

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>

The MIT License (MIT)

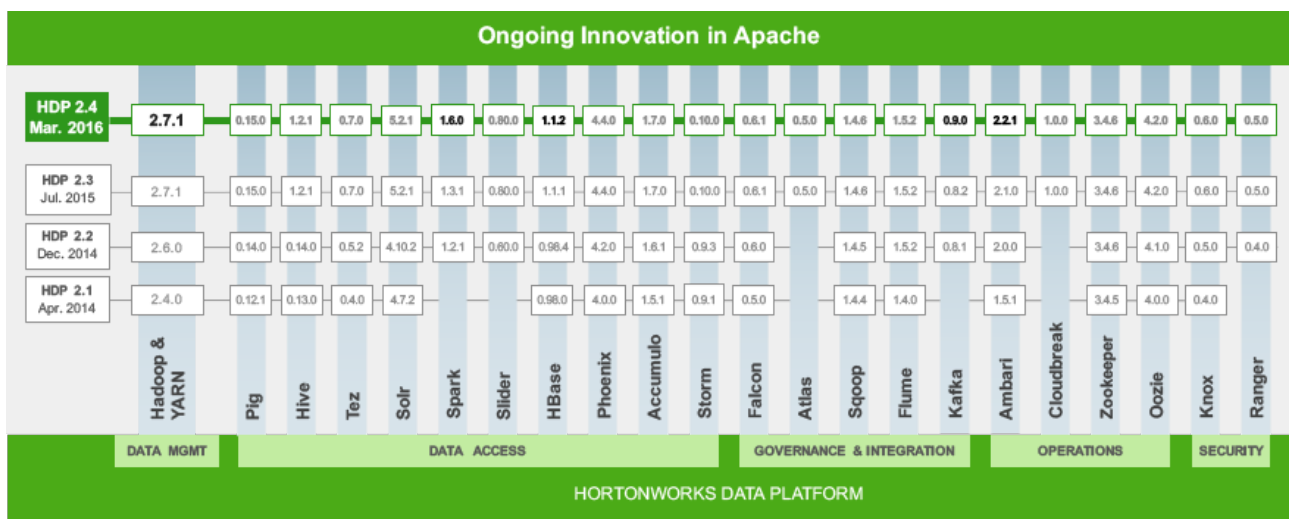
Copyright (c) 2016. Data Science Lab, University of Kachchh.

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

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 starting 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).

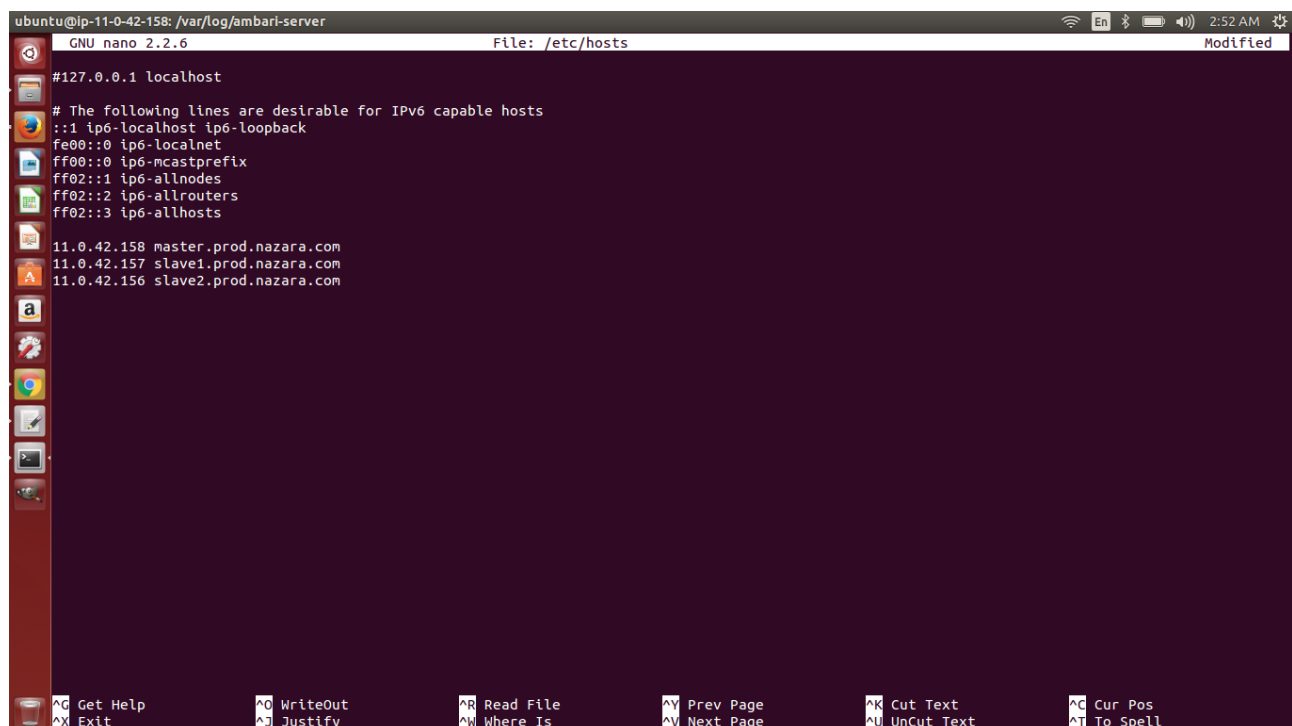
- comment localhost line
- provide master and slave ip address and name server
- Ex.

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]

A screenshot of a terminal window on an Ubuntu system. The terminal title bar shows 'ubuntu@ip-11-0-42-158: /var/log/ambari-server'. The window displays the contents of the /etc/hosts file being edited with the nano text editor. The file content includes a commented-out localhost entry, IPv6 loopback and multicast addresses, and three entries for the Hadoop cluster nodes: 11.0.42.158 master.prod.nazara.com, 11.0.42.157 slave1.prod.nazara.com, and 11.0.42.156 slave2.prod.nazara.com. The nano editor interface shows the file name 'File: /etc/hosts' and a 'Modified' status. The bottom status bar of the nano editor displays various keyboard shortcuts like '^O Get Help', '^X Exit', '^W WriteOut', '^J Justify', '^R Read File', '^M Where Is', '^Y Prev Page', '^V Next Page', '^K Cut Text', '^U UnCut Text', '^C Cur Pos', and '^T To Spell.

[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,

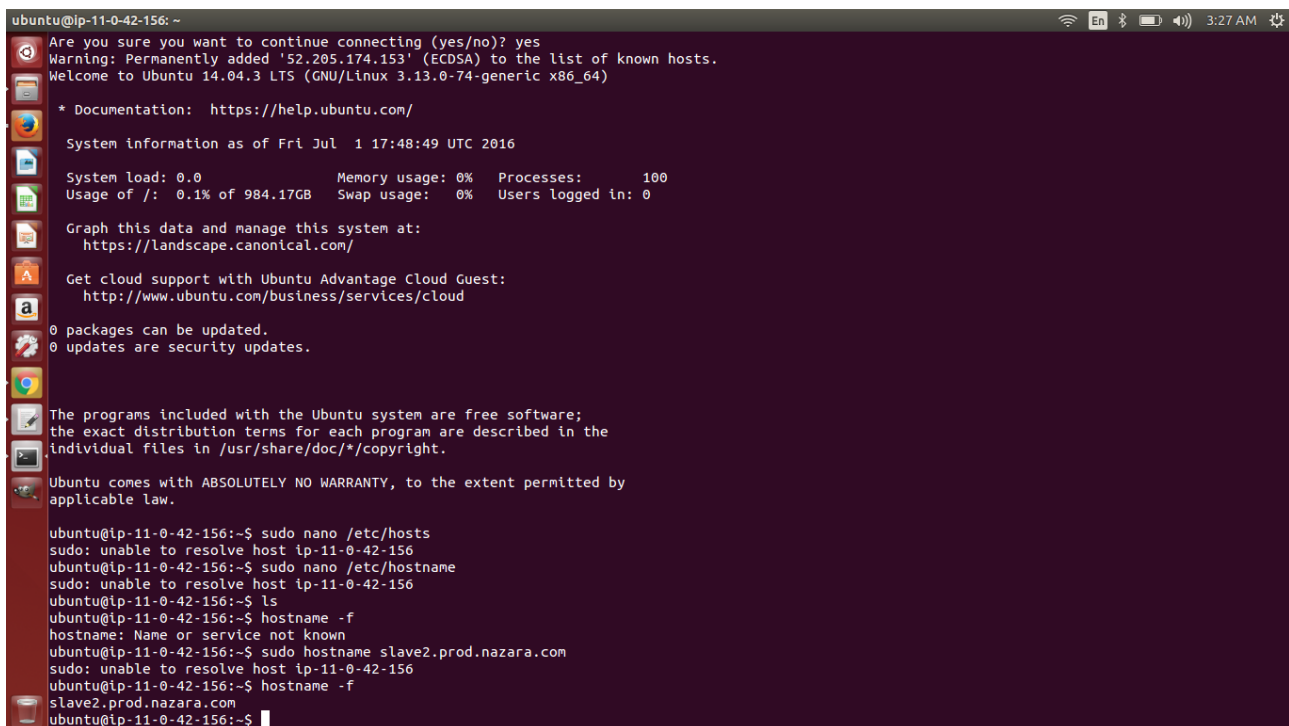
Sudo hostname master.prod.nazara.com

If it is slave node then example,

Sudo hostname slave1.prod.nazara.com

For validation,

\$ hostname -f



```
ubuntu@ip-11-0-42-156: ~
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added '52.205.174.153' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 14.04.3 LTS (GNU/Linux 3.13.0-74-generic x86_64)

* Documentation:  https://help.ubuntu.com/

System information as of Fri Jul  1 17:48:49 UTC 2016

System load: 0.0               Memory usage: 0%   Processes:   100
Usage of /:  0.1% of 984.17GB  Swap usage:  0%   Users logged in: 0

Graph this data and manage this system at:
https://landscape.canonical.com/

Get cloud support with Ubuntu Advantage Cloud Guest:
http://www.ubuntu.com/business/services/cloud

0 packages can be updated.
0 updates are security updates.

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

ubuntu@ip-11-0-42-156:~$ sudo nano /etc/hosts
sudo: unable to resolve host ip-11-0-42-156
ubuntu@ip-11-0-42-156:~$ sudo nano /etc/hostname
sudo: unable to resolve host ip-11-0-42-156
ubuntu@ip-11-0-42-156:~$ ls
ubuntu@ip-11-0-42-156:~$ hostname -f
hostname: Name or service not known
ubuntu@ip-11-0-42-156:~$ sudo hostname slave2.prod.nazara.com
sudo: unable to resolve host ip-11-0-42-156
ubuntu@ip-11-0-42-156:~$ hostname -f
slave2.prod.nazara.com
ubuntu@ip-11-0-42-156:~$
```

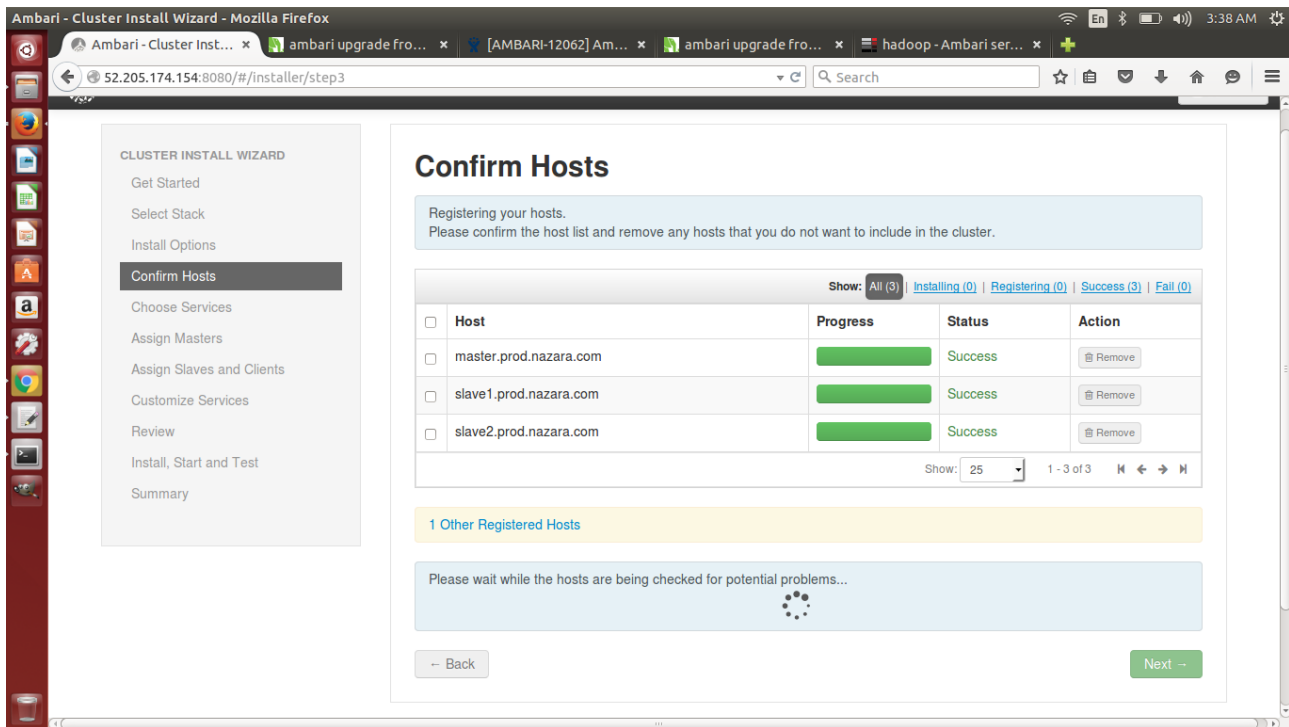
[Figure 3]: Set hostnme to appropriate name server address.

Appendix B : Ambari UI Setup

1) Confirm hosts

- User needs to configure how many masters and slaves are being required according to requirements.

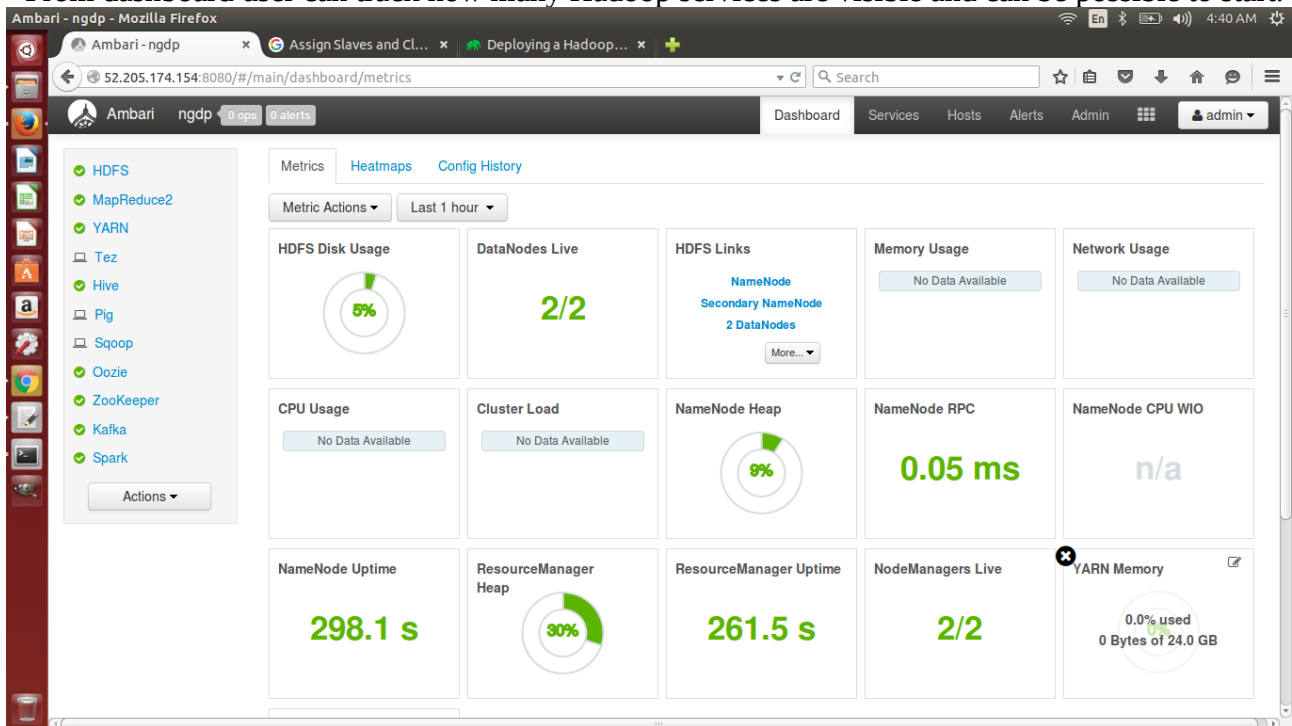
- **Note:** make sure everything setup has been done as mentioned in **Appendix-A**



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

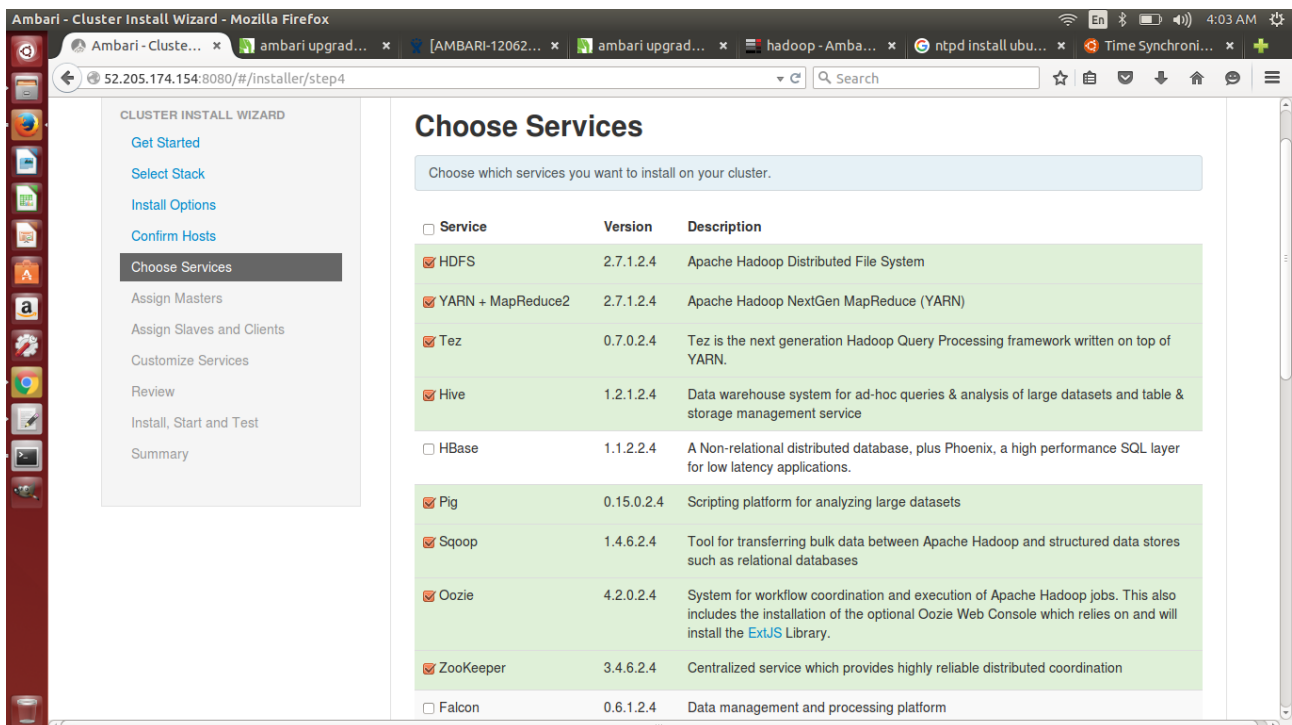
2) Ambari UI Dashboard

- 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.

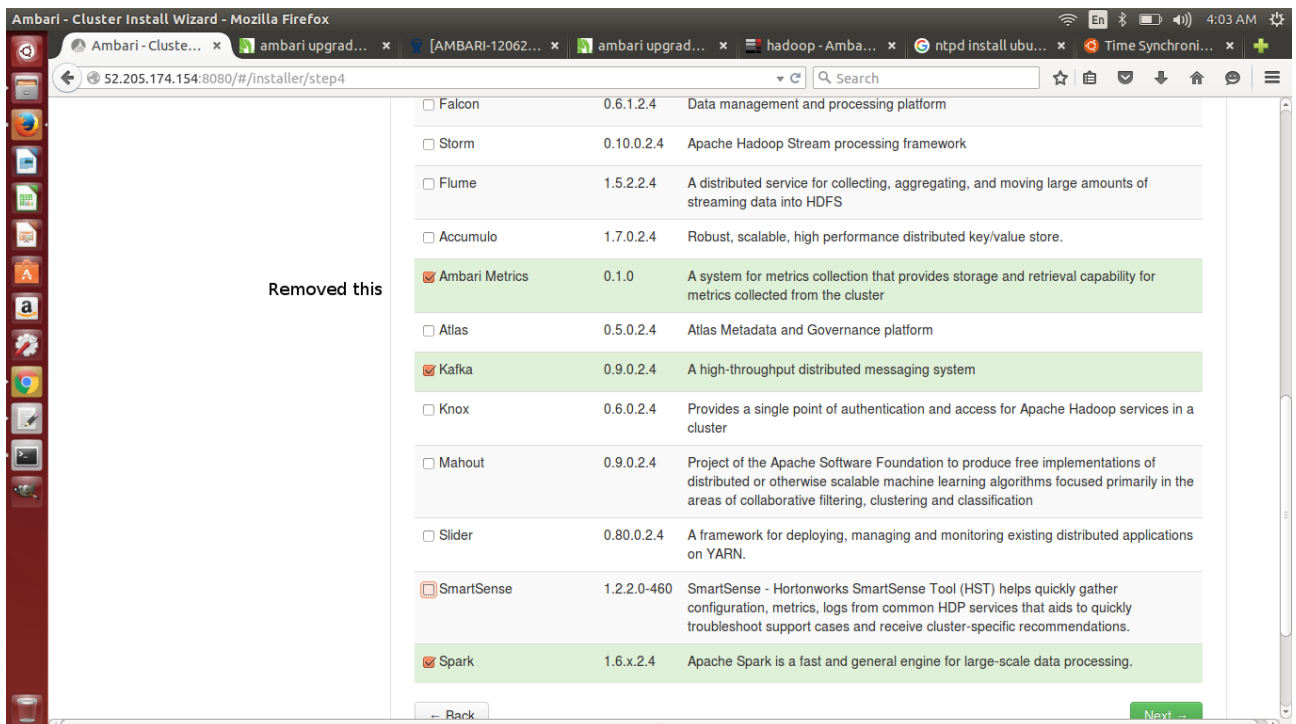
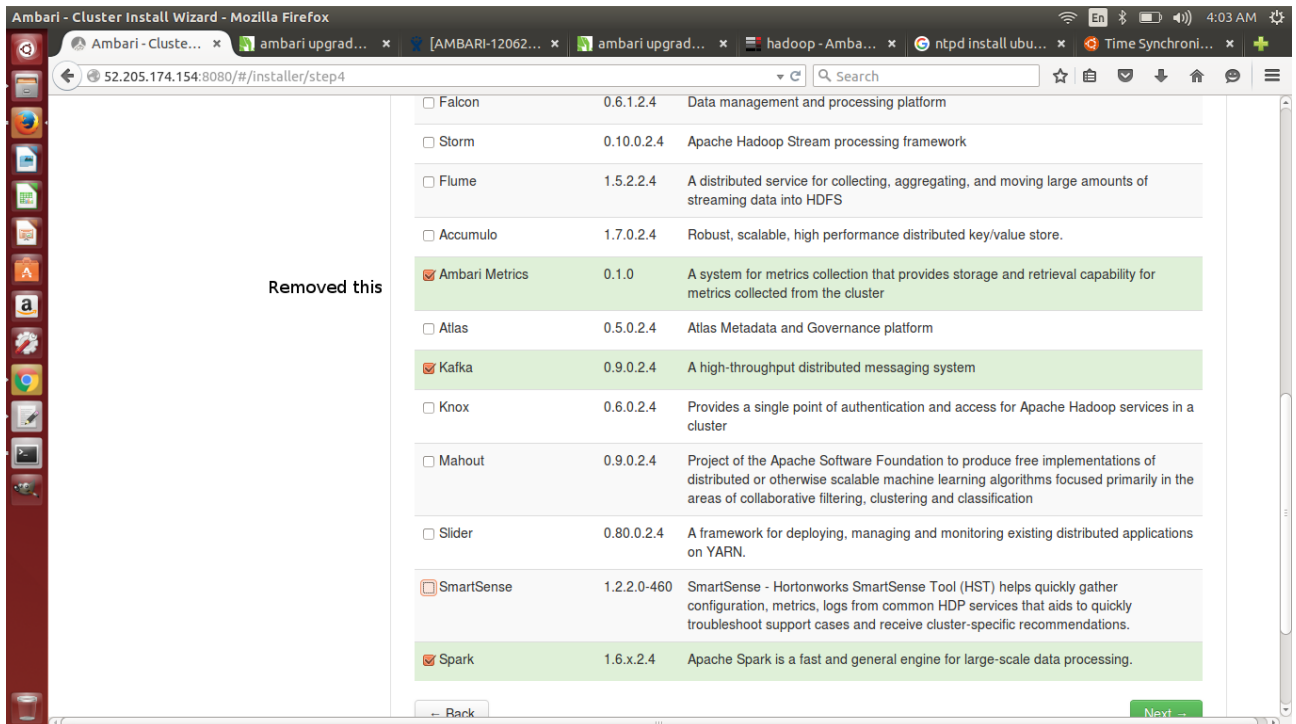


[Figure 4]: Ambari Dashboard

3) Selected Hadoop Ecosystem services

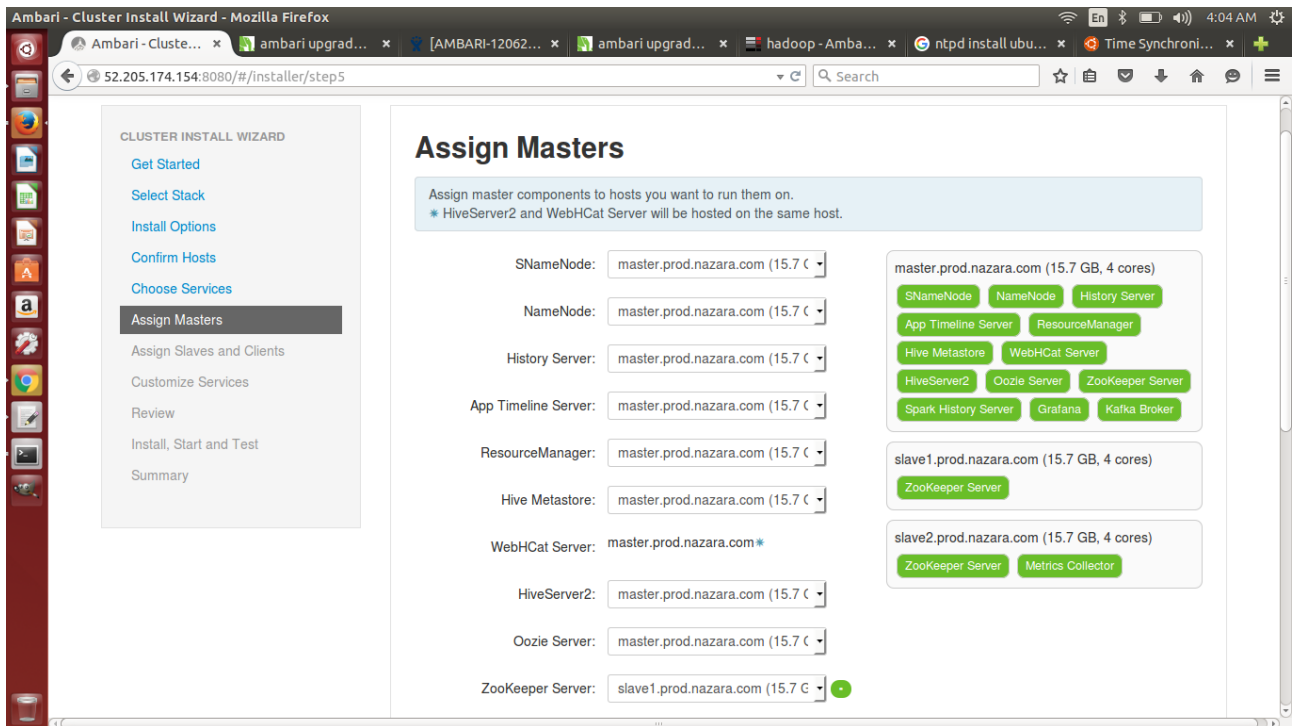


[Figure 5]: Choose Hadoop related services with Ambari



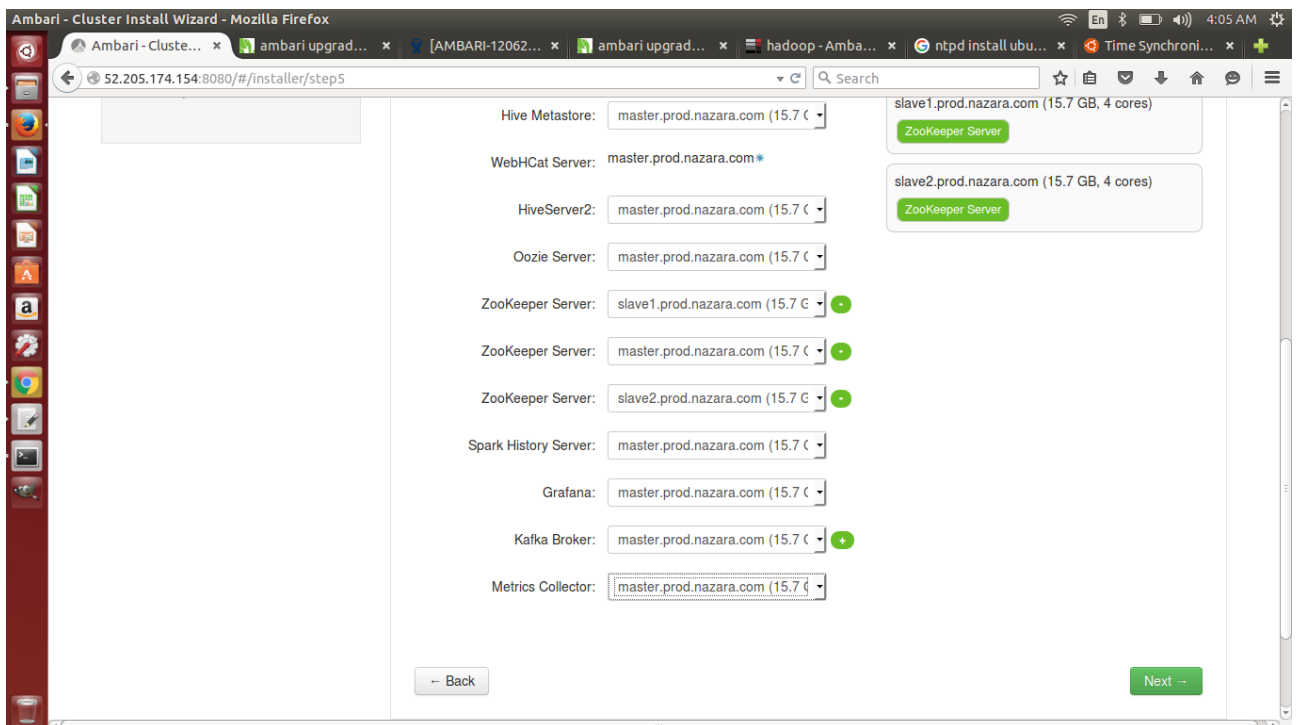
[Figure 6]: Choose Hadoop related services with Ambari

4) Assign Hadoop Masters



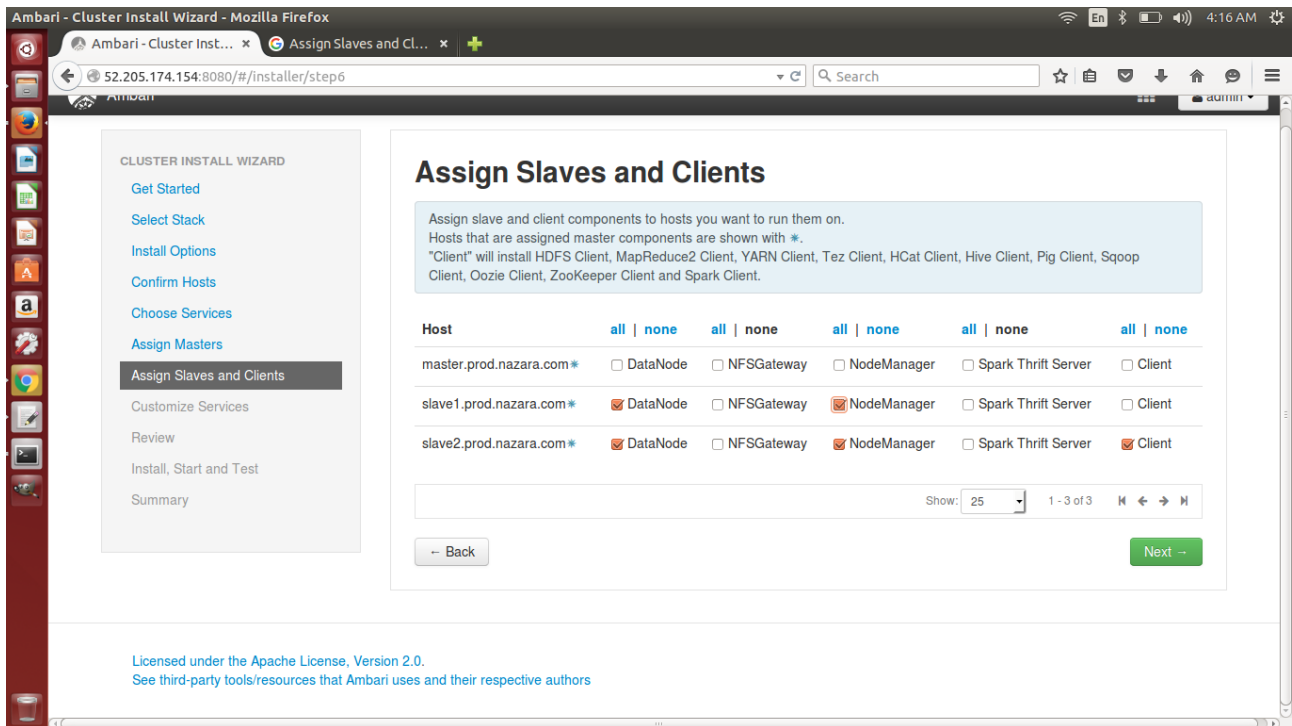
[Figure 7]: Assign Master / Slave node in Hadoop related services

5) Assign Slaves and Client



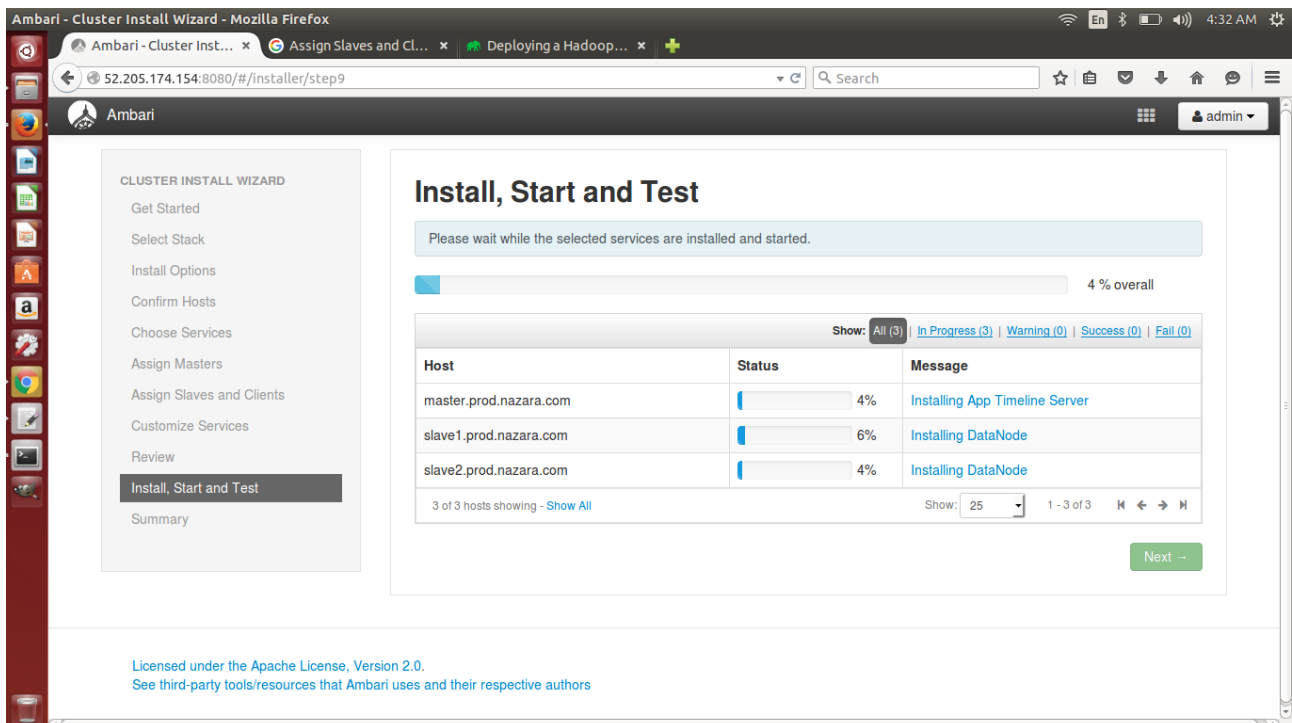
[Figure 8]: Assign Master / Slave node in Hadoop related services

6) Install Master and slaves



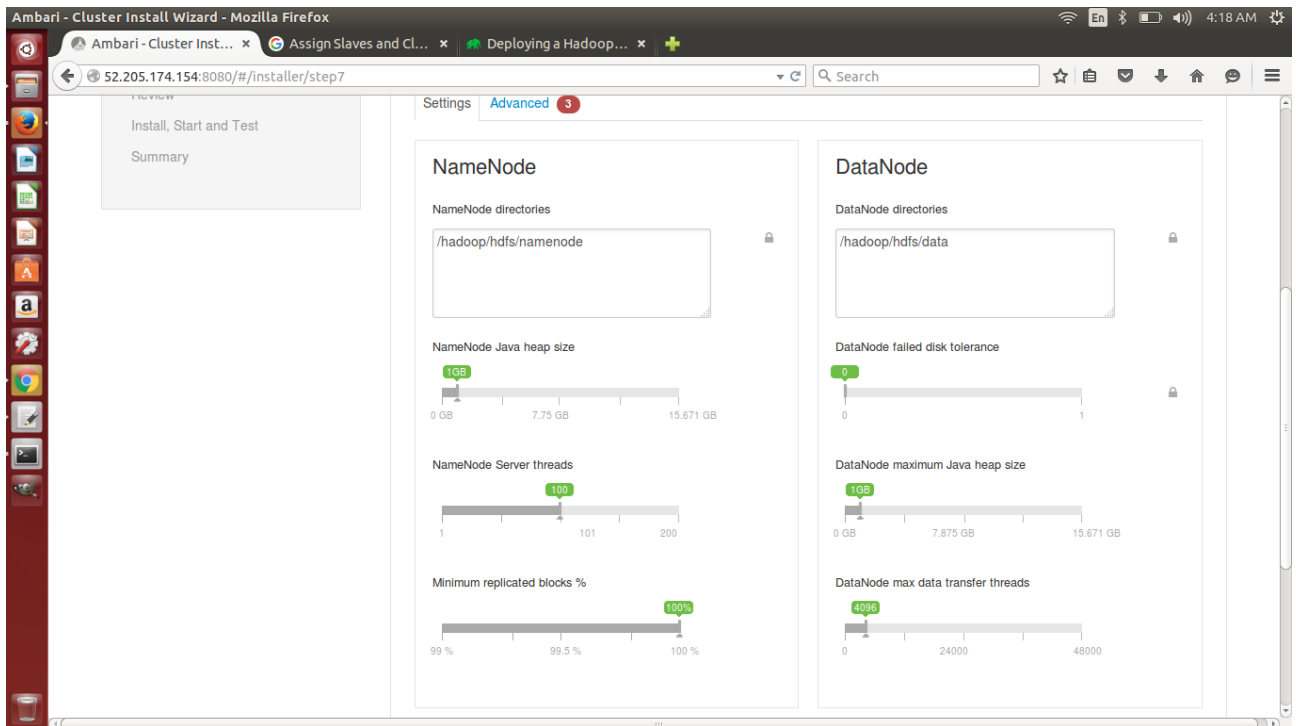
[Figure 9]: Assign slaves and clients

6) Install , Start and Test Masters and Slaves



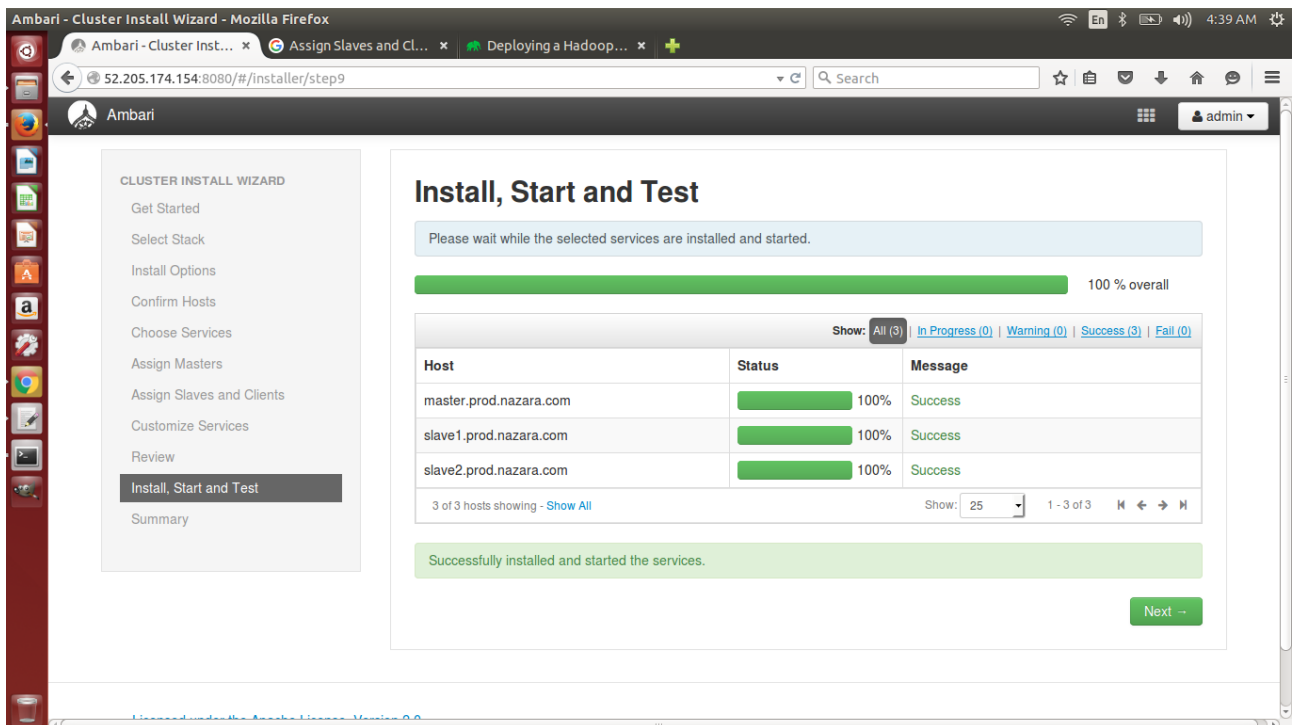
[Figure 10]: Install, Start and Test Masters and Slaves.

7) Namenode and Datanode configuration on Ambari



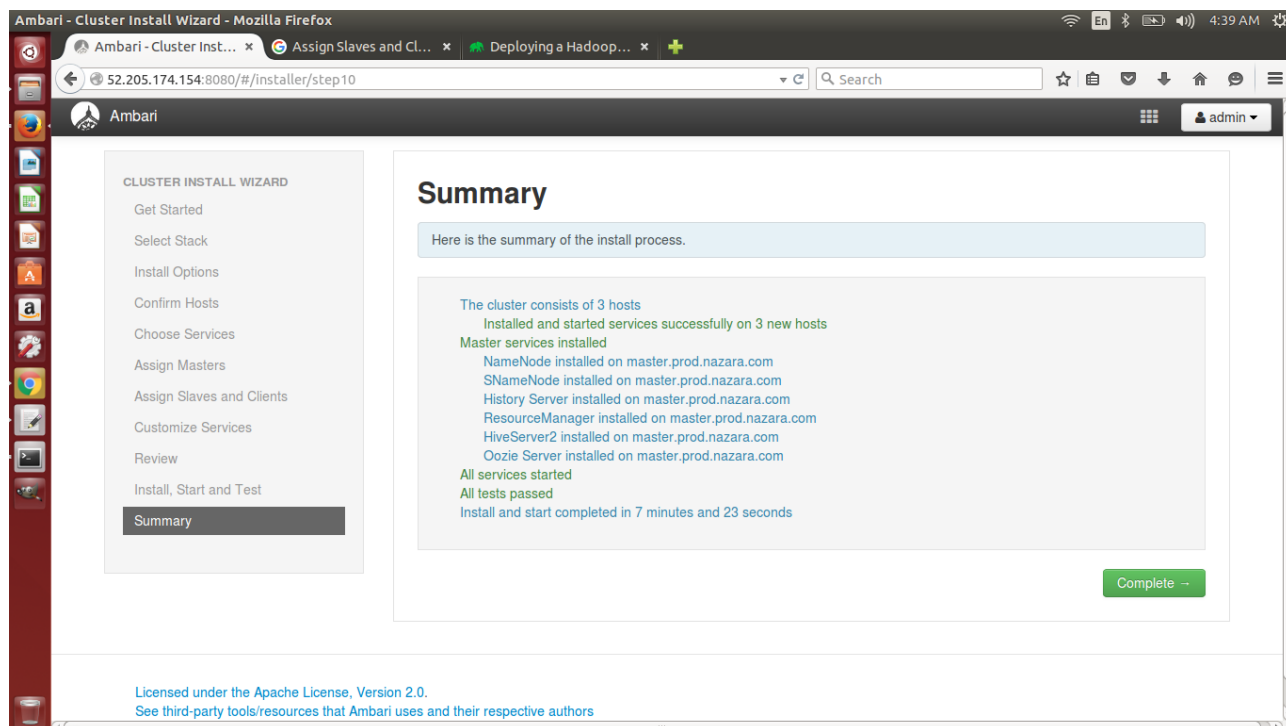
[Figure 11]: Namenode and Datanode Advanced configuration on Ambari.

8) Testing Master and Slave



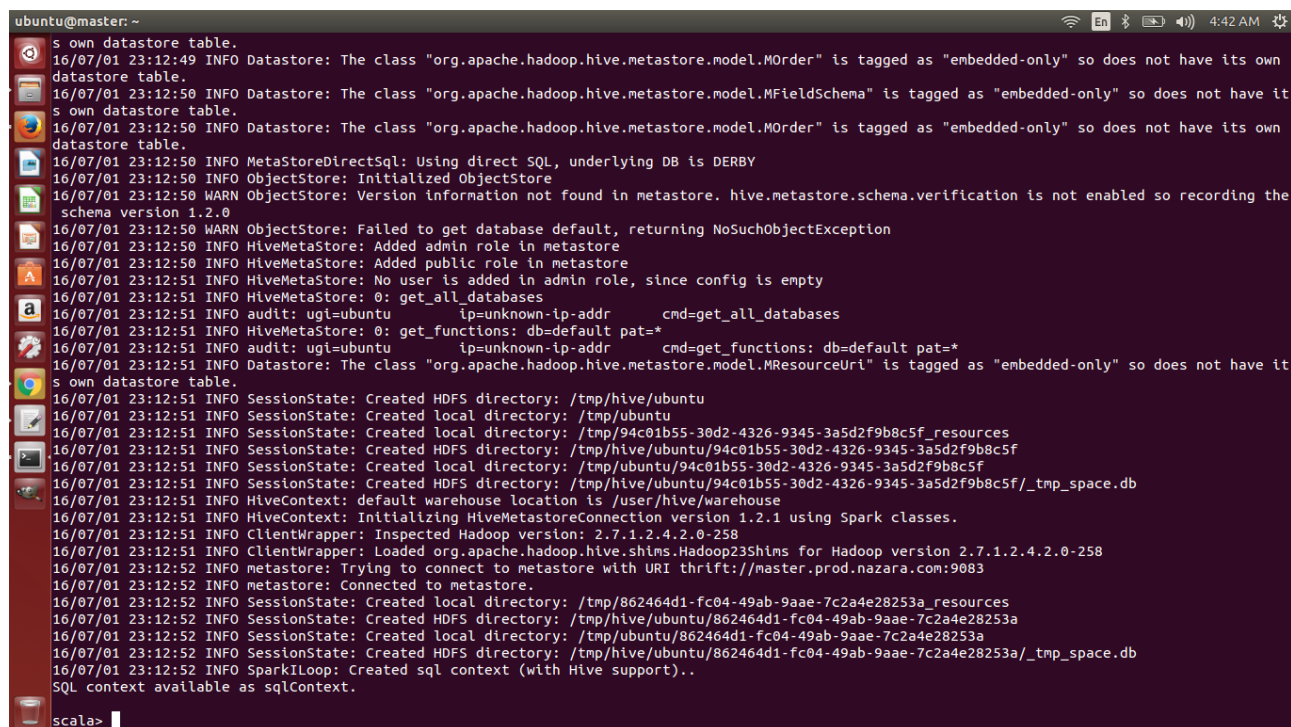
[Figure 12]: Install Start and Test Hadoop masters and Slaves.

9) Installation Summary Process



[Figure 13]: Ambari Installation Summary

10) starting spark-shell from Ambari



[Figure 14]: starting spark-shell from 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