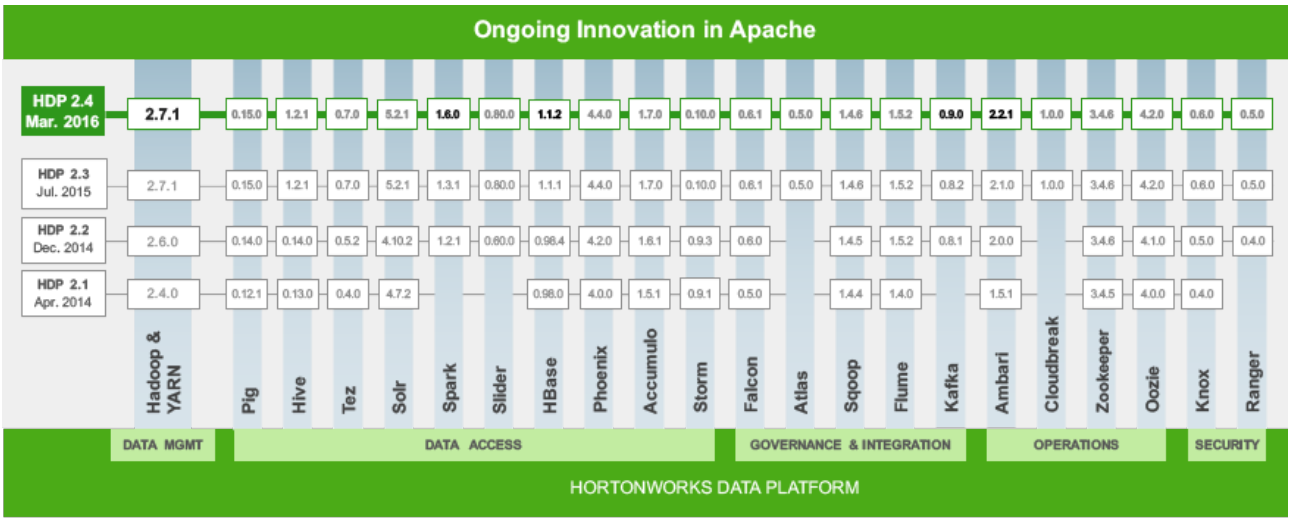
Multi Node Apache Hadoop Ecosystem Enterprise Setup and Configuration Ubuntu 14.04 Machine's with Hortonworks HDP 2.4 Distribution

Data Science Lab, The Department of Computer Science, KSKV Kachchh University.

Web: http://cs.kutchuni.edu.in

HDP is the industry's only true secure, enterprise-ready open source Apache™ Hadoop® distribution based on a centralized architecture (YARN). HDP addresses the complete needs of data-at-rest, powers real-time customer applications and delivers robust analytics that accelerate decision making and innovation.



[Figure 1] : Hortonworks HDP 2.4 Data Platform

Installation of HDP 2.4 with Ambari 2.2.2.0

Step 1: List Ambari 2.2.2.0 from Public repository

sudo wget -nv http://public-repo-1.hortonworks.com/ambari/ubuntu14/2.x/updates/2.2.2.0/ambari.list -O /etc/apt/sources.list.d/ambari.list

Step 2: Accessing keyserver key

sudo apt-key adv --recv-keys --keyserver keyserver.ubuntu.com B9733A7A07513CAD

Step 3: Update the Ubuntu Repository

sudo apt-get update

Step 4: do package handling utility -- cache manipulator

sudo apt-cache showpkg ambari-server

sudo apt-cache showpkg ambari-agent

sudo apt-cache showpkg ambari-metrics-assembly

Step 5: Install Ambari Server

sudo apt-get install ambari-server

Step 6: Setup Ambari Server Database and configuration

sudo ambari-server setup

Step 7: Install Ambari agent

sudo apt-get install ambari-agent

Step 8: Start Ambari Server

sudo ambari-server start

Step 9: Start Ambari Agent

sudo ambari-agent start

Step 10: for Checking status of Ambari server and Ambari agent

sudo ambari-server status

sudo ambari-agent status

Step 11: for stopping Ambari Server and Ambari agent

sudo ambari-server stop

sudo ambari-agent stop

Step 12: Login Ambari

http://hostname:8080

*Step 13: Start All services [**Note:** make sure everything setup has been done as mentioned in **Appendix-A**]*

Step 14: For starding individual Hadoop Eco system service's by shell

/usr/hdp/2.4.2.0-258/

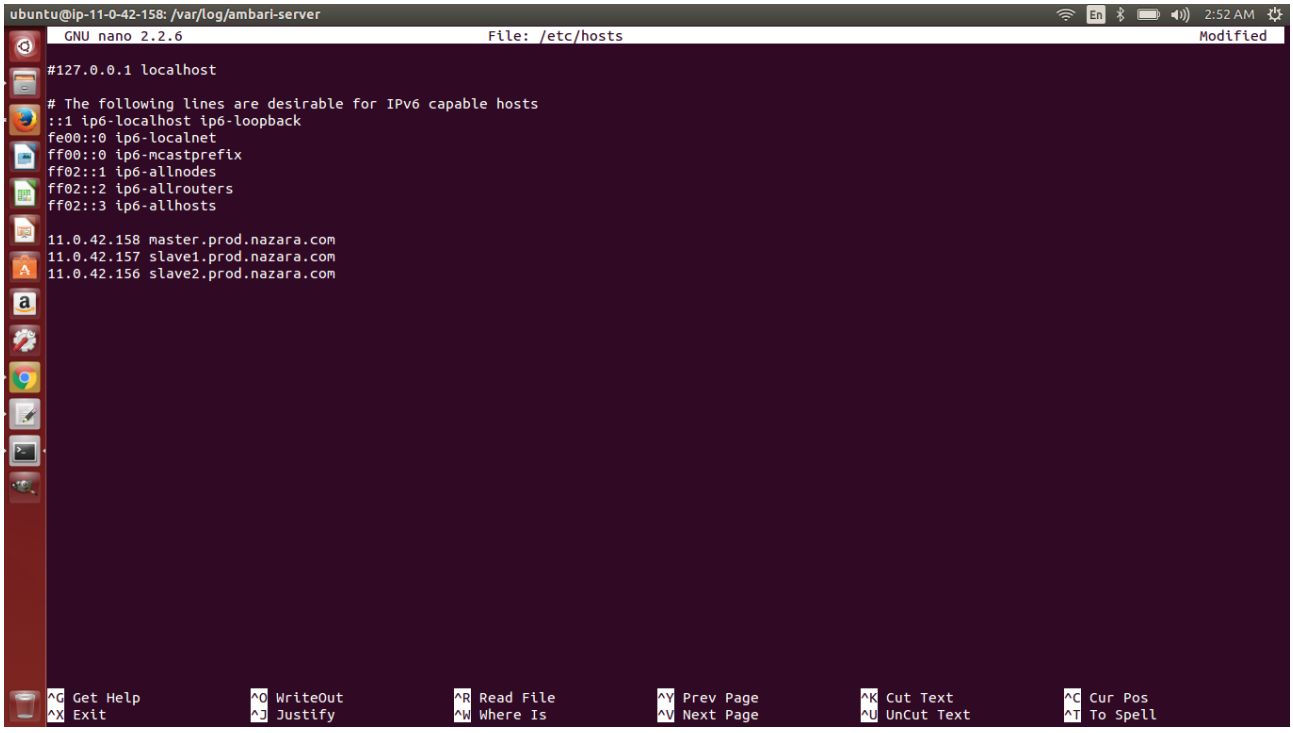
Appendix A : Configuration

1) /etc/hosts

Follow following steps in every cluster of Hadoop (master and slaves).

- comment localhost line
- provide master and slave ip address and name server
- Ex.
11.0.42.158 master.prod.nazara.com
11.0.42.157 slave1.prod.nazara.com
11.0.42.156 slave2.prod.nazara.com

[Note: above are the private IP addresses , currently we get from AWS]



[Figure 2]: /etc/hosts configuration on every Hadoop nodes (master and slaves).

2) /etc/hostname

Configure below steps on every hadoop node (master + slaves)

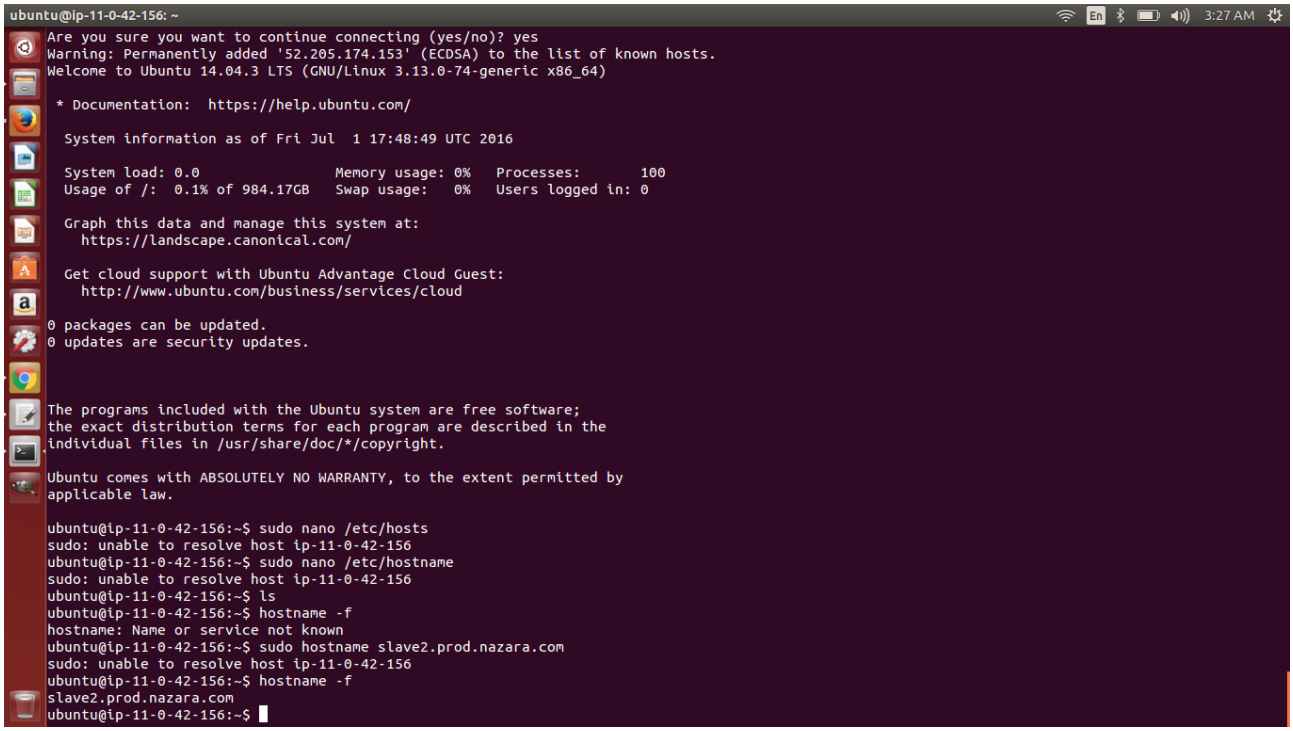
- open /etc/hostname
- comment out by default given IP Address
- put individual machine's own Mapped server name
- example,
If it is Master node, we will put
11.0.42.158 master.prod.nazara.com
If it is Slave node, we will put
11.0.42.157 slave1.prod.nazara.com
and accordingly for other slaves.

3) Register hostname

- Go to every Hadoop Nodes(master, slaves)
- Provide appropriate hostname for every master and slaves.
- Example.

If it is master node then example,
Sudo hostname master.prod.nazara.com
If it is slave node then example,
Sudo hostname slave1.prod.nazara.com

For validation,
$ hostname – f

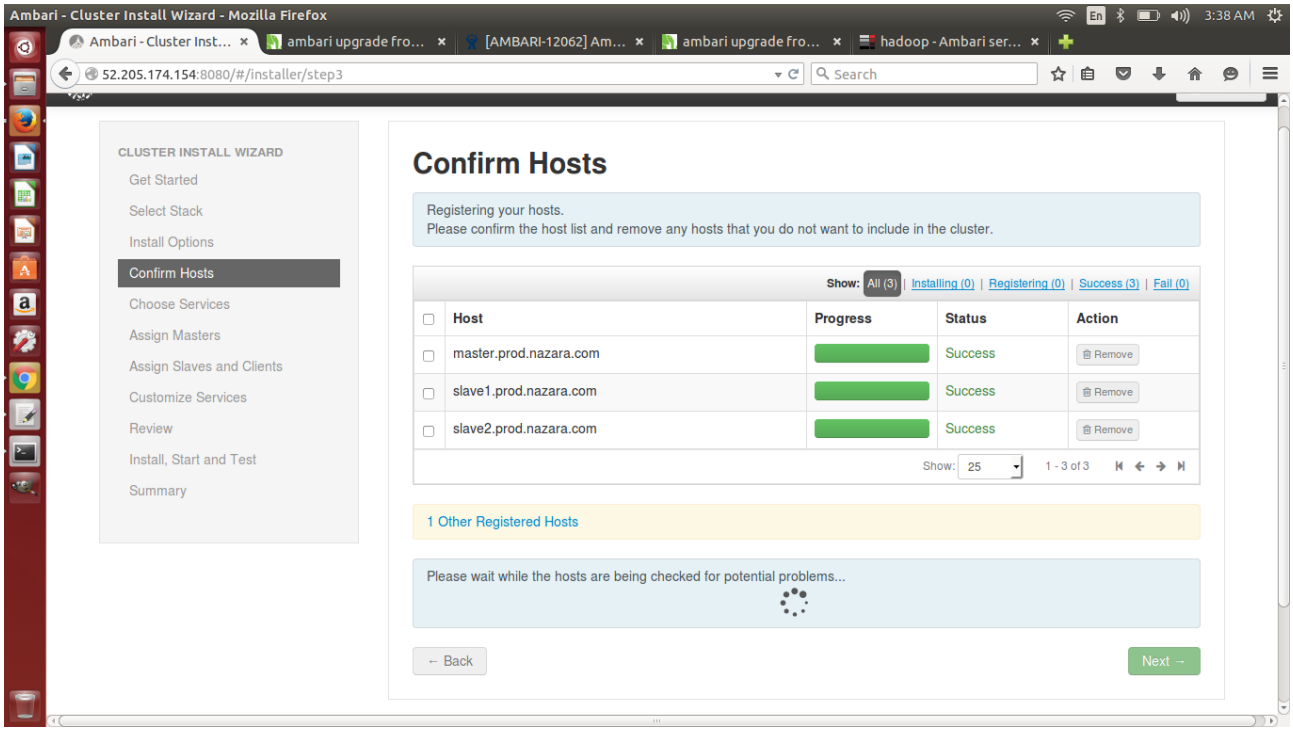


[Figure 3]: Set hostnme to appropriate name server address.

Appendix B : Ambari UI Setup

1) Confirm hosts

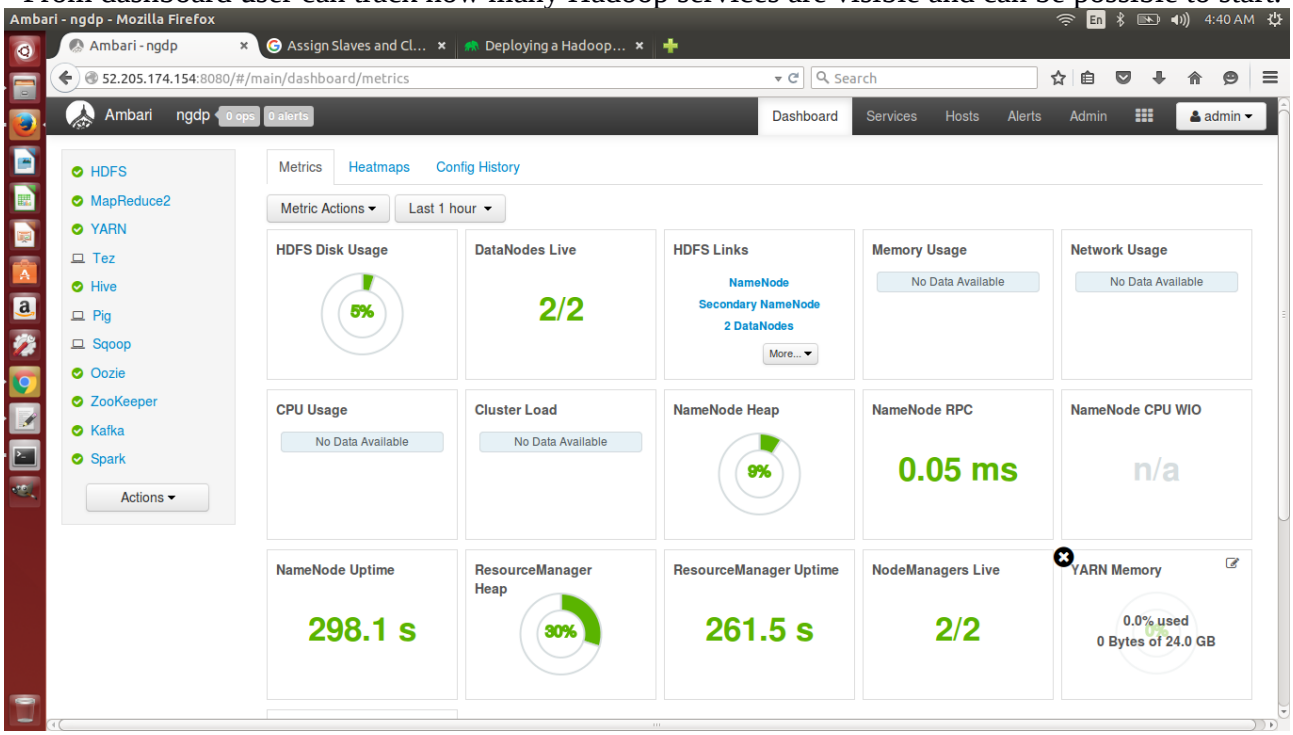
- User needs to configure how many masters and slaves are being required according to requirements.
- ***Note:*** *make sure everything setup has been done as mentioned in* ***Appendix-A***



[Figure 4]: Ambari Master and slave nodes Configurations.

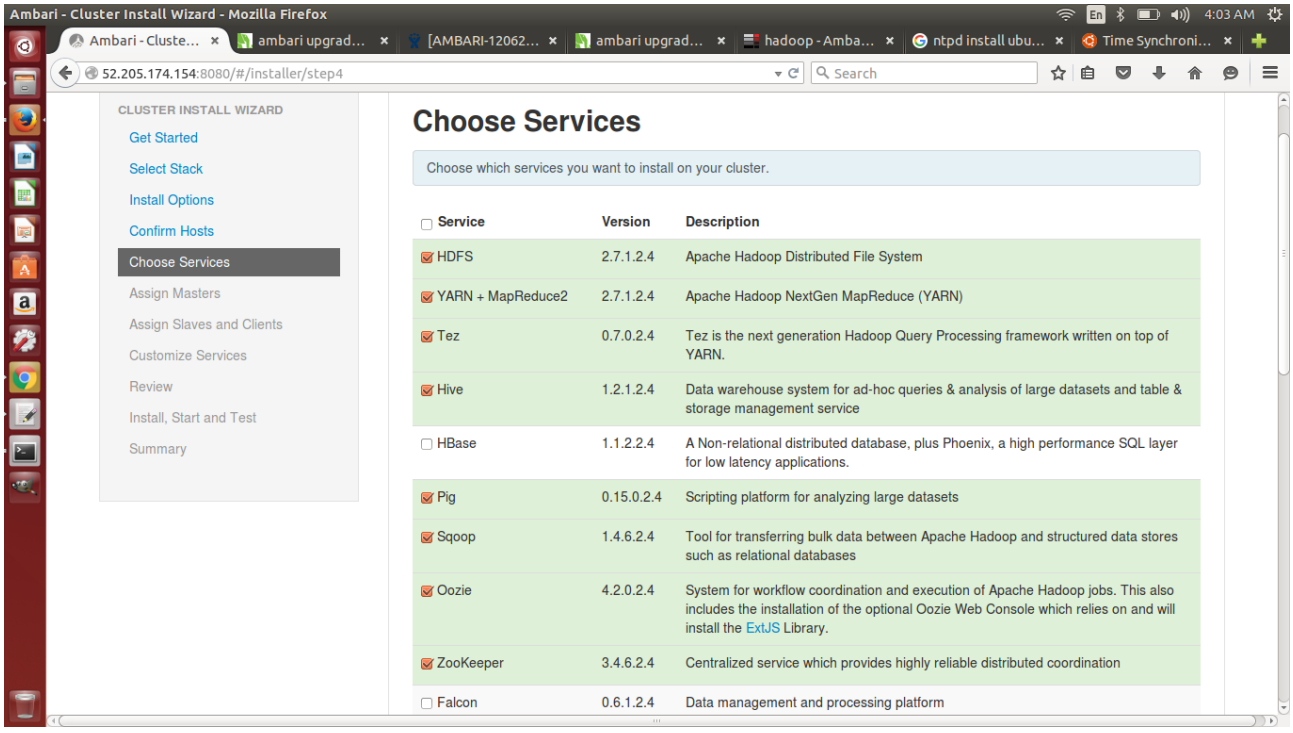
2) Ambari UI Dashboard

- After accessing Ambari UI from http://hostname:8080 with admin / admin or appropriate credentials user will get redirected to Dashboard.
- From dashboard user can track how many Hadoop services are visible and can be possible to start.

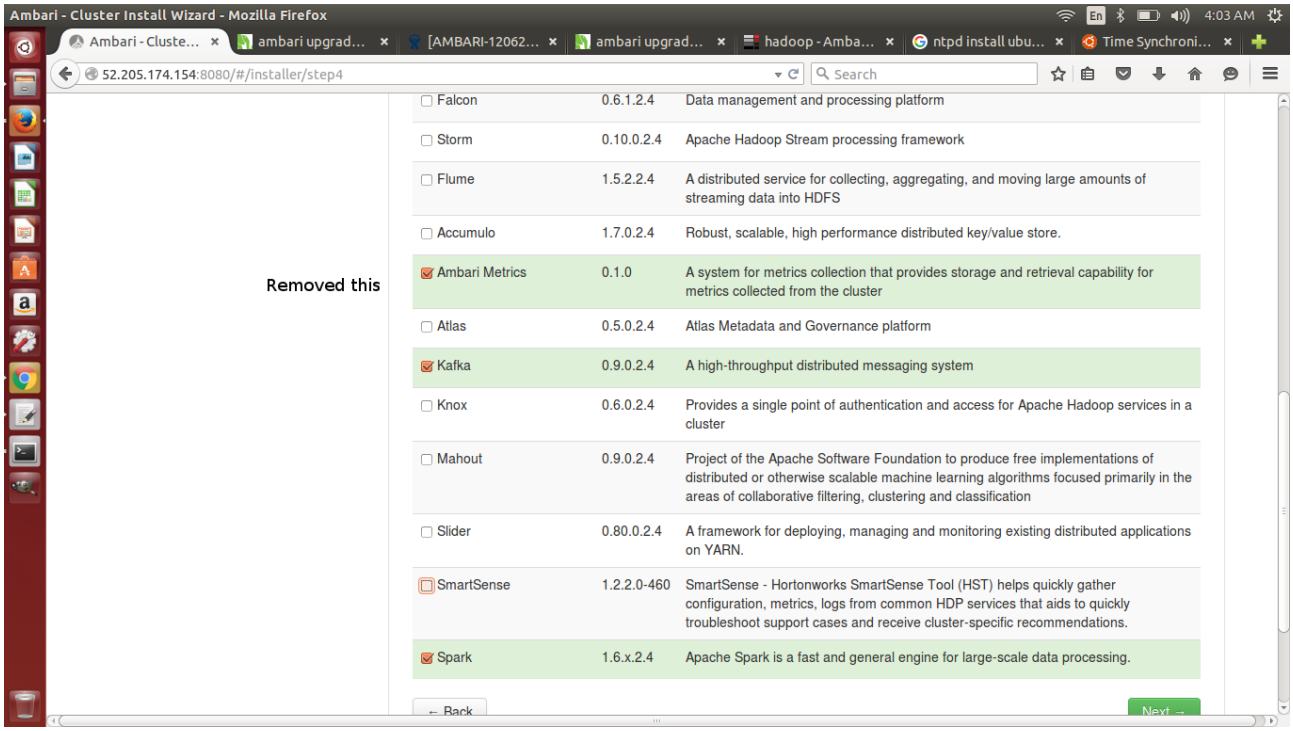


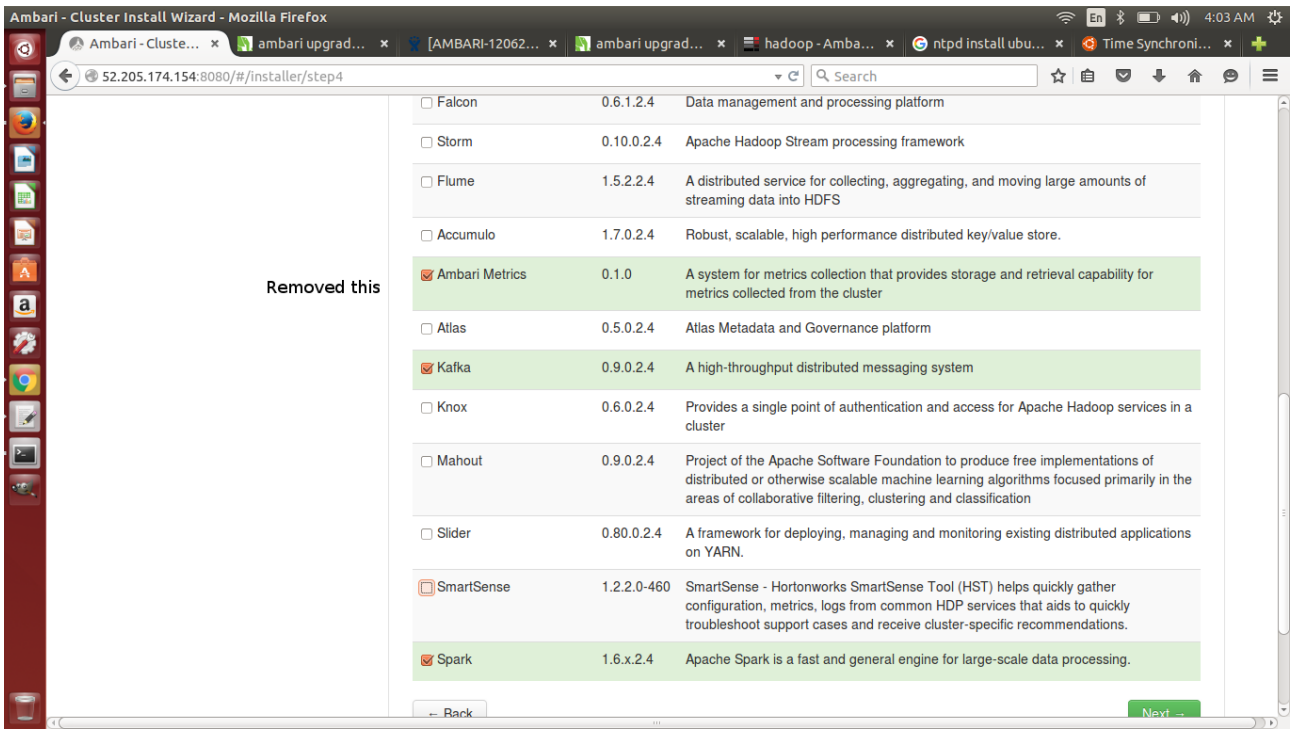
[Figure 4]: Ambari Dashboard

3) Selected Hadoop Ecosystem services



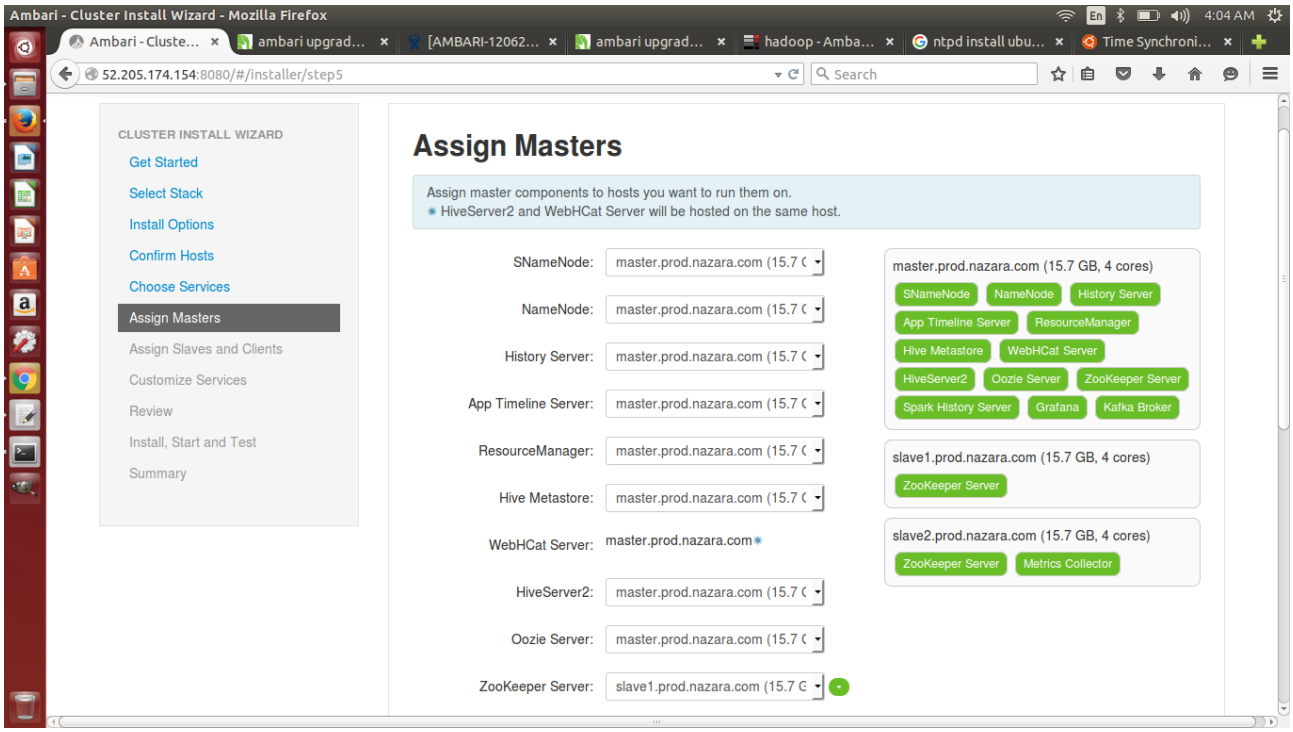
[Figure 5]: Choose Hadoop related services with Ambari





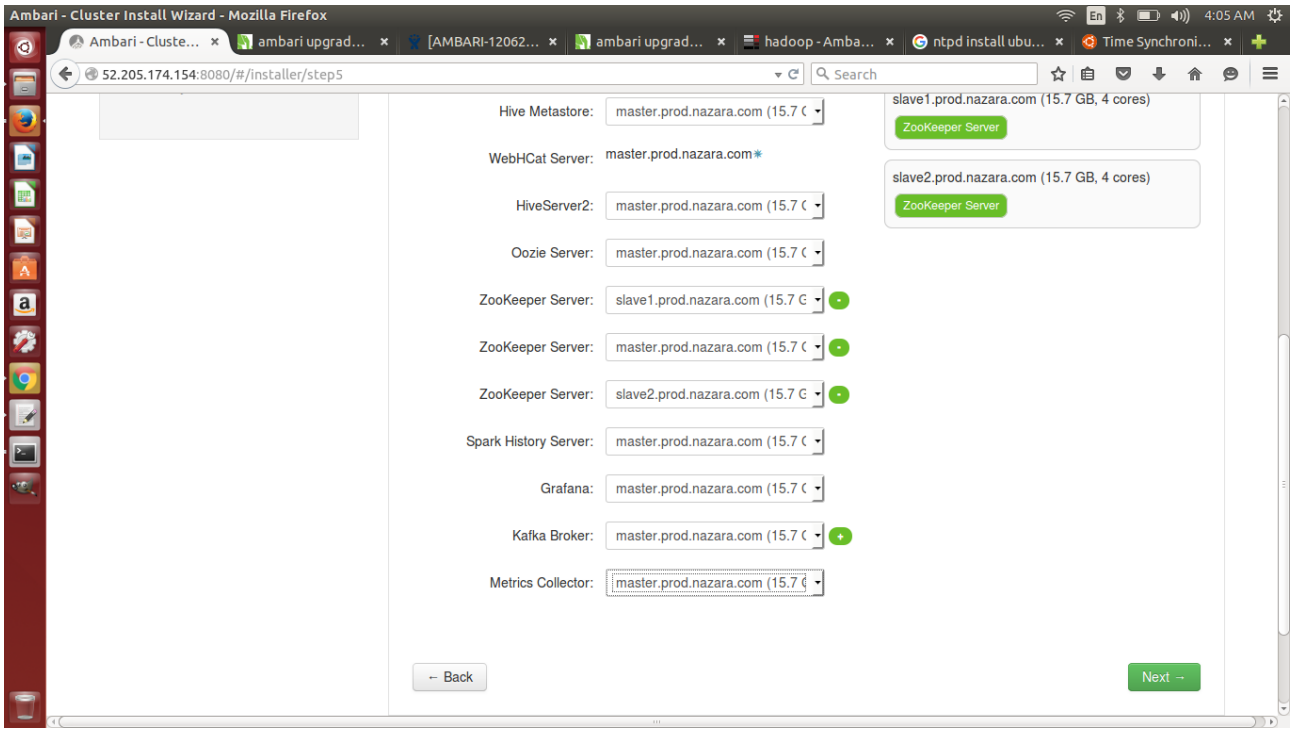
[Figure 6]: Choose Hadoop related services with Ambari

4) Assign Hadoop Masters



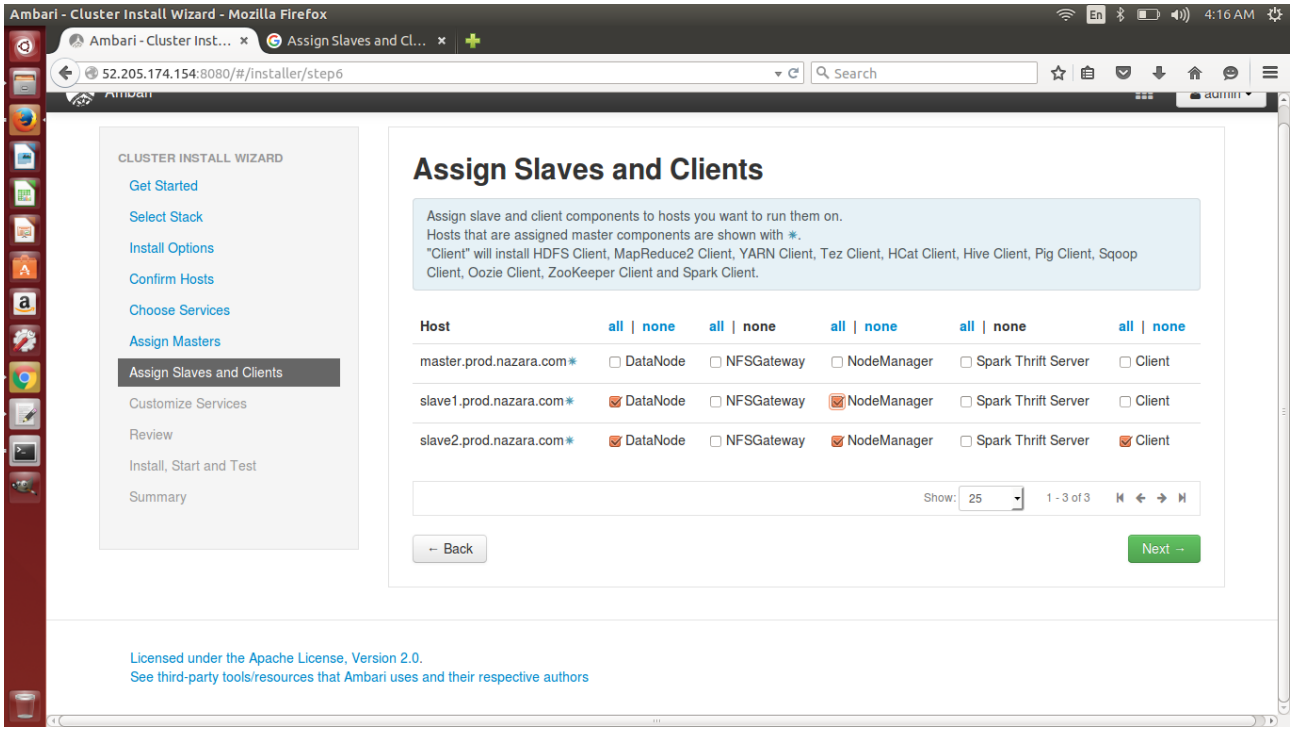
[Figure 7]: Assign Master / Slave node in Hadoop related services

5) Assign Slaves and Client



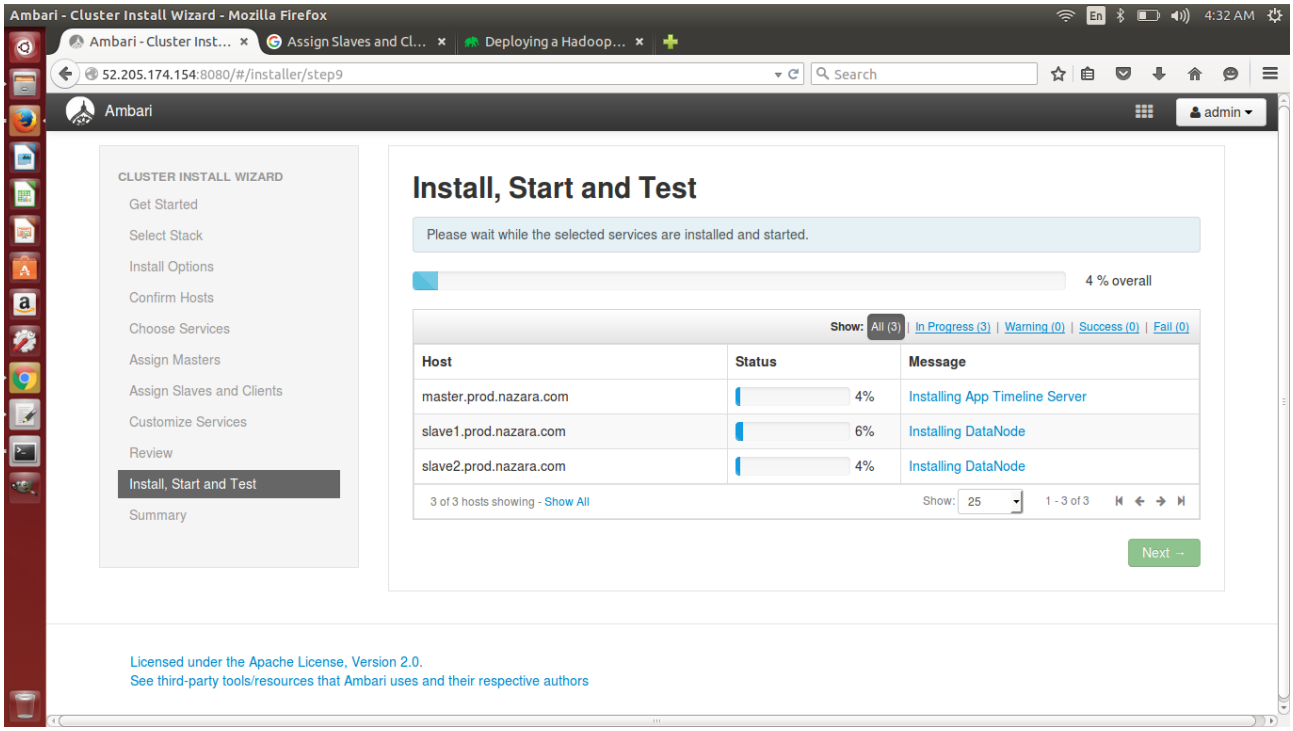
[Figure 8]: Assign Master / Slave node in Hadoop related services

6) Install Master and slaves



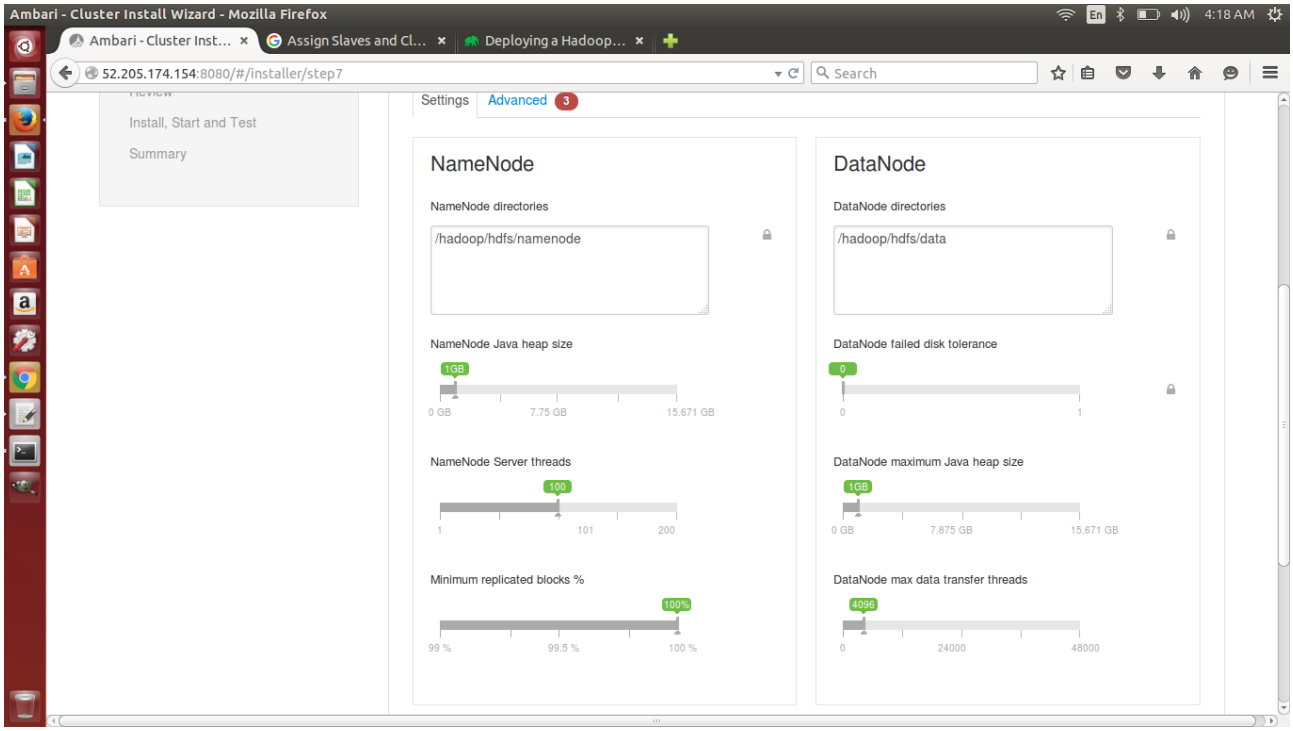
[Figure 9]: Assign slaves and clients

6) Install , Start and Test Masters and Slaves



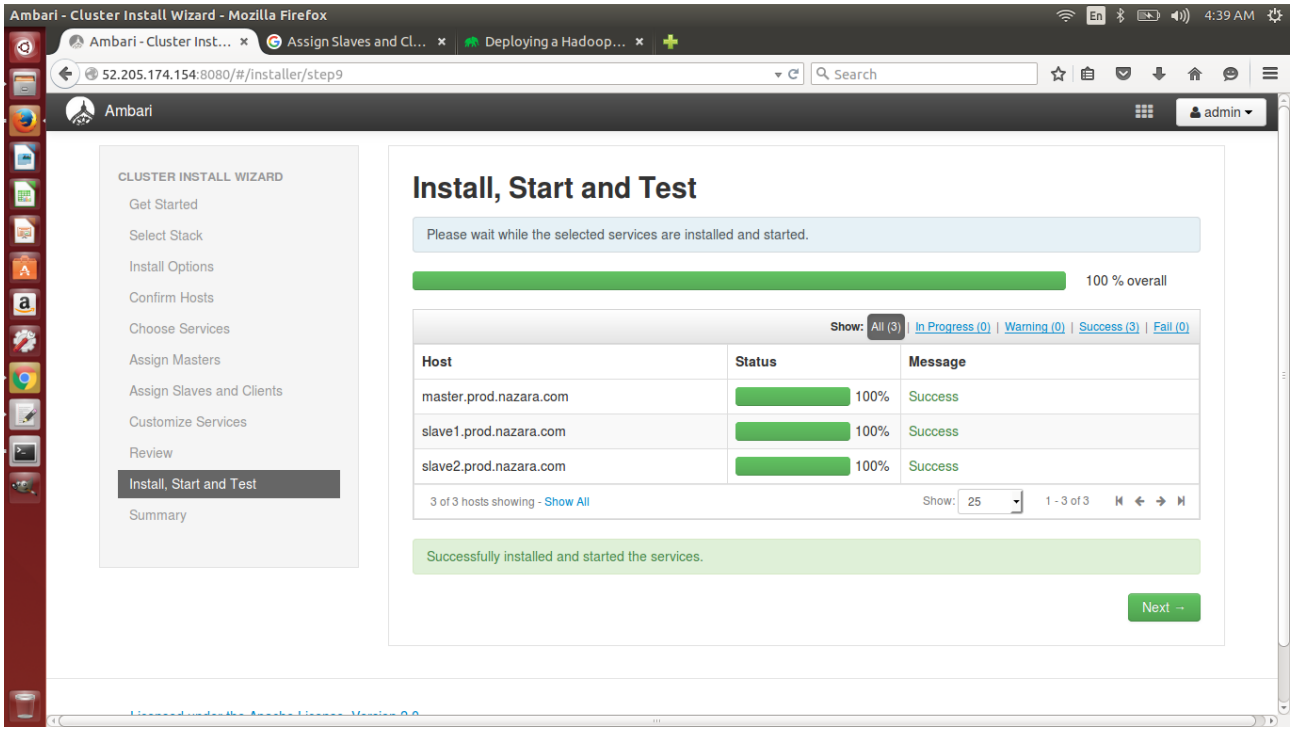
[Figure 10]: Install, Start and Test Masters and Slaves.

7) Namenode and Datanode configuration on Ambari



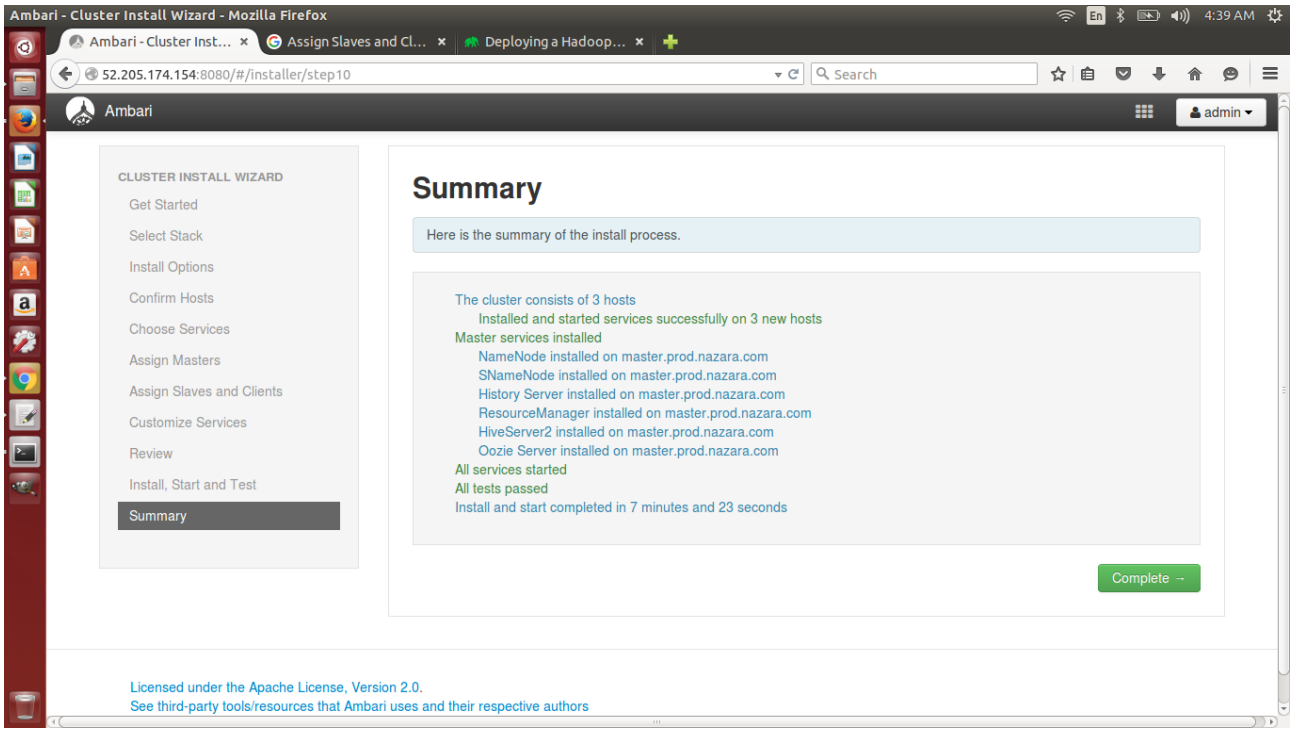
[Figure 11]: Namenode and Datanode Advanced configuration on Ambari.

8) Testing Master and Slave



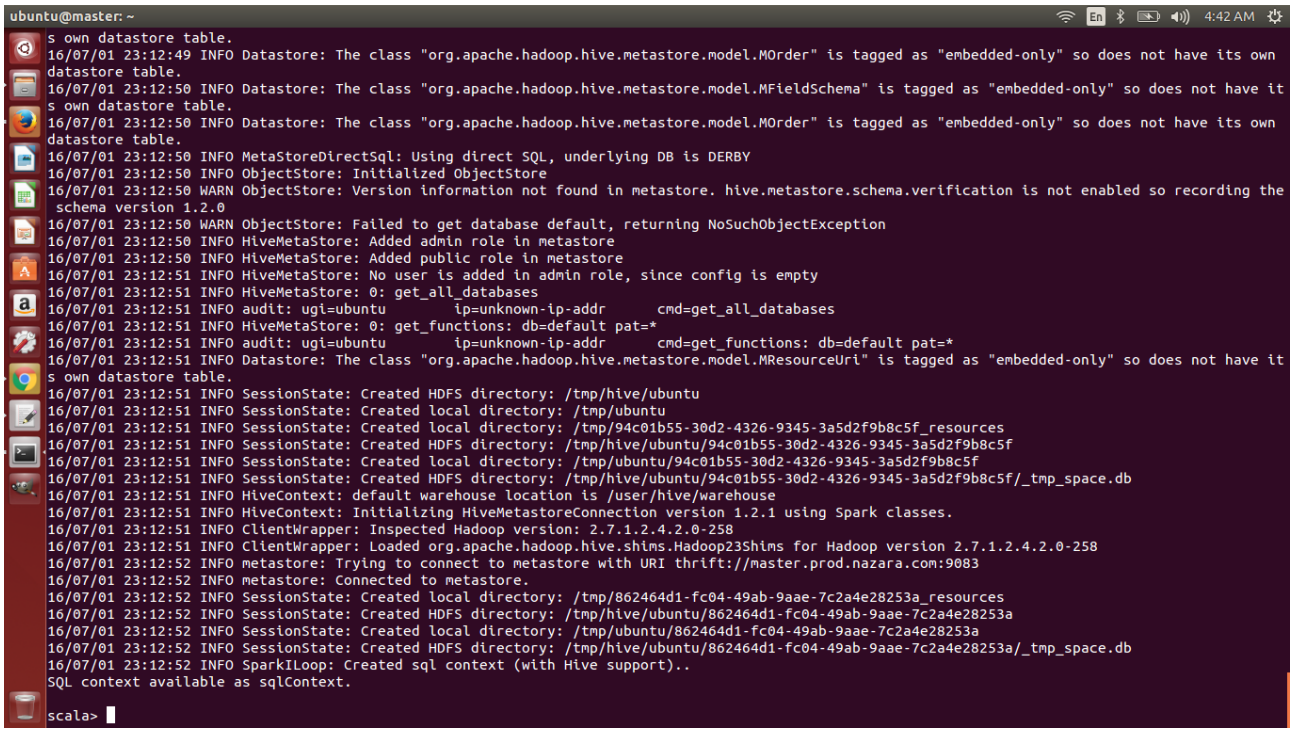
[Figure 12]: Install Start and Test Hadoop masters and Slaves.

9) Installation Summary Process



[Figure 13]: Ambari Installation Summary

10) starting spark-shell from Ambari



[Figure 14]: starting spark-shell fromo Ambari

Below possible Errors that might occure:

1) Ambari Server java process died with exitcode 255. Check /var/log/ambari-server/ambari-server.out for more information.

Solution: sudo ambari-server upgrade

2) ERROR: Exiting with exit code -1

Solution: changed hostname in /etc/hosts and value available at /etc/hostname should be there. Ex. 11.0.42.158 master.prod.nazara.com

3) Host Checks found 2 issues on 3 hosts.
After manually resolving the issues, click Rerun Checks.
To manually resolve issues on each host run the HostCleanup script (Python 2.6 or greater is required)

Solution:
sudo apt install ntp
sudo service ntp restart

4) If anything goes wrong, and if DevOps failed to troubleshoot

Solution: In worst case only

```
sudo ambari-server reset
```