Lab 2) Perform setting up and Installing Hadoop in its three operating modes as follow:

1. **Standalone**
2. **Pseudo Distributed**
3. **Fully Distributed**

Hadoop is written in Java, so you will need to have Java installed on your machine, version 6 or later. Sun's JDK is the one most widely used with Hadoop, although others have been reported to work.

Hadoop runs on Unix and on Windows. Linux is the only supported production platform, but other flavors of Unix (including Mac OS X) can be used to run Hadoop for development. Windows is only supported as a development platform, and additionally requires Cygwin to run. During the Cygwin installation process, you should include the openssh package if you plan to run Hadoop in pseudo-distributed mode

ALGORITHM

STEPS INVOLVED IN INSTALLING HADOOP IN STANDALONE MODE:-

1. Command for installing ssh is **"sudo apt-get install ssh"**.
2. Command for key generation is **ssh-keygen –t rsa –P " ".**
3. Store the key into rsa.pub by using the command **cat $HOME/.ssh/id_rsa.pub >>**

   $HOME/.ssh/authorized_keys
4. Extract the java by using the command **tar xvfz jdk-8u60-linux-i586.tar.gz.**
5. Extract the eclipse by using the command **tar xvfz eclipse-jee-mars-R-linux-gtk.tar.gz**
6. Extract the hadoop by using the command **tar xvfz hadoop-2.7.1.tar.gz**

7. Move the java to **/usr/lib/jvm/** and eclipse to **/opt/** paths. Configure the java path in the eclipse.ini file
8. Export java path and hadoop path in ./bashrc
9. Check the installation successful or not by checking the java version and hadoop version
10. Check the hadoop instance in standalone mode working correctly or not by using an implicit hadoop jar file named as word count.
11. If the word count is displayed correctly in part-r-00000 file it means that standalone mode is installed successfully.

ALGORITHM

STEPS INVOLVED IN INSTALLING HADOOP IN PSEUDO DISTRIBUTED MODE:-

1. In order install pseudo distributed mode we need to configure the hadoop configuration files resides in the directory /home/lendi/hadoop-2.7.1/etc/hadoop.
2. First configure the hadoop-env.sh file by changing the java path.
3. Configure the core-site.xml which contains a property tag, it contains name and value. Name as fs.defaultFS and value as hdfs://localhost:9000
4. Configure hdfs-site.xml.
5. Configure yarn-site.xml.
6. Configure mapred-site.xml before configure the copy mapred-site.xml.template to mapred-site.xml.
7. Now format the name node by using command hdfs namenode –format.
8. Type the command start-dfs.sh,start-yarn.sh means that starts the daemons like NameNode,DataNode,SecondaryNameNode ,ResourceManager,NodeManager.

9. Run JPS which views all daemons. Create a directory in the hadoop by using command hdfs dfs –mkdr /csedir and enter some data into lendi.txt using command nano lendi.txt and copy from local directory to hadoop using command hdfs dfs –copyFromLocal lendi.txt /csedir/and run sample jar file wordcount to check whetherpseudo distributed mode is working or not.

10. Display the contents of file by using command hdfs dfs –cat /newdir/part-r-0000

FULLY DISTRIBUTED MODE INSTALLATION: ALGORITHM

1. Stop all single node clusters

   $stop-all.sh

2. Decide one as NameNode (Master) and remaining as DataNodes(Slaves).

3. Copy public key to all three hosts to get a password less SSH access

   $ssh-copy-id –I $HOME/.ssh/id_rsa.pub lendi@l5sys24

4. Configure all Configuration files, to name Master and Slave Nodes.

   $cd $HADOOP_HOME/etc/hadoop

   $nano core-site.xml

   $ nano hdfs-site.xml

5. Add hostnames to file slaves and save it.

   $ nano slaves

6. Configure $ nano yarn-site.xml

7. Do in Master Node

   $ hdfs namenode –format

   $ start-dfs.sh

   $start-yarn.sh

8. Format NameNode

9. Daemons Starting in Master and Slave Nodes

10. END

INPUT

ubuntu @localhost> jps

OUTPUT:

Data node, name nodem Secondary name node, NodeManager, Resource Manager

II) Using Web Based Tools to Manage Hadoop Set-upDESCRIPTION

Hadoop set up can be managed by different web-based tools, which can be easy for theuser to identify the running daemons. Few of the tools used in the real world are:

a) Apache Ambari

b) Horton Works

c) Apache Spark

LIST OF CLUSTERS IN HADOOP

Apache Hadoop Running at Local Host

Hadoop NameNode localhost:9000 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Hadoop NameNode localhost:9...

localhost:50070/dfshealth.jsp

NameNode 'localhost:9000' (active)

Started:	Fri Dec 13 13:12:58 EST 2013
Version:	2.2.0, 1529768
Compiled:	2013-10-07T06:28Z by hortonmu from branch-2.2.0
Cluster ID:	CID-a29c04fc-ee6f-411d-a3a9-32908b0091a5
Block Pool ID:	BP-759972770-10.0.0.1-1386957724751

Browse the filesystem
NameNode Logs

Cluster Summary

Security is *OFF*
1 files and directories, 0 blocks = 1 total.
Heap Memory used 65.92 MB is 27% of Commited Heap Memory 237.31 MB. Max Heap Memory is 444.50 MB.
Non Heap Memory used 29.35 MB is 95% of Commited Non Heap Memory 30.69 MB. Max Non Heap Memory is 214 MB.

Configured Capacity	:	49.22 GB
DFS Used	:	24 KB
Non DFS Used	:	9.48 GB
DFS Remaining	:	39.73 GB
DFS Used%	:	0.00%
DFS Remaining%	:	80.73%
Block Pool Used	:	24 KB

Browsing HDFS - Mozilla Firefox

Browsing HDFS

localhost:50070/explorer.html#/

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

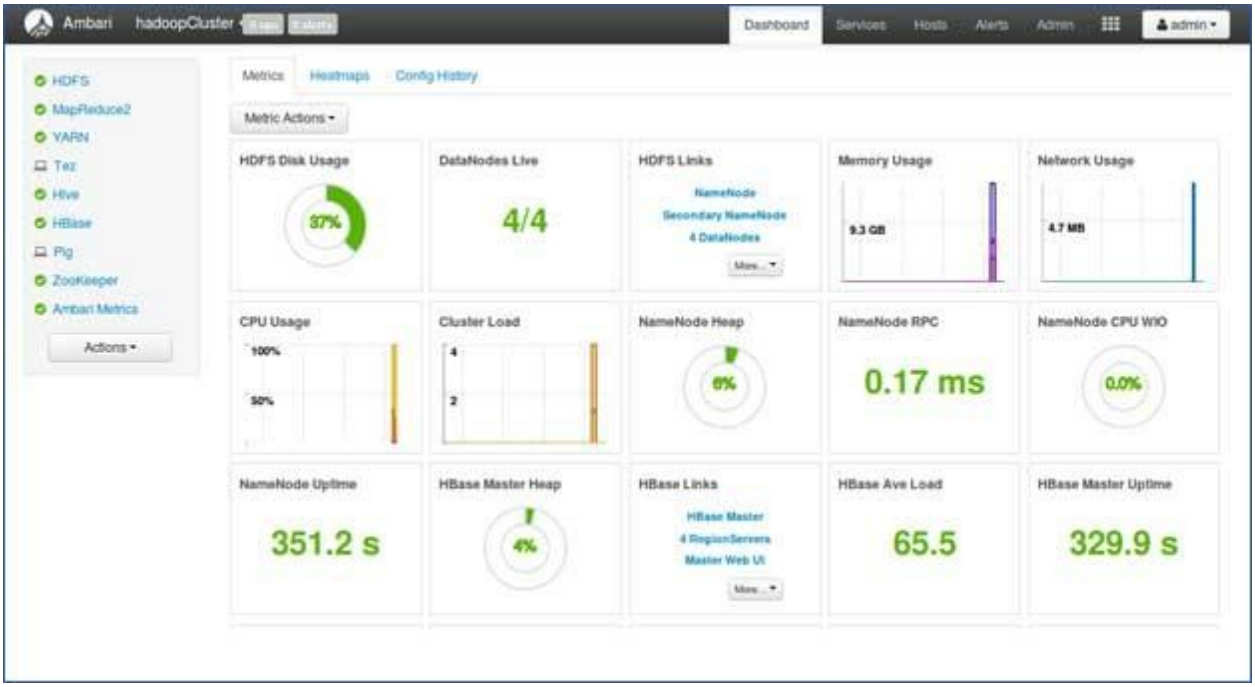
Browse Directory

/ Go!

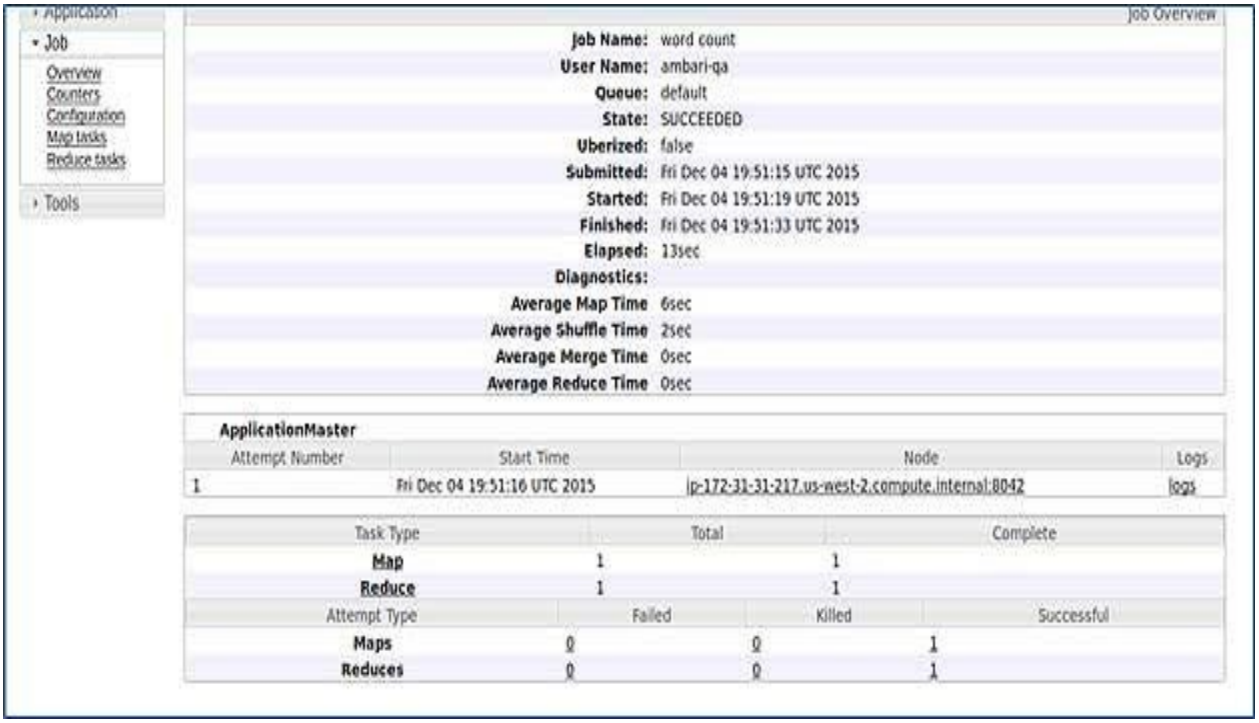
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	lendi	supergroup	0 B	Wed 17 Aug 2016 02:44:00 AM EDT	0	0 B	lendi_english
drwxr-xr-x	lendi	supergroup	0 B	Wed 17 Aug 2016 02:17:48 AM EDT	0	0 B	sadhana
drwxr-xr-x	lendi	supergroup	0 B	Sat 13 Aug 2016 01:31:42 AM EDT	0	0 B	shakes
drwxr-xr-x	lendi	supergroup	0 B	Sat 13 Aug 2016 01:35:59 AM EDT	0	0 B	shakes1
drwx------	lendi	supergroup	0 B	Sat 13 Aug 2016 01:19:03 AM EDT	0	0 B	tmp
drwxr-xr-x	lendi	supergroup	0 B	Sat 13 Aug 2016 12:09:18 AM EDT	0	0 B	ukcp
drwxr-xr-x	lendi	supergroup	0 B	Sat 13 Aug 2016 01:25:55 AM EDT	0	0 B	ukcp1
drwxr-xr-x	lendi	supergroup	0 B	Sat 13 Aug 2016 01:28:06 AM EDT	0	0 B	ukcp2
drwxr-xr-x	lendi	supergroup	0 B	Sat 13 Aug 2016 01:19:20 AM EDT	0	0 B	user

Hadoop, 2015.

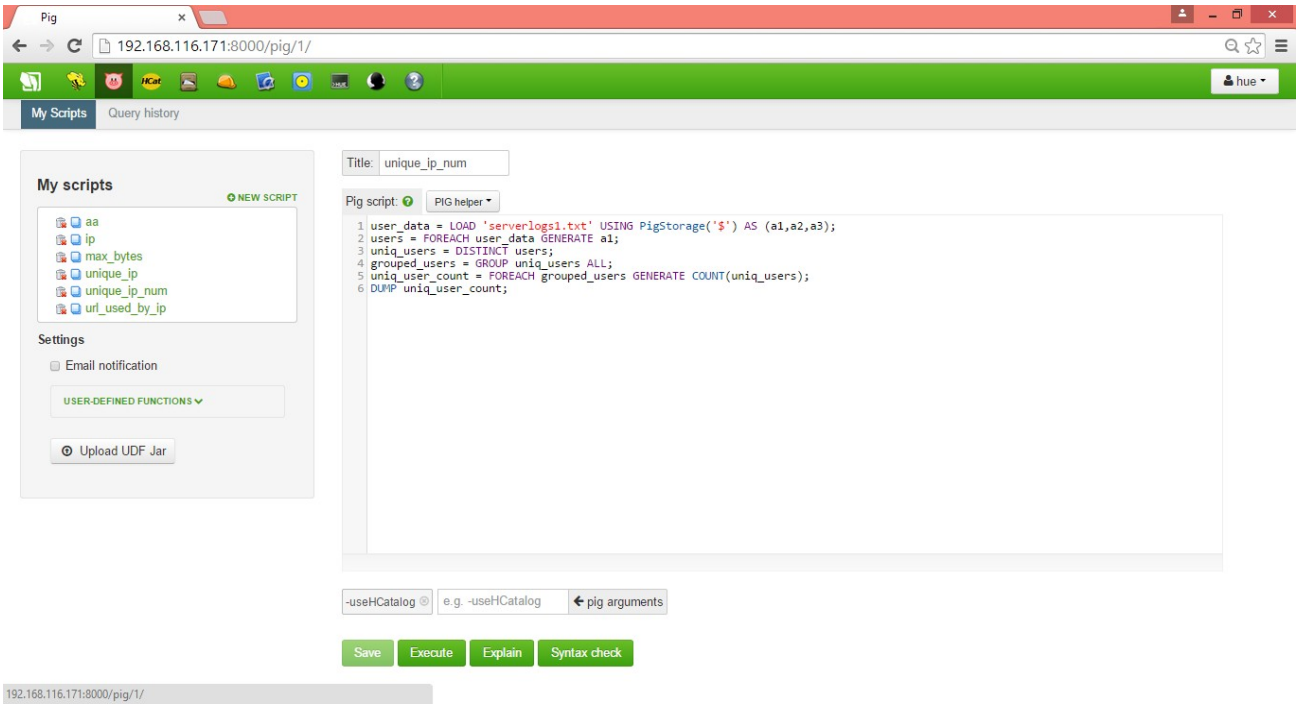
AMBARI Admin Page for Managing Hadoop Clusters



AMBARI Admin Page for Viewing Hadoop Map Reduce Jobs



Horton Works Tool for Managing Map Reduce Jobs in Apache Pig



Running Map Reduce Jobs in Horton Works for Pig Latin Scrip