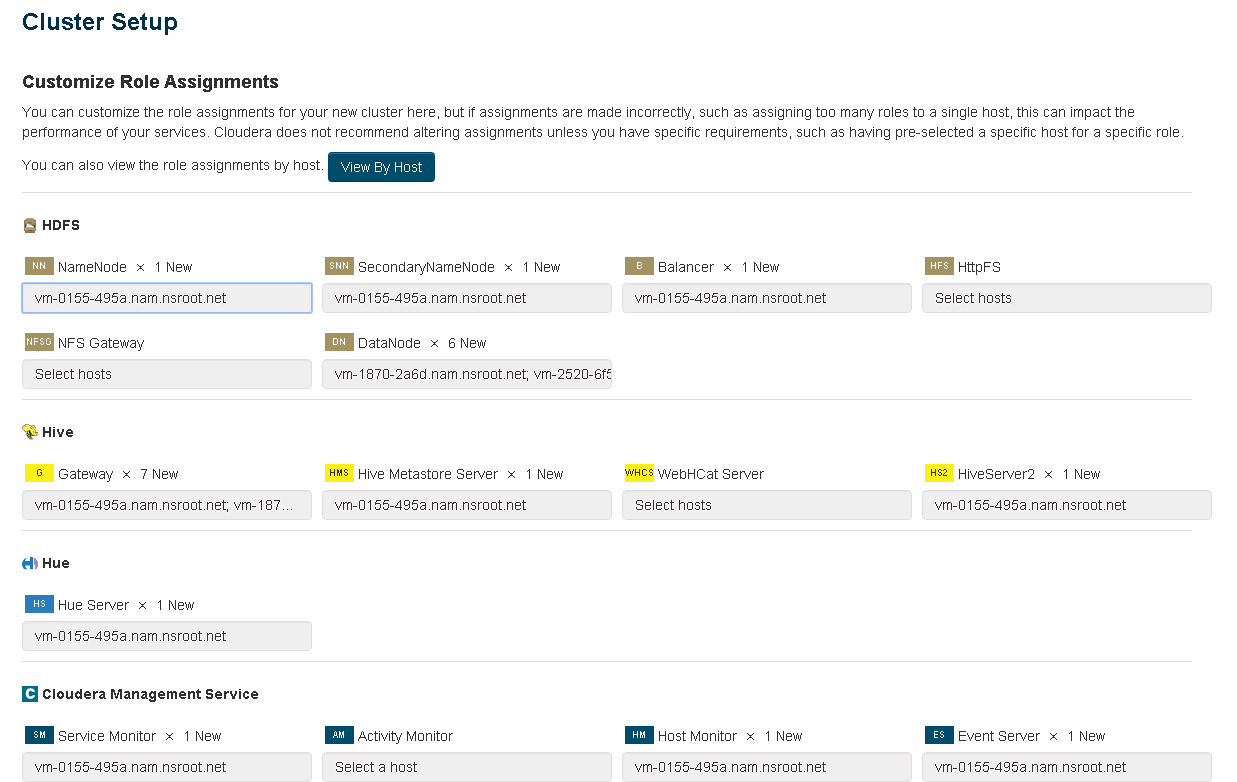
Wait for all installation completed.



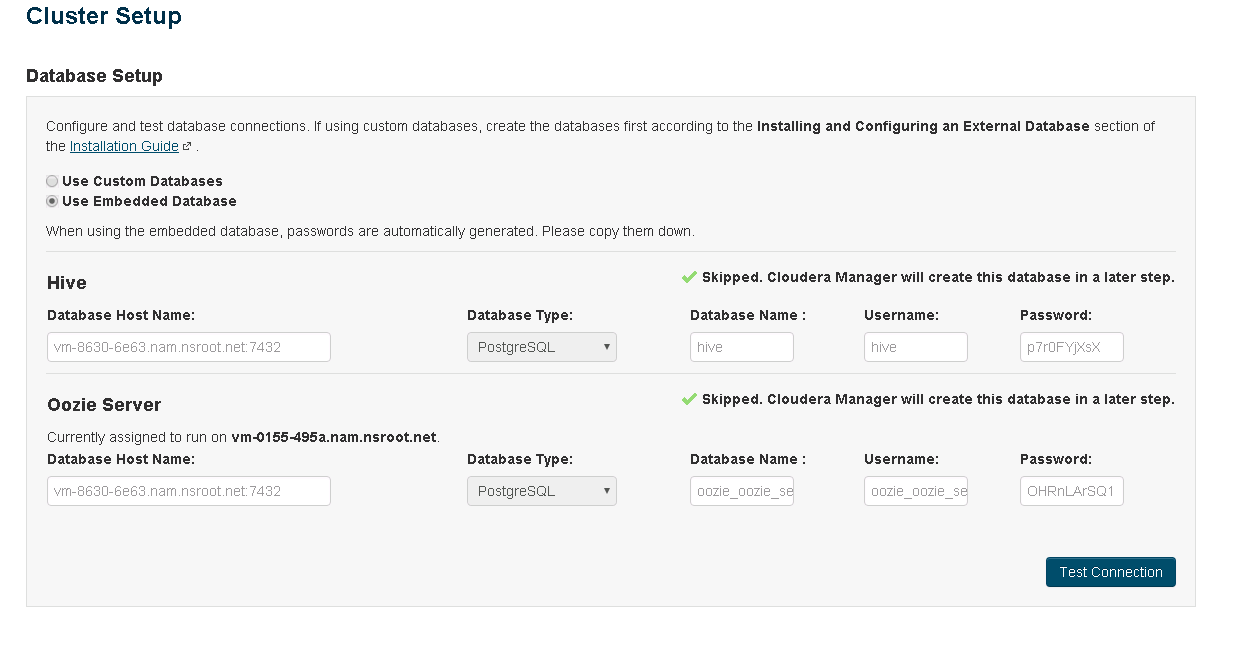
Click continue button to continue



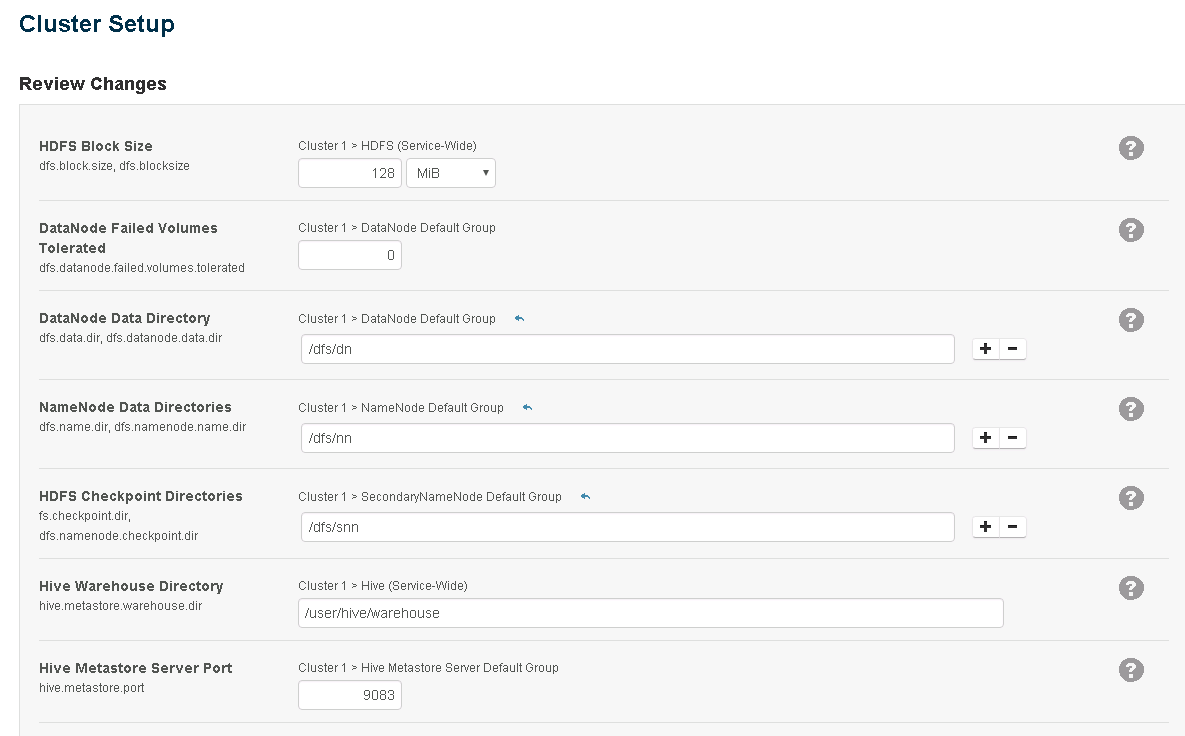
Choose all HDP components you want to use and click continue button to continue



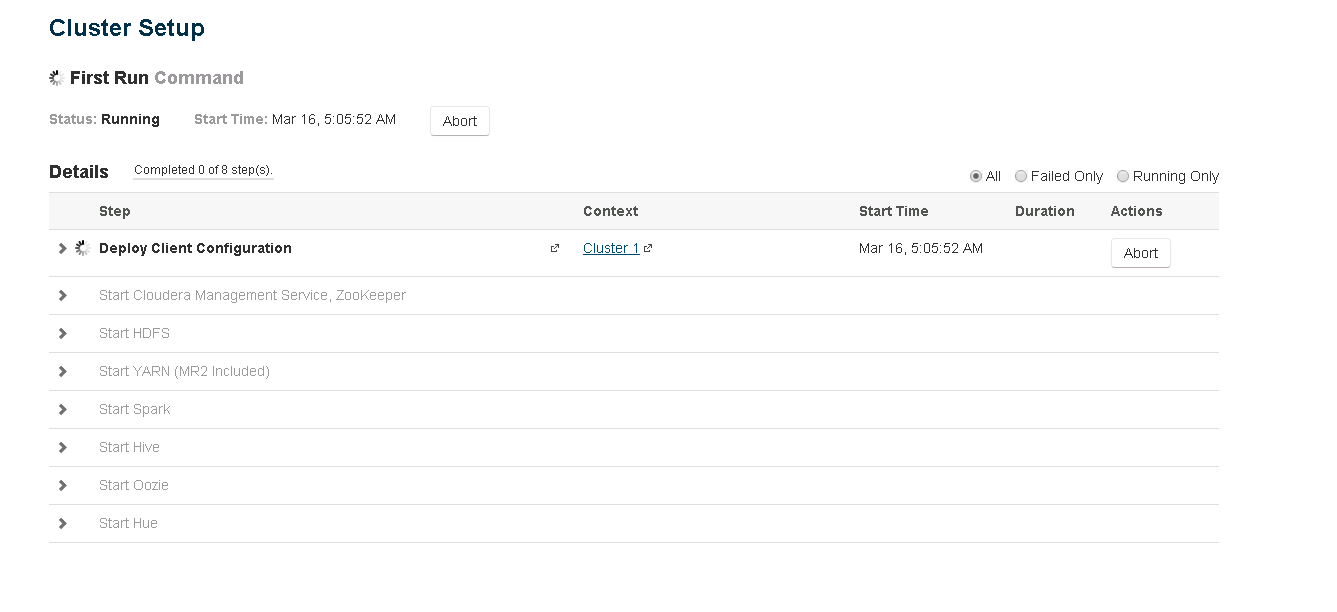
Allocation node to HDP components( or use default value )and then click continue button to continue



Use embed database or custom database and then click continue button to continue



Update default value and then click continue button to continue



Start HDF components.

