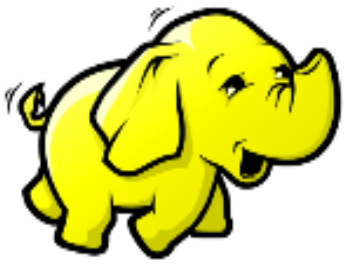
Wisdom Sprouts

We foster knowledge here

BIG DATA HADOOP

Lab Manual



Yarn Installation:

Step 1: Stop all services

```
for service in /etc/init.d/hadoop*
> do
> sudo $service stop
> done
```

Expected Output

```
Stopping Hadoop jobtracker:                                [  OK  ]
stopping jobtracker
Stopping Hadoop tasktracker:                               [  OK  ]
stopping tasktracker
Stopping Hadoop datanode:                                  [  OK  ]
stopping datanode
Stopping Hadoop namenode:                                  [  OK  ]
stopping namenode
Stopping Hadoop secondarynamenode:                         [  OK  ]
stopping secondarynamenode
```

Step 2:Remove hadoop-0.20-conf-pseudosudo

```
sudo yum remove hadoop-0.20-conf-pseudo hadoop-0.20-mapreduce-*
```

Expected Output

```
Lot of Dependencies will be resolved then the following will be
displayed.

Transaction Summary
================================================================================
============
Remove       10 Package(s)

Installed size: 341 M
Is this ok [y/N]: y
```

After confirming a yes then lot of dependencies will be resolved then the following will be displayed.

```
Removed:
  hadoop-0.20-conf-pseudo.i686 0:2.0.0+1475-1.cdh4.4.0.p0.23.el6
  hadoop-0.20-mapreduce.i686 0:2.0.0+1475-1.cdh4.4.0.p0.23.el6
  hadoop-0.20-mapreduce-jobtracker.i686 0:2.0.0+1475-
1.cdh4.4.0.p0.23.el6
  hadoop-0.20-mapreduce-tasktracker.i686 0:2.0.0+1475-
1.cdh4.4.0.p0.23.el6

Dependency Removed:
  hadoop-client.i686 0:2.0.0+1475-1.cdh4.4.0.p0.23.el6
  hive.noarch 0:0.10.0+198-1.cdh4.4.0.p0.15.el6
  hive-hbase.noarch 0:0.10.0+198-1.cdh4.4.0.p0.15.el6
  hive-jdbc.noarch 0:0.10.0+198-1.cdh4.4.0.p0.15.el6
  pig.noarch 0:0.11.0+42-1.cdh4.6.0.p0.13.el6
  sqoop.noarch 0:1.4.3+62-1.cdh4.4.0.p0.15.el6

Complete!
```

Step 3: Download the CDH4 Package

```
sudo wget http://archive.cloudera.com/cdh4/one-click-
install/redhat/6/i386/cloudera-cdh-4-0.i386.rpm
```

Expected Output

```
--2016-02-24 01:56:25--  http://archive.cloudera.com/cdh4/one-click-
install/redhat/6/i386/cloudera-cdh-4-0.i386.rpm
Resolving archive.cloudera.com... 43.249.75.167
Connecting to archive.cloudera.com|43.249.75.167|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 9136 (8.9K) [application/x-redhat-package-manager]
Saving to: “cloudera-cdh-4-0.i386.rpm.1”

100%[======================================>] 9,136       --.-K/s
in 0.01s

2016-02-24 01:56:25 (853 KB/s) - “cloudera-cdh-4-0.i386.rpm.1” saved
[9136/9136]
```

Step 4: Install the RPM

```
sudo yum --nogpgcheck localinstall cloudera-cdh-4-0.i386.rpm
```

Expected Output

```
Loaded plugins: fastestmirror, refresh-packagekit, security
Setting up Local Package Process
Examining cloudera-cdh-4-0.i386.rpm: cloudera-cdh-4-0.i386
cloudera-cdh-4-0.i386.rpm: does not update installed package.
Nothing to do
```

Step 5: To install Hadoop with YARN

```
sudo yum install hadoop-conf-pseudo
```

Expected Output

```
Lot of packages will be downloaded.
Then confirmation message will be asked for.
Finally you will see this message
Installed:
  hadoop-conf-pseudo.i686 0:2.0.0+1612-1.cdh4.7.1.p0.12.el6

Dependency Installed:
  hadoop-mapreduce-historyserver.i686 0:2.0.0+1612-
1.cdh4.7.1.p0.12.el6
  hadoop-yarn-nodemanager.i686 0:2.0.0+1612-1.cdh4.7.1.p0.12.el6
  hadoop-yarn-resourcemanager.i686 0:2.0.0+1612-1.cdh4.7.1.p0.12.el6

Dependency Updated:
  hadoop.i686 0:2.0.0+1612-1.cdh4.7.1.p0.12.el6
  hadoop-hdfs.i686 0:2.0.0+1612-1.cdh4.7.1.p0.12.el6
  hadoop-hdfs-datanode.i686 0:2.0.0+1612-1.cdh4.7.1.p0.12.el6
  hadoop-hdfs-namenode.i686 0:2.0.0+1612-1.cdh4.7.1.p0.12.el6
  hadoop-hdfs-secondarynamenode.i686 0:2.0.0+1612-
1.cdh4.7.1.p0.12.el6
  hadoop-mapreduce.i686 0:2.0.0+1612-1.cdh4.7.1.p0.12.el6
  hadoop-yarn.i686 0:2.0.0+1612-1.cdh4.7.1.p0.12.el6

Complete!
```

Step 6: To view the files

```
rpm -ql hadoop-conf-pseudo
```

Expected Output

```
/etc/hadoop/conf.pseudo
/etc/hadoop/conf.pseudo/README
/etc/hadoop/conf.pseudo/core-site.xml
/etc/hadoop/conf.pseudo/hadoop-env.sh
/etc/hadoop/conf.pseudo/hadoop-metrics.properties
/etc/hadoop/conf.pseudo/hdfs-site.xml
/etc/hadoop/conf.pseudo/log4j.properties
/etc/hadoop/conf.pseudo/mapred-site.xml
/etc/hadoop/conf.pseudo/yarn-site.xml
```

Step 7: Format the NameNode

```
sudo -u hdfs hdfs namenode -format
```

Expected Output

```
After lot of stuff you will see the following message
16/02/24 02:05:27 INFO namenode.NNStorage: Storage directory
/var/lib/hadoop-hdfs/cache/hdfs/dfs/name has been successfully
formatted.
16/02/24 02:05:27 INFO namenode.FSImage: Saving image file
/var/lib/hadoop-
hdfs/cache/hdfs/dfs/name/current/fsimage.ckpt_0000000000000000000
using no compression
16/02/24 02:05:27 INFO namenode.FSImage: Image file of size 119
saved in 0 seconds.
16/02/24 02:05:27 INFO namenode.NNStorageRetentionManager: Going to
retain 1 images with txid >= 0
16/02/24 02:05:27 INFO util.ExitUtil: Exiting with status 0
16/02/24 02:05:27 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at
localhost.localdomain/127.0.0.1
************************************************************/
```

Step 8: Start HDFS Services

```
for service in /etc/init.d/hadoop-hdfs-*
```

Expected Output

```
> do
> sudo $service start
> done
Output:
Starting Hadoop datanode:                                  [  OK  ]
starting datanode, logging to /var/log/hadoop-hdfs/hadoop-hdfs-
datanode-localhost.localdomain.out
Starting Hadoop namenode:                                  [  OK  ]
starting namenode, logging to /var/log/hadoop-hdfs/hadoop-hdfs-
namenode-localhost.localdomain.out
Starting Hadoop secondarynamenode:                         [  OK  ]
starting secondarynamenode, logging to /var/log/hadoop-hdfs/hadoop-
hdfs-secondarynamenode-localhost.localdomain.out
```

Step 9: Check if HDFS Services started

```
sudo /usr/java/latest/bin/jps
```

Expected Output

```
4097 Jps
3870 SecondaryNameNode
3765 NameNode
2340 HMaster
3954 DataNode
```

Step 10: (If Data Node doesn't start)

```
sudo rm -r /var/lib/hadoop-hdfs/cache/hdfs/dfs/data/
```

Expected Output

Step 11: Remove the old /tmp if it exists

```
sudo -u hdfs hadoop fs -rm -r /tmp
```

Expected Output

Step 12: Create a new /tmp directory and set permissions

```
sudo -u hdfs hadoop fs -mkdir /tmp
sudo -u hdfs hadoop fs -chmod -R 1777 /tmp
```

Expected Output

Step 13: Create the staging directory and set permissions

```
sudo -u hdfs hadoop fs -mkdir /tmp/hadoop-yarn/staging
sudo -u hdfs hadoop fs -chmod -R 1777 /tmp/hadoop-yarn/staging
```

Expected Output

Step 14: Create the done_intermediate directory under the staging directory and set permissions

```
sudo -u hdfs hadoop fs -mkdir /tmp/hadoop-
yarn/staging/history/done_intermediate

sudo -u hdfs hadoop fs -chmod -R 1777 /tmp/hadoop-
yarn/staging/history/done_intermediate
```

Expected Output

Step 15: Change ownership on the staging directory and subdirectory

```
sudo -u hdfs hadoop fs -chown -R mapred:mapred /tmp/hadoop-
yarn/staging
```

Expected Output

Step 16: Create the /var/log/hadoop-yarn directory and set ownership

```
sudo -u hdfs hadoop fs -mkdir /var/log/hadoop-yarn
sudo -u hdfs hadoop fs -chown yarn:mapred /var/log/hadoop-yarn
```

Expected Output

Step 17: Verify the HDFS File Structure

```
sudo -u hdfs hadoop fs -ls -R /
```

Expected Output

```
drwxrwxrwt   - hdfs supergroup          0 2016-02-24 03:03 /tmp
drwxr-xr-x   - hdfs supergroup          0 2016-02-24 03:03
/tmp/hadoop-yarn
drwxrwxrwt   - mapred mapred              0 2016-02-24 03:04
/tmp/hadoop-yarn/staging
drwxr-xr-x   - mapred mapred              0 2016-02-24 03:04
/tmp/hadoop-yarn/staging/history
drwxrwxrwt   - mapred mapred              0 2016-02-24 03:04
/tmp/hadoop-yarn/staging/history/done_intermediate
drwxr-xr-x   - hdfs   supergroup          0 2016-02-24 03:05 /var
drwxr-xr-x   - hdfs   supergroup          0 2016-02-24 03:05
/var/log
drwxr-xr-x   - yarn   mapred              0 2016-02-24 03:05
/var/log/hadoop-yarn
```

Step 18.1: Start YARN

```
sudo service hadoop-yarn-resourcemanager start
```

Expected Output

```
Starting Hadoop resourcemanager:                           [  OK  ]
starting resourcemanager, logging to /var/log/hadoop-yarn/yarn-yarn-
resourcemanager-localhost.localdomain.out
```

Step 18.2:

```
sudo service hadoop-yarn-nodemanager start
```

Expected Output

```
Starting Hadoop nodemanager:                               [  OK  ]
starting nodemanager, logging to /var/log/hadoop-yarn/yarn-yarn-
nodemanager-localhost.localdomain.out
```

Step 18.3:

```
sudo service hadoop-mapreduce-historyserver start
```

Expected Output

```
Starting Hadoop historyserver:                             [  OK  ]
starting historyserver, logging to /var/log/hadoop-mapreduce/yarn-
mapred-historyserver-localhost.localdomain.out
```

Step 19:Check if services have started

```
sudo /usr/java/latest/bin/jps
```

Expected Output

```
3870 SecondaryNameNode
4488 ResourceManager
3765 NameNode
2340 HMaster
4964 Jps
4777 NodeManager
3954 DataNode
4904 JobHistoryServer
```

Step 20:Create User Directories

```
sudo -u hdfs hadoop fs -mkdir /user/wisdom
sudo -u hdfs hadoop fs -chown wisdom /user/wisdom
```

Expected Output

Step 21:Create a sample input

```
hadoop fs -mkdir input
hadoop fs -put /etc/hadoop/conf/*.xml input
hadoop fs -ls input
```

Expected Output

```
Found 4 items
```

```
-rw-r--r--   1 wisdom supergroup       1458 2016-02-24 03:12
input/core-site.xml
-rw-r--r--   1 wisdom supergroup       1875 2016-02-24 03:12
input/hdfs-site.xml
-rw-r--r--   1 wisdom supergroup       1549 2016-02-24 03:12
input/mapred-site.xml
-rw-r--r--   1 wisdom supergroup       2361 2016-02-24 03:12
input/yarn-site.xml
```

Step 22:Set HADOOP_MAPRED_HOME for user wisdom

```
export HADOOP_MAPRED_HOME=/usr/lib/hadoop-mapreduce
```

Expected Output

Step 23:Run an example Hadoop job

```
hadoop jar /usr/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar
grep input output23 'dfs[a-z.]+'
```

Expected Output

```
After lot of stuff you will finally see something like this
File Input Format Counters
              Bytes Read=320
       File Output Format Counters
              Bytes Written=150
```

Step 24:Check Output

```
hadoop fs -cat output23/part-r-00000
```

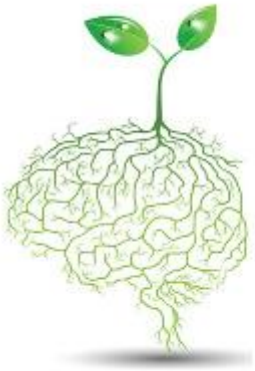
Expected Output

```
1      dfs.safemode.min.datanodes
1      dfs.safemode.extension
1      dfs.replication
1      dfs.namenode.name.dir
1      dfs.namenode.checkpoint.dir
1      dfs.datanode.data.dir
```

Step 25:Check Browser

http://localhost:8088

Expected Output



WISDOM SPROUTS

We Foster Knowledge Here

Contact Us

You can call us on +91-976-220-6123

Or write to us at helpdesk@wisdomsprouts.com

Or Visit www.wisdomsprouts.com