Document 1: Cloudera Manager Installation

Objective: Install and configure Cloudera Manager Free edition

- Learn about CDH and Cloudera Manager
- Install Cloudera Manager Free Edition
- Connect to Cloudera Manager Admin Console (website)
- Install CDH4.1.2
- Run some post-install verification commands

CDH4.1.2 & Cloudera Manager Information:

CDH consists of 100% open source Apache Hadoop plus nine other open source projects from the Hadoop ecosystem. CDH is thoroughly tested and certified to integrate with the widest range of operating systems, hardware and databases.

Components' Versions included in CDH 4.1.2:
https://ccp.cloudera.com/display/DOC/CDH+Version+and+Packaging+Information#CDHVersionandPackagingInformation-CDHVersion4.1.2Packaging

There are two editions of CM: Free and Enterprise. CM Free can be used for clusters under 50 nodes and helps with deployment and service/configuration management. CM Enterprise can be used for unlimited nodes and includes more features like CPU + RAM monitoring, service monitoring, security, multi- cluster management, log management, events + alerting and tech support integration.

For Detailed comparison between both the versions:
http://www.cloudera.com/content/cloudera/en/products/cloudera-manager.html

CDH4 contains both MapReduce gen-one (specifically version 0.20.2) and MapReduce next- gen (YARN). Cloudera strongly advices to only use MR-v1 in production at this time. MR-v2 is considered alpha code at the moment and is not designed to be run in production. It is only being early-released with CDH4 for early adopters to start exploring and learning. YARN should be production ready by approximately mid to late 2013. It is not possible to run both MR-v1 and MR-v2 in the cluster simultaneously.

Reference Documentation

"Ports Used by Components of CDH4"
https://ccp.cloudera.com/display/FREE4DOC/Configuring+Ports+for+Cloudera+Manager+Free+Edition

Installing Putty (Windows):

If you're on Windows, install PuTTY, a free telnet/SSH client. With PuTTY, you can connect to the Cloudera VM from a lightweight client in Windows and open multiple cmd- line sessions to the VM.

Download PuTTY from:
http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html

Look for the file named putty.exe under "Windows on Intel x86":

Binaries

The latest release version (beta 0.62). This will generally be a version I think is reasonably likely to work we bug, before reporting it to me.

For Windows on Intel x86

PuTTY:	putty.exe	(or by FTP)	(RSA sig)	(DSA sig)
PuTTYtel:	puttytel.exe	(or by FTP)	(RSA sig)	(DSA sig)
PSCP:	pscp.exe	(or by FTP)	(RSA sig)	(DSA sig)
PSFTP:	psftp.exe	(or by FTP)	(RSA sig)	(DSA sig)
Plink:	plink.exe	(or by FTP)	(RSA sig)	(DSA sig)
Pageant:	pageant.exe	(or by FTP)	(RSA sig)	(DSA sig)
PuTTYgen:	puttygen.exe	(or by FTP)	(RSA sig)	(DSA sig)

A .ZIP file containing all the binaries (except PuTTYtel), and also the help files

Zip file:	putty.zip	(or by FTP)	(RSA sig)	(DSA sig)

A Windows installer for everything except PuTTYtel

Installer:	putty-0.62-installer.exe	(or by FTP)	(RSA sig)	(DSA sig)

After starting PuTTY, enter the IP address of the VM where Cloudera is being hosted into PuTTY. The connection type will be SSH and the port will be 22.

In our case, the VM is being hosted on a local machine at 172.16.90.25, which cannot be accessed from outside the campus as NATing and Port Forwarding are disabled by institute's rules. So we would be using Teamviewer as the alternative. Nevertheless this guide is prepared keeping ideal case in mind.

The username and password for login are:

Login as: **project**
cloudera@172.16.90.25 password: **project123**

Connecting to the VM via Terminal/iTerm2 (Mac):

Open up your terminal app of choice and type in:
ssh root@172.16.90.25

Type in the password when prompted and accept the rsa2 key if needed.

Install Cloudera Manager Free Edition:

First download CM Free from cloudera:

```
[root@cdh4-cm-vm1 ~]# pwd
/root
[root@cdh4-cm-vm1 ~]# wget
http://archive.cloudera.com/cm4/installer/latest/cloudera-manager-
installer.bin
--2012-08-27 04:34:16--
http://archive.cloudera.com/cm4/installer/latest/cloudera-manager-
installer.bin
Resolving archive.cloudera.com... 75.101.147.51
Connecting to archive.cloudera.com|75.101.147.51|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 504807 (493K) [application/octet-stream]
Saving to: âcloudera-manager-installer.binâ

100%[======================================>] 504,807     1.32M/s   in
0.4s

2012-08-27 04:34:17 (1.32 MB/s) - âcloudera-manager-installer.binâ
```

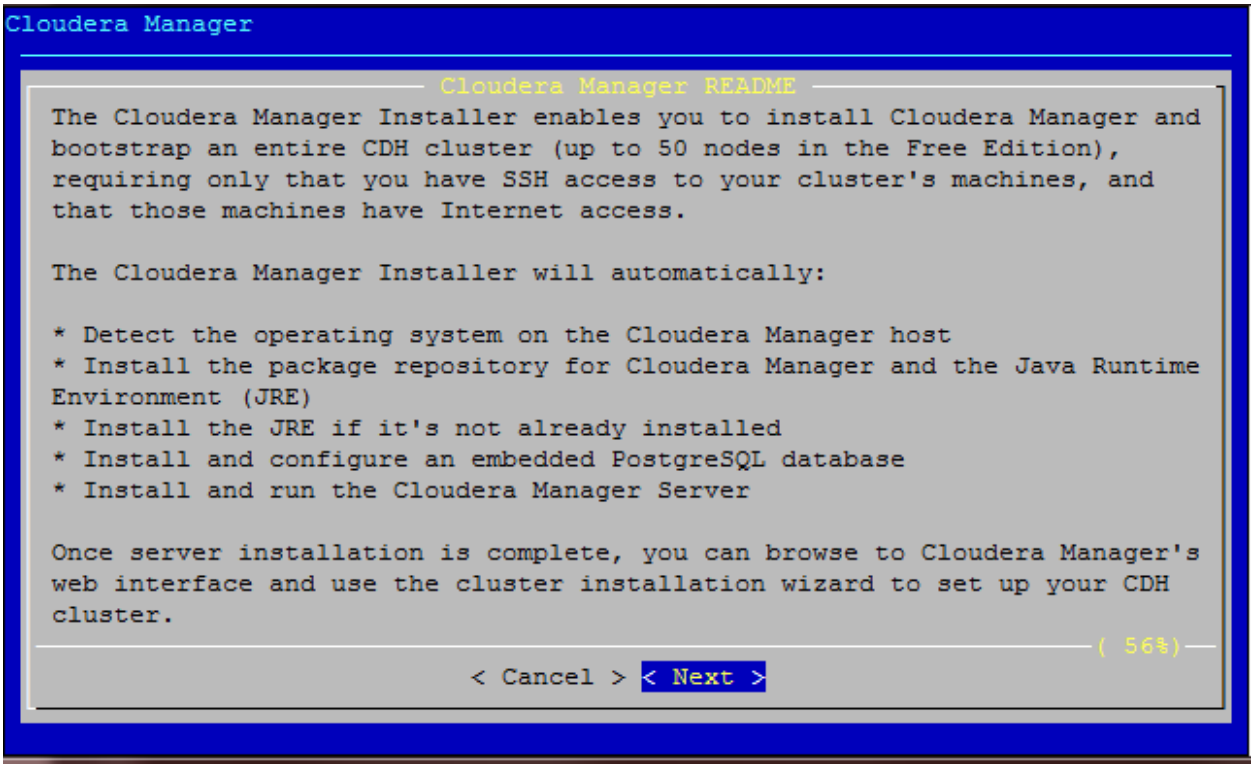
Change the installer file to have executable permissions:

```
[root@cdh4-cm-vm1 ~]# ls -l
total 500
-rw-r--r-- 1 root root 504991 Jan 18 07:05 cloudera-manager-installer.bin

[root@cdh4-cm-vm1 ~]# chmod u+x cloudera-manager-installer.bin

[root@cdh4-cm-vm1 ~]# ls -l
total 500
-rwxr--r-- 1 root root 504991 Jan 18 07:05 cloudera-manager-installer.bin
```

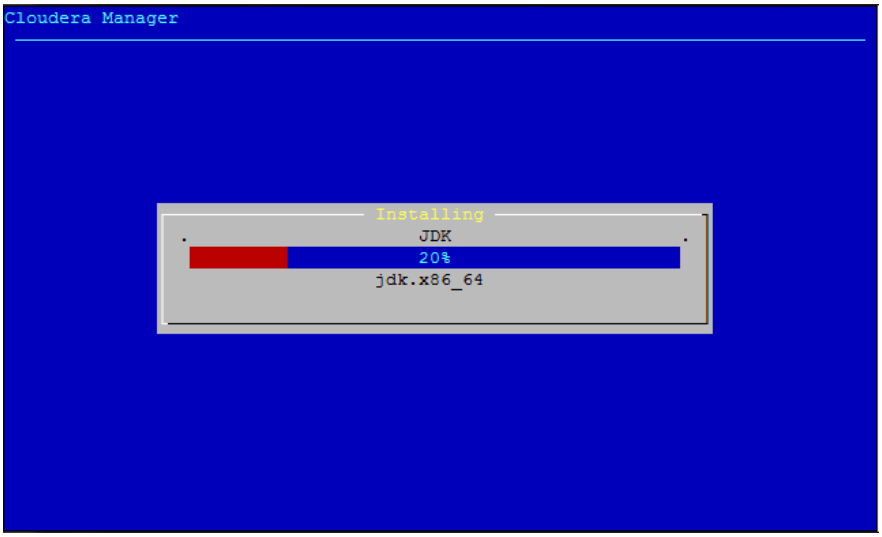
Run the Cloudera Manager installer:

```
[root@cdh4-cm-vm1 ~]# ./cloudera-manager-installer.bin
```

Press **Enter** to choose **Next** for the Readme file:

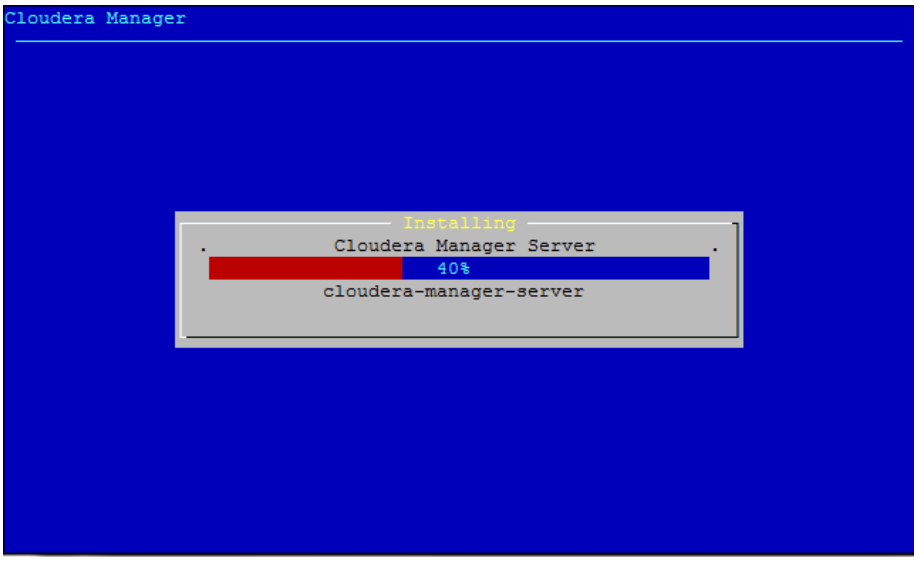


The Cloudera Manager installation now starts:



The Oracle JDK installation will take about 2 – 3 minutes and the installer may appear to be stuck at 20%. This is normal.

Then the Cloudera Manager Server installation will take a few minutes to install:



The installer will also install the embedded PostgreSQL database server.

When the installer completes, **make a note of the IP address and port** that it requests you to go to and then hit **OK.**

You should now be returned to the linux cmd prompt.

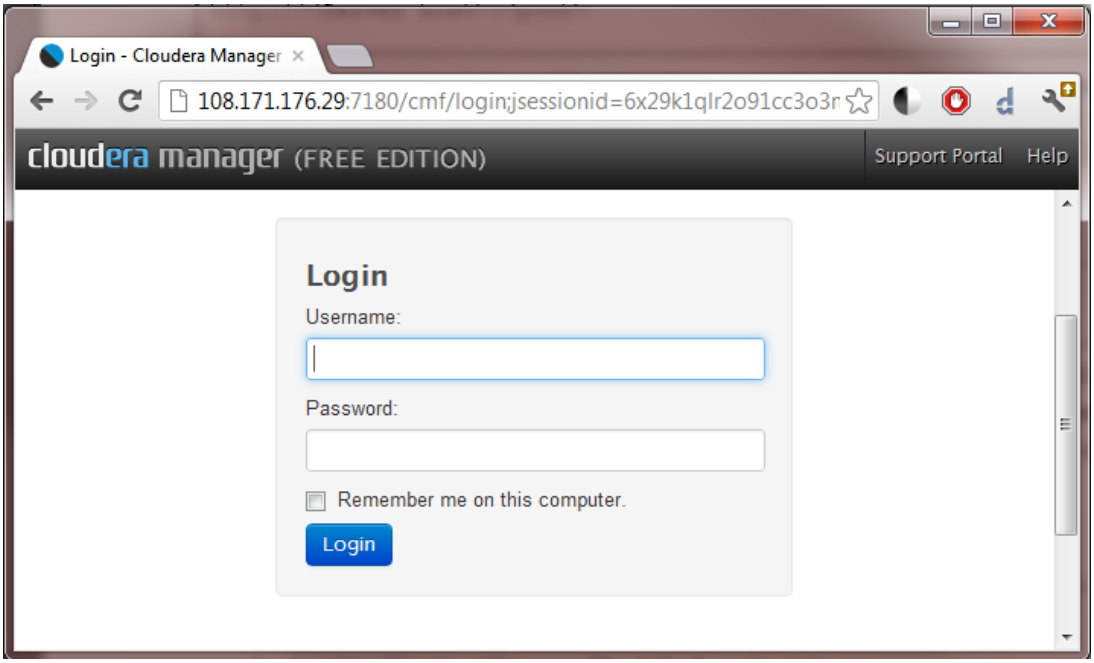
At the cmd prompt, verify that just Java got installed:

```
[root@cdh4-cm-vm1 ~]# java -version
java version "1.6.0_31"
Java(TM) SE Runtime Environment (build 1.6.0_31-b04)
Java HotSpot(TM) 64-Bit Server VM (build 20.6-b01, mixed mode)
```

Connect to the Cloudera Manager Admin Console:

So far, only Cloudera Manager has been installed on the VM. CDH4 has not been installed yet.

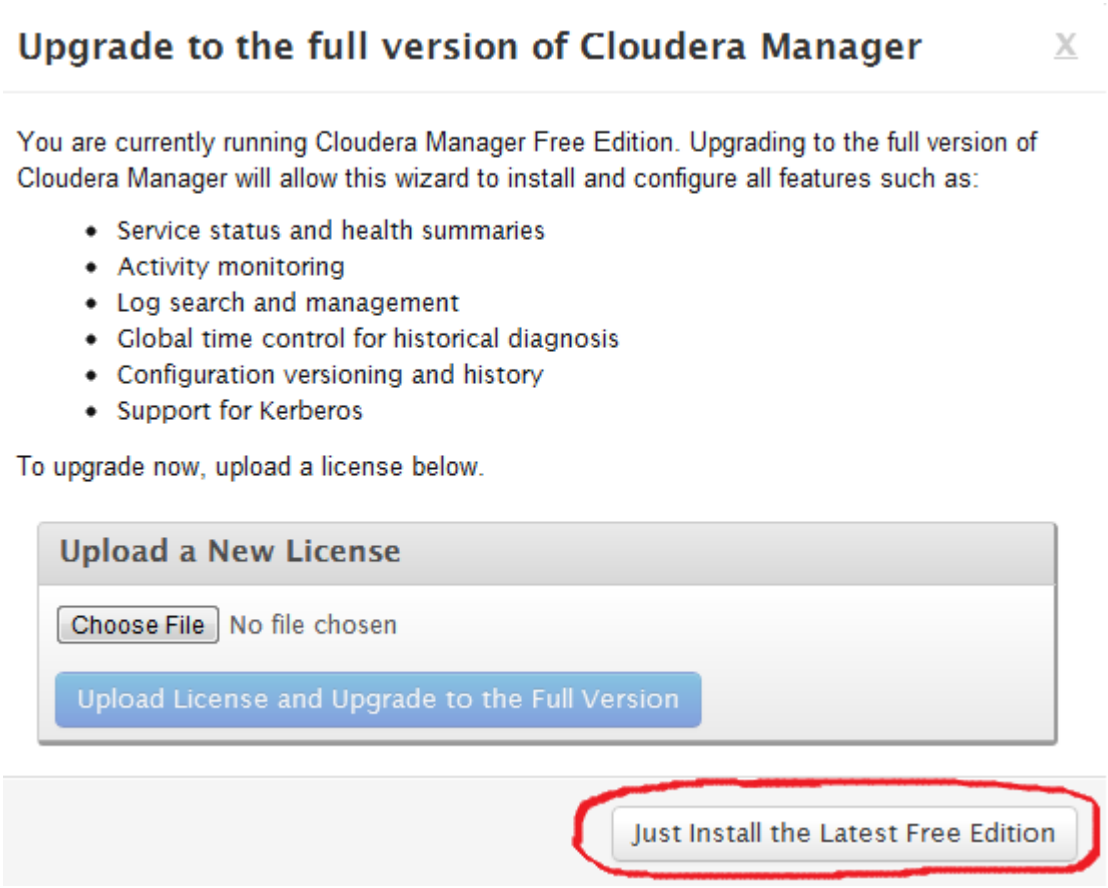
From Chrome go to the server URL for Cloudera Manager:
http://"<IP that you have noted fown in the previous step">:7180



Log in to Cloudera Manager using the default credentials:

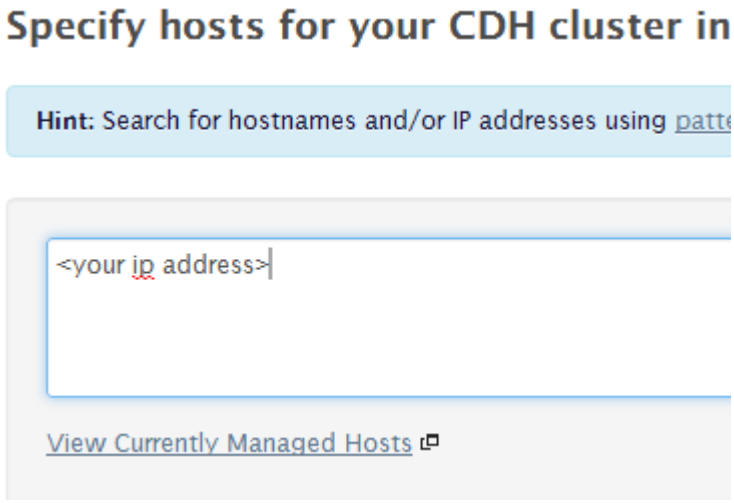
Username: **admin (now changed to cloudera)**
Password: **admin (now changed to wh!tesw@n)**

The Cloudera Manager Enterprise Edition pop up now launches, asking for an enterprise key. Click on "**Just Install the Latest Free Edition**" to bypass this screen.



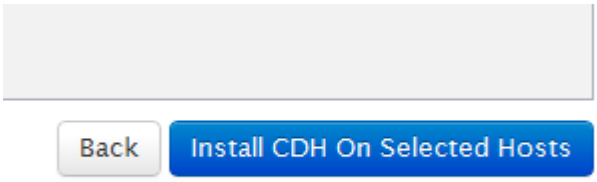
On the **Thank You** screen, just click **Continue** at the bottom to Continue.

Type in your IP address on the hosts page and click **Search**:



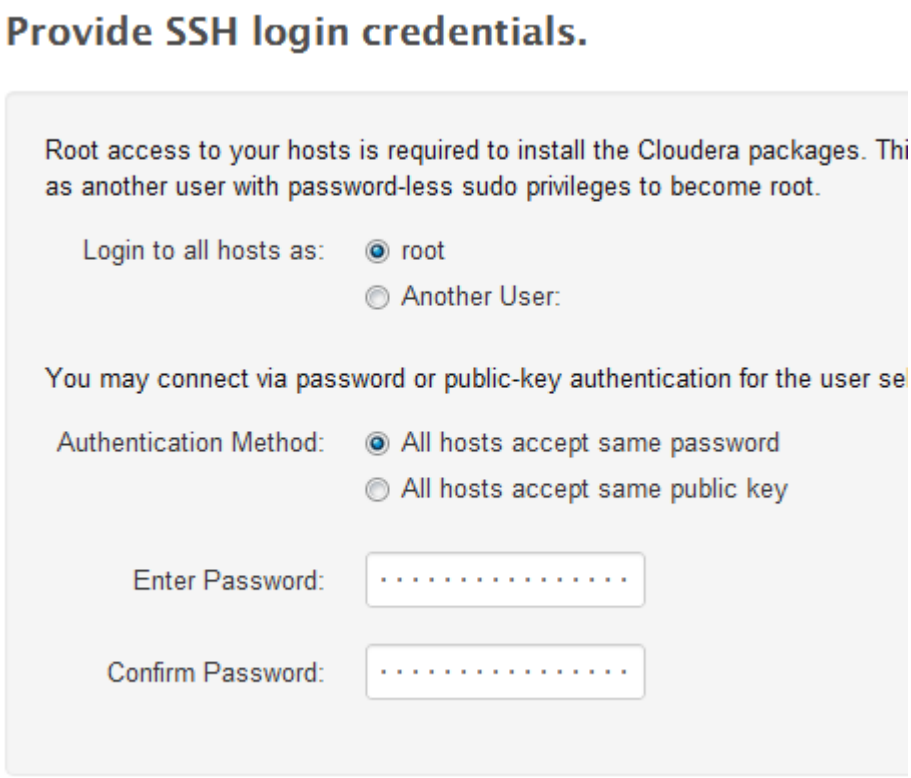
Your VM should now be found and automatically checked

Finally, click on **Install CDH On Selected Hosts**.

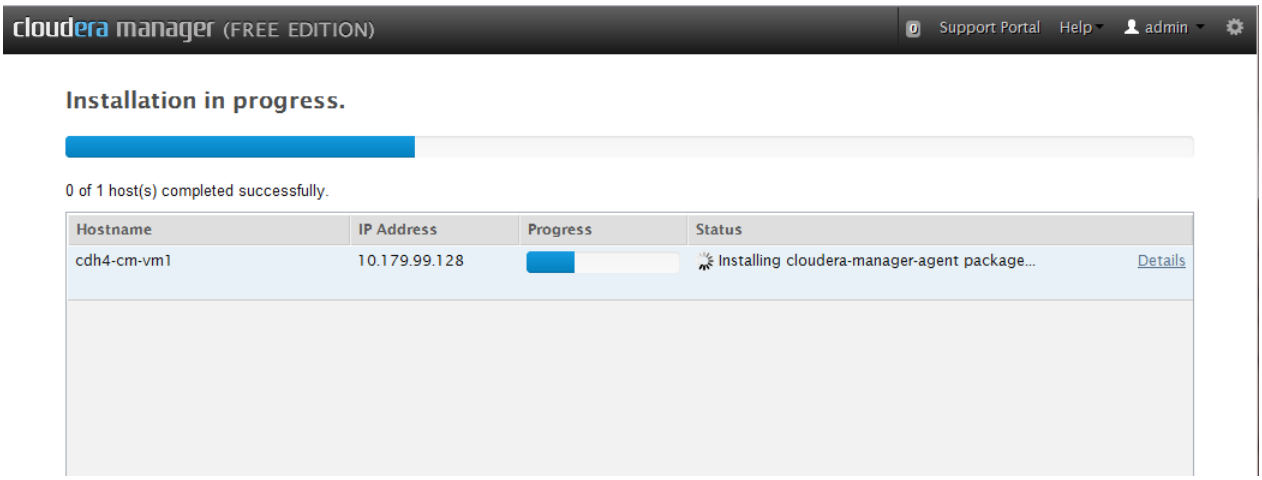


Choose **CDH4** / **CDH 4.1.2** / **None** / **Matched repository** to install the latest version of CDH4 and click **Continue**

In the SSH screen, login to all hosts as **root**. Choose the Authentication Method of "**All hosts accept same password**". Leave the SSH Port at 22 and the Number of simultaneous installations at 10. **The password is the root password of the system we are willing to install the setup – project123 in our case**. Then click "**Start Installation**".



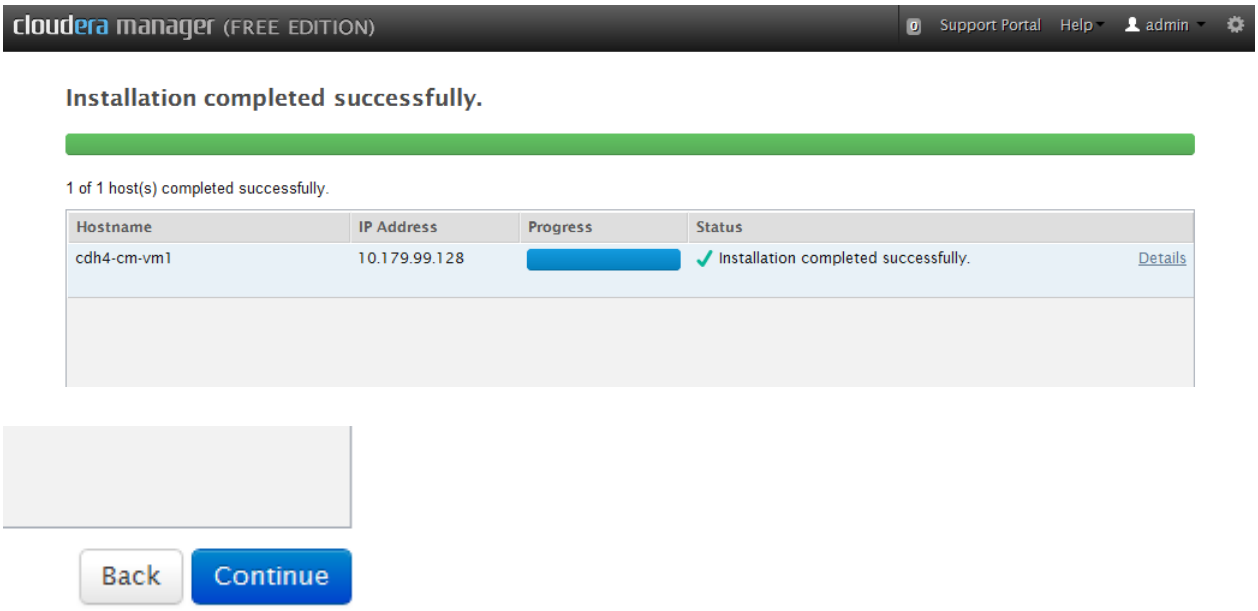
The installation will now kick off:



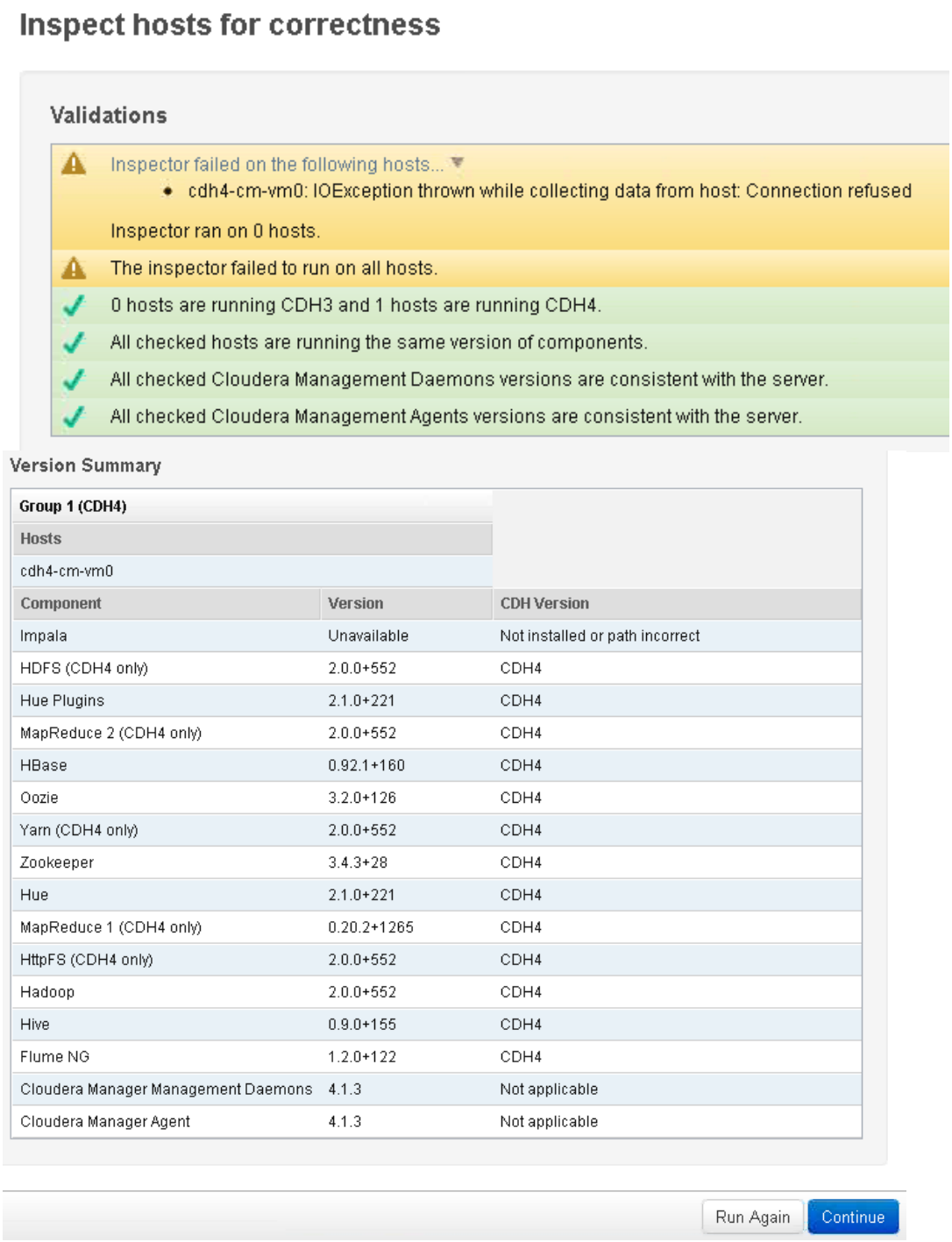
The installation will take about 5 minutes to complete. It is downloading CDH4 and installing it.

Consider browsing through some of those links at the top of this document while you wait.

When the Continue button appears at the bottom of the screen, the installation has completed, however the services have not started yet. Click Continue.



The Host Inspector page now appears, with a warning/error about failing to find the host. Ignore this and **continue**.



Inspect hosts for correctness

Validations

- Inspector failed on the following hosts...
  - cdh4-cm-vm0: IOException thrown while collecting data from host: Connection refused

  Inspector ran on 0 hosts.
- The inspector failed to run on all hosts.
- 0 hosts are running CDH3 and 1 hosts are running CDH4.
- All checked hosts are running the same version of components.
- All checked Cloudera Management Daemons versions are consistent with the server.
- All checked Cloudera Management Agents versions are consistent with the server.

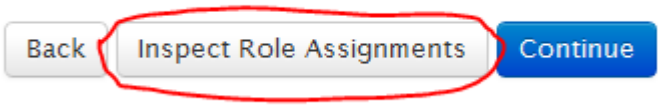
Version Summary

Group 1 (CDH4)		
Hosts		
cdh4-cm-vm0		
Component	**Version**	**CDH Version**
Impala	Unavailable	Not installed or path incorrect
HDFS (CDH4 only)	2.0.0+552	CDH4
Hue Plugins	2.1.0+221	CDH4
MapReduce 2 (CDH4 only)	2.0.0+552	CDH4
HBase	0.92.1+160	CDH4
Oozie	3.2.0+126	CDH4
Yarn (CDH4 only)	2.0.0+552	CDH4
Zookeeper	3.4.3+28	CDH4
Hue	2.1.0+221	CDH4
MapReduce 1 (CDH4 only)	0.20.2+1265	CDH4
HttpFS (CDH4 only)	2.0.0+552	CDH4
Hadoop	2.0.0+552	CDH4
Hive	0.9.0+155	CDH4
Flume NG	1.2.0+122	CDH4
Cloudera Manager Management Daemons	4.1.3	Not applicable
Cloudera Manager Agent	4.1.3	Not applicable

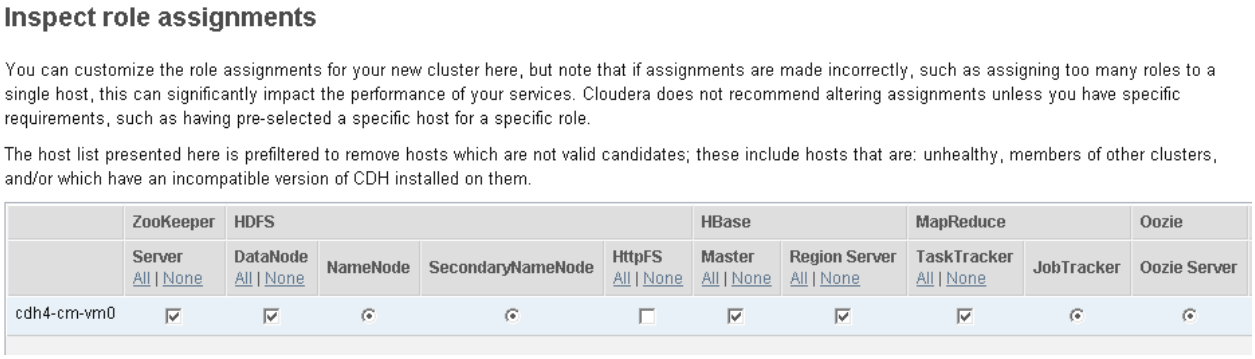
Run Again | Continue

On the “Choose the services that you want to start on your cluster” page, select **CDH4** and **All Services**.

Click **Inspect Role Assignments**.

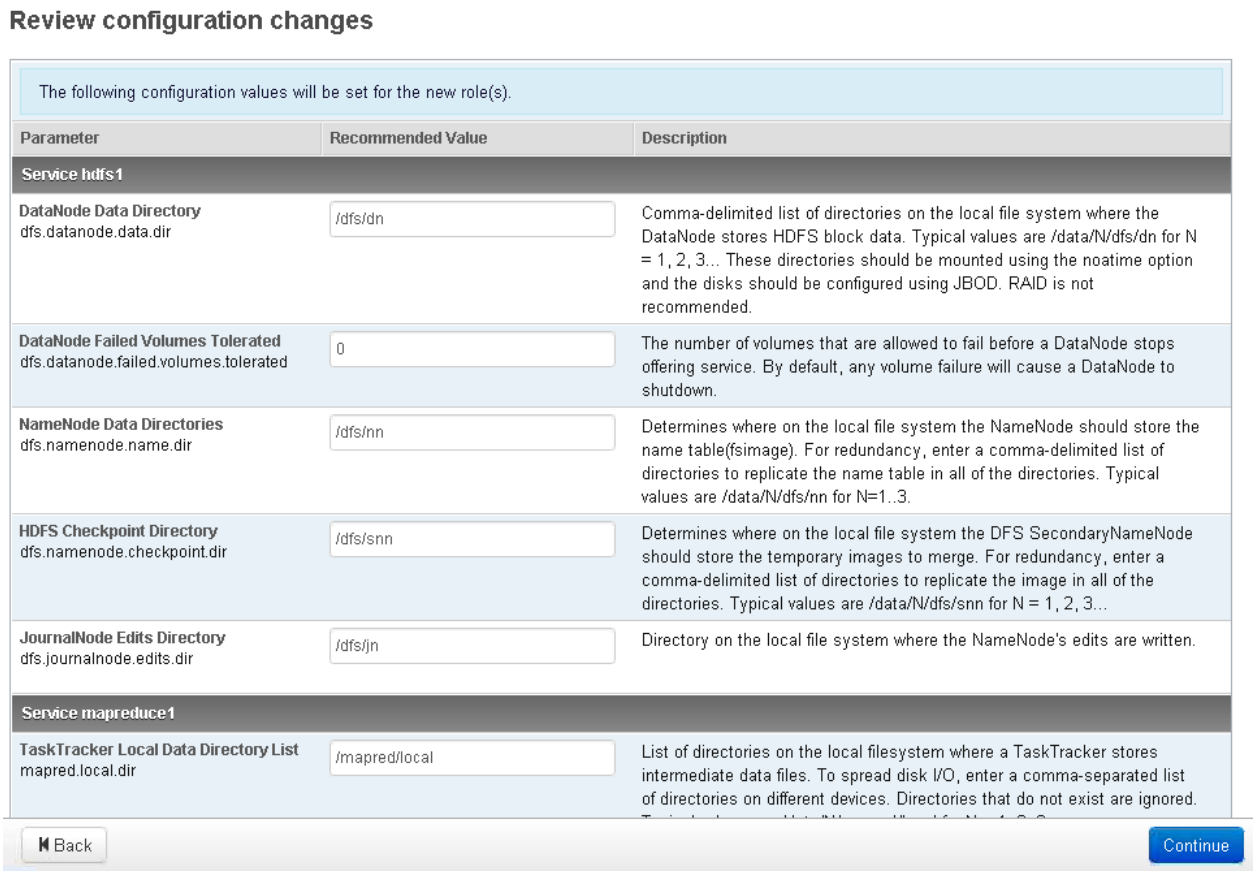


We won’t be changing the location of where we want to run these roles, but in a production cluster deployment, this is where you can re-shuffle the roles and customize where each role will run.

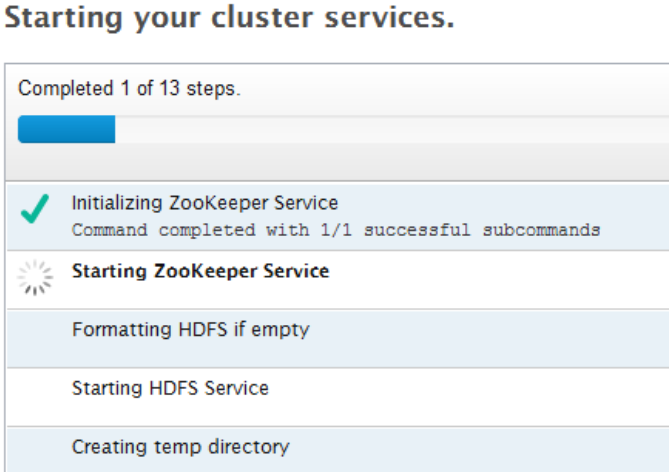


Click **Continue**.

On the “Review configuration changes” page, leave all of the default settings and just click **Continue**. The CM Installer will create all of those directories for you since they don’t currently exist.

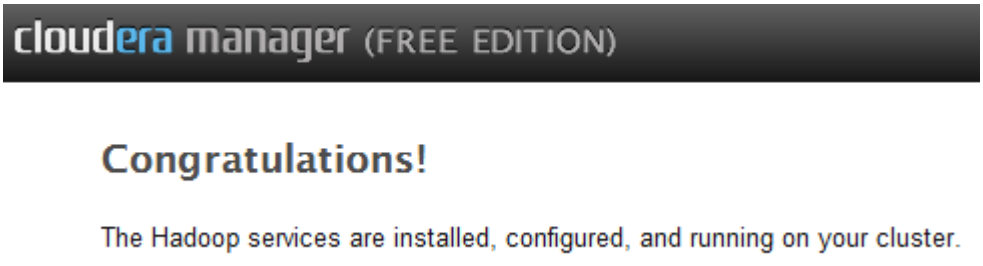


The cluster services will now begin to start. This process can take 3 to 5 minutes.

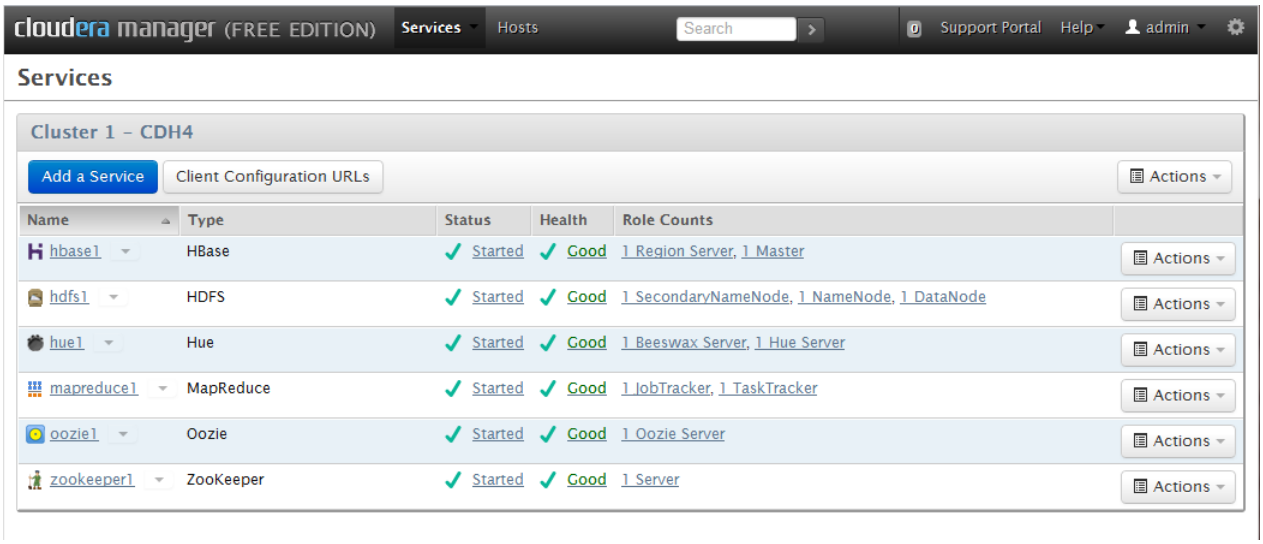


When the starting process is finished, click **Continue**.

You will now see a Congratulations screen. Pat yourself on the back and **Continue ;) :D**:



You should now see all of the Services running with **Good** health:



Final steps :

Now, return the PuTTy or Terminal window and do a final check to ensure that all the default Hadoop daemons are running.

```
[root@cdh4-cm-vm1 ~]# /usr/java/jdk1.6.0_31/bin/jps
7668 NameNode
8632 JobTracker
9199 Jps
7558 QuorumPeerMain
8842 Bootstrap
7736 SecondaryNameNode
8607 TaskTracker
8200 HMaster
8235 HRegionServer
2665 Main
7696 DataNode
9086 RunJar
```

```
[root@cdh4-cm-vm1 ~]# free -m
             total       used       free     shared    buffers     cached
Mem:          4010       3989         21          0         75       1742
-/+ buffers/cache:       2171       1839
Swap:         1953          0       1953
```

Finally, take note that this installation took 1.6 GB of disk space

```
[root@cdh4-cm-vm0 ~]# df -Th
Filesystem    Type    Size  Used Avail Use% Mounted on
/dev/xvda1    ext3    158G  2.5G  148G   2% /
tmpfs         tmpfs   1.9G     0  1.9G   0% /dev/shm
```

Possible Issues: (which we faced here)

HBase may crash. Since we are running Hadoop in a memory constrained environment, it is possible that HBase services may crash. This is okay and expected. Feel free to stop HBase service from the Cloudera Manager GUI's main page by clicking "Actions" and Stop next to the HBase service. We can start it when needed.