

1. Recommended Platform

OS : ubuntu 12.04

Hadoop: Apache Hadoop(hadoop-2.2.0 or above)

2. Install java

2.1) Update the source list

command:-sudo apt-get update

2.2) install open jdk

command:-sudo apt-get install openjdk-6-jdk

3. Configure SSH (shown below in screenshot):-

3.1) install open SSH Server-Client

command:-sudo apt-get install openssh-server openssh-client

3.2) Generate key value pair

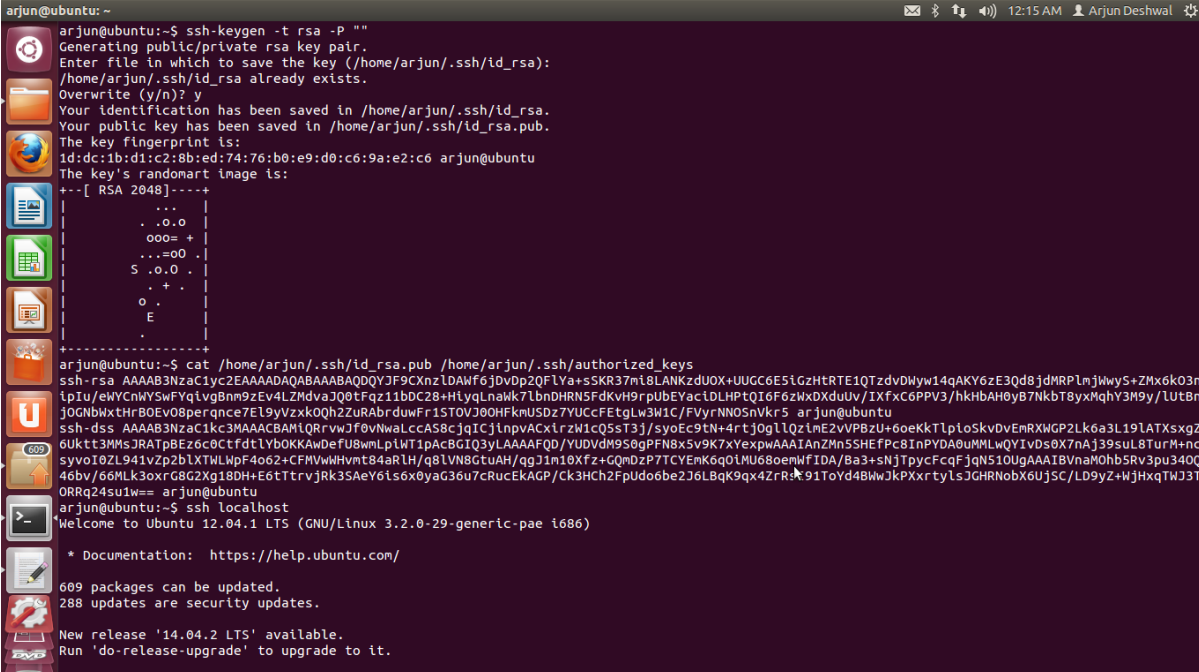
command:-ssh-keygen -t rsa -P ""

3.3)configure password-less SSH

command:-cat \$HOME/.ssh/id_rsa.pub >> \$HOME/.ssh/authorized_keys

3.4)check by SSH to localhost

command :-ssh localhost



```
arjun@ubuntu:~$ ssh-keygen -t rsa -P ""
Generating public/private rsa key pair.
Enter file in which to save the key (/home/arjun/.ssh/id_rsa):
/home/arjun/.ssh/id_rsa already exists.
Overwrite (y/n)? y
Your identification has been saved in /home/arjun/.ssh/id_rsa.
Your public key has been saved in /home/arjun/.ssh/id_rsa.pub.
The key fingerprint is:
id:dc:1b:d1:c2:8b:ed:74:76:b0:e9:d0:c6:9a:e2:c6 arjun@ubuntu
The key's randomart image is:
+--[ RSA 2048 ]-----+
|          .o.o          |
|         .ooo=+        |
|        ...=oo-       |
|       S .o.o .       |
|      .+ .            |
|     .+ .            |
|    .+ .            |
|   .+ .            |
|  .+ .            |
| .+ .            |
|+-----+
arjun@ubuntu:~$ cat /home/arjun/.ssh/id_rsa.pub >> /home/arjun/.ssh/authorized_keys
ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQDQYJF9Cknz1DAWf6jDvDp2QFLYa+s5KR37mi8LANKzdUOX+UUGC6E5iGzHrTE1QTzdvdWYw14qAKY6zE3Qd8jdMRPLnjWwyS+ZMx6k03n
ipIu/eMYCnWYswFYqIvgBnm9zEv4LZMdvajQ0tFqz11bDC28+HilyqLnawK7lbnDHRN5FdKvH9rpubEYac1DLHPTQI6FzWxDXduUv/IXfxc6PPV3/hkHbAH0yB7NkbT8yxMqhY3M9y/LUTBn
j0GNbwxthRBOEv08perqnce7E19yVzx0Qh2ZuRABrduwFr1STOVJ90HFkmUSDz7YUCCFetGLw3W1C/FVyrNNOsnVkr5 arjun@ubuntu
ssh-dss AAAAB3NzaC1kc3MAAACBAmlQRrvwJf0vNwaLccA58cjICjInpvACxlrzH1CQ5sT3j/syoEc9tn+4rtj0gl1QziE2vVPBzU+6oeKkTlp0SkvDvEnRXWGP2Lk6a3L19LATxsxgZ
6Uktt3MMsJRATpBEz6c0CtfdtlyBOKKAwDefU8wmlpW1pAcBG1Q3ylAAAFQD/YUDVdM9S0gPFN8x5v9K7xYexpwAAAIAnZMn5SHEFPc8InPYDA0uMMLwQYIvDs0X7nAj39suL8TurM+nc
syvoI0ZL941vZp2bLXTLWpF4o62+CFMVWmHvmt84aR1H/q8LVN8GtuAH/qgJ1n10Xfz+GQmDZP7TCYEmK6q0LMU680enWfIDA/Ba3+sNjTpycFcqFjQNS10UgAAAIBVnaM0hb5Rv3pu340Q
46bv/66MLK3oxrG8G2Xg18DH+E6tTtrvjRk3SAeY6Is6x8yAC36u7cRucEKAGP/Ck3HCh2Fpudo6be2J6LbQ9qx4ZrR:91ToYd4BmwJkPXrtyLsJGHRNobX6Uj5C/LD9yZ+WjHxqTJWJ3T
ORRq24su1w== arjun@ubuntu
arjun@ubuntu:~$ ssh localhost
Welcome to Ubuntu 12.04.1 LTS (GNU/Linux 3.2.0-29-generic-pae i686)

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

609 packages can be updated.
288 updates are security updates.

New release '14.04.2 LTS' available.
Run 'do-release-upgrade' to upgrade to it.
```

4. Download Hadoop

download Apache hadoop

5. Install Hadoop

5.1)The downloaded tar file is extracted.

command:- tar xzf hadoop-2.2.0.tar.gz

5.2) Goto the folder

command:- cd hadoop-2.2.0

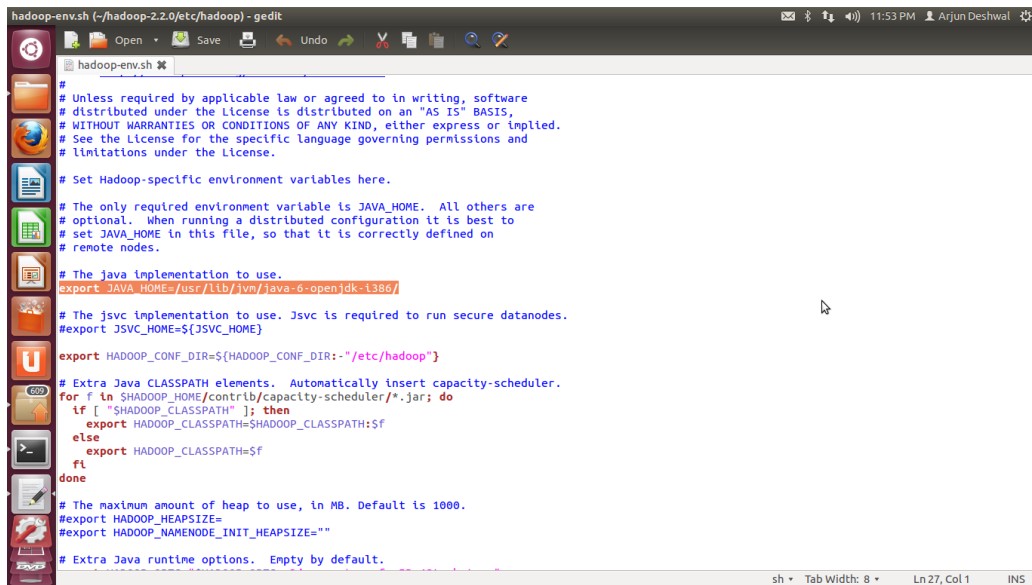
6. Setup / Configure

6.1) Edit configuration file (hadoop-env.sh) and set JAVA_HOME. (Fig 1)

6.1.1) open file (hadoop-env.sh) folder: `-$HOME/hadoop.2.2.0/etc/hadoop/hadoop-env.sh`

6.1.2) set path:-

command: `-export JAVA_HOME=/usr/lib/jvm/java-6-openjdk-i386/`



```
hadoop-env.sh (~/.hadoop-2.2.0/etc/hadoop) - gedit
hadoop-env.sh
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.

# Set Hadoop-specific environment variables here.

# The only required environment variable is JAVA_HOME. All others are
# optional. When running a distributed configuration it is best to
# set JAVA_HOME in this file, so that it is correctly defined on
# remote nodes.

# The java implementation to use.
export JAVA_HOME=/usr/lib/jvm/java-6-openjdk-i386/

# The jsvc implementation to use. Jsvc is required to run secure datanodes.
export JSVC_HOME=${JSVC_HOME}

export HADOOP_CONF_DIR=${HADOOP_CONF_DIR:-"/etc/hadoop"}

# Extra Java CLASSPATH elements. Automatically insert capacity-scheduler.
for f in $HADOOP_HOME/contrib/capacity-scheduler/*.jar; do
  if [ "$HADOOP_CLASSPATH" ]; then
    export HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$f
  else
    export HADOOP_CLASSPATH=$f
  fi
done

# The maximum amount of heap to use, in MB. Default is 1000.
#export HADOOP_HEAPSIZE=
#export HADOOP_NAMENODE_INIT_HEAPSIZE=""

# Extra Java runtime options. Empty by default.
```

Fig 1

6.2) Edit configuration file (core-site.xml)(shown in Fig 2):

```
<configuration>
<property>
<name>fs.default.name</name>
<value>hdfs://localhost:9000</value>
</property>
</configuration>
```

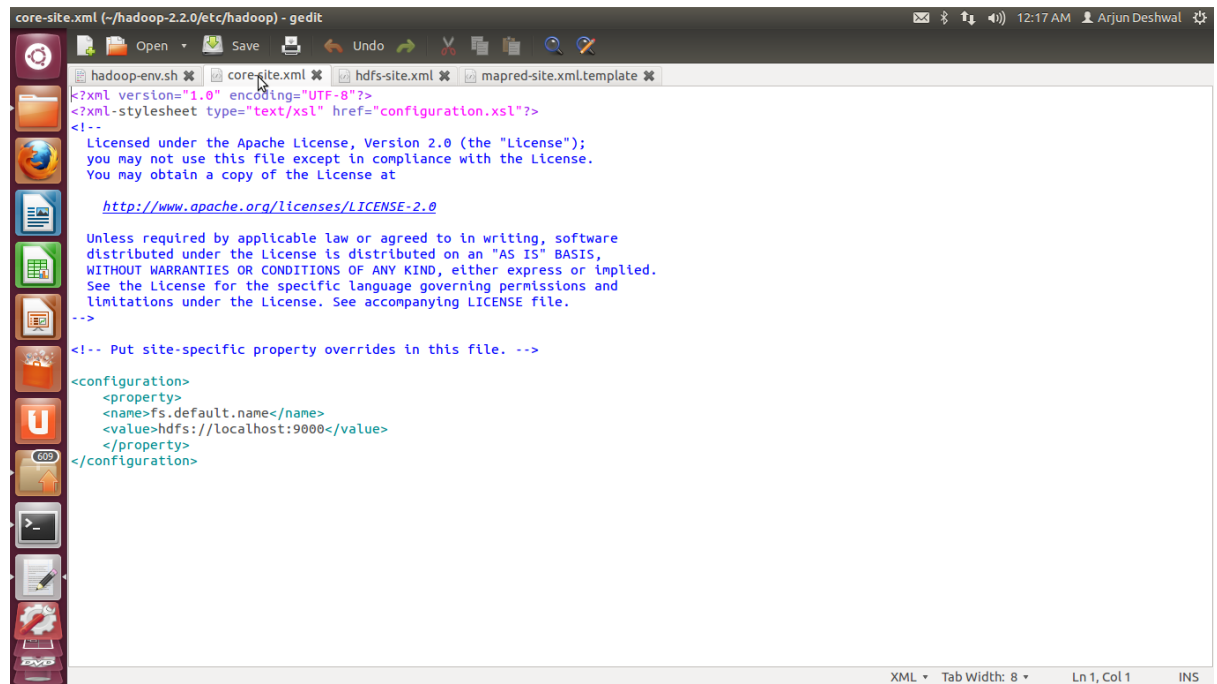


Fig 2

6.3) Edit configuration file (hdfs-site.xml) (shown in Fig 3)

```
<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
</configuration>
```

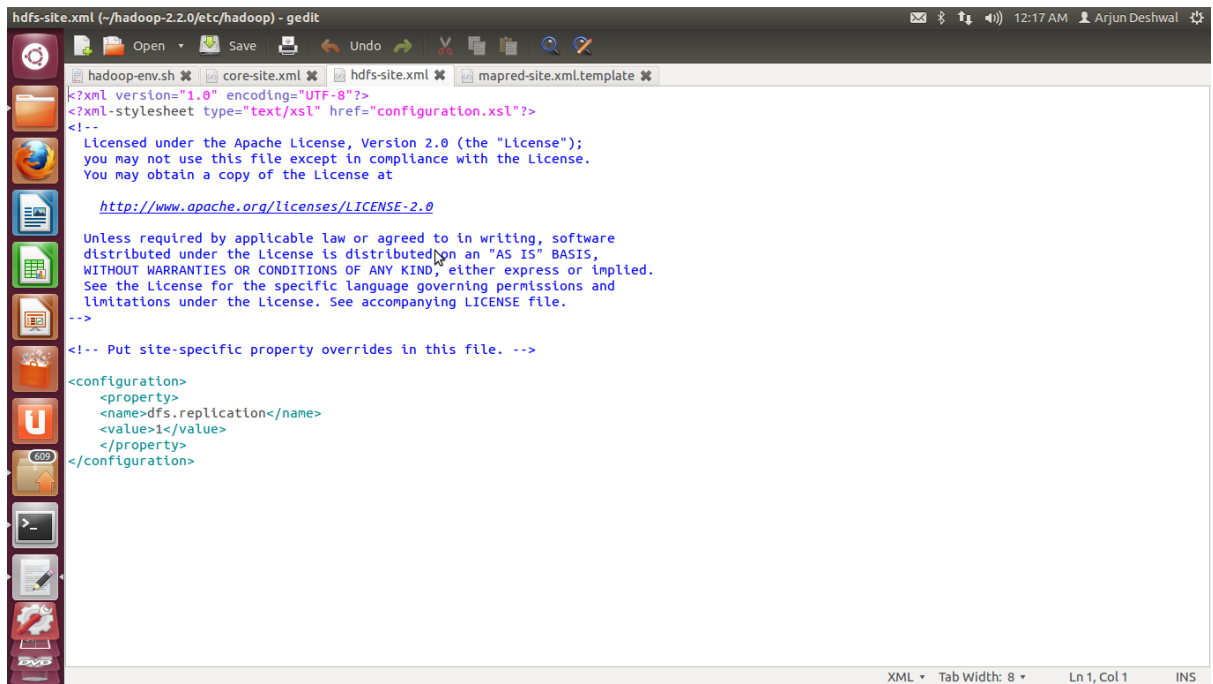


Fig 3

6.4) Edit configuration file (mapred-site.xml) (shown in Fig 4)

```
<configuration>
<property>
<name>mapred.job.tracker</name>
<value>localhost:9001</value>
</property>
</configuration>
```

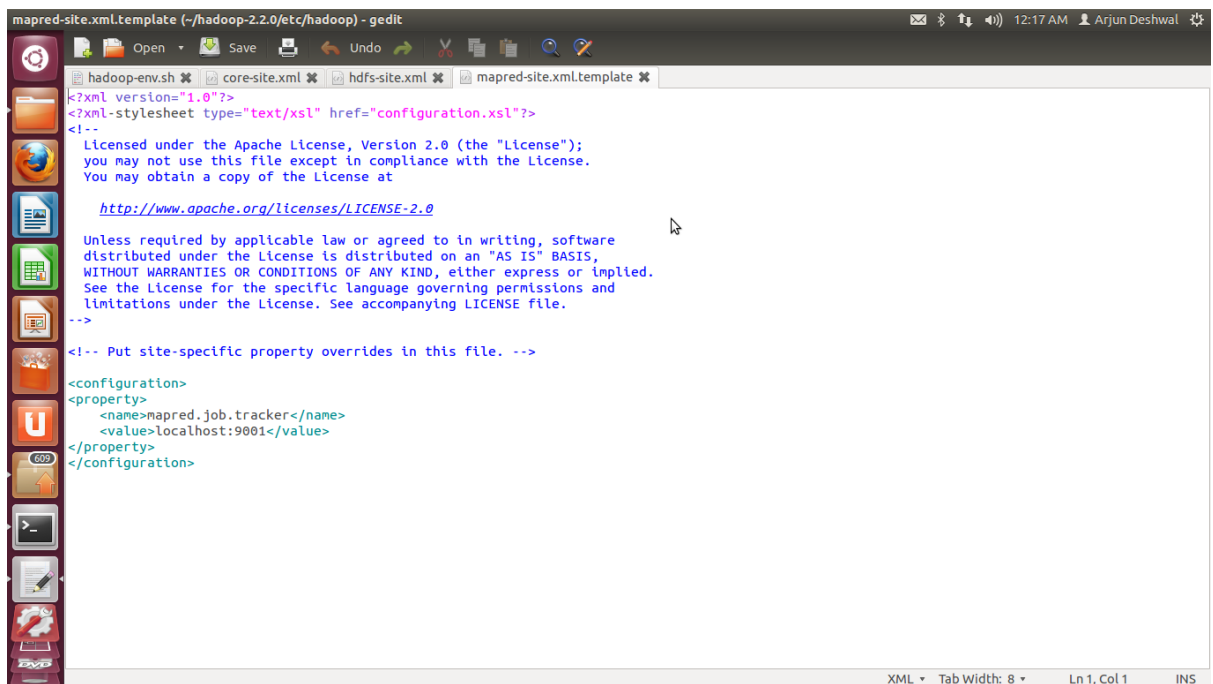
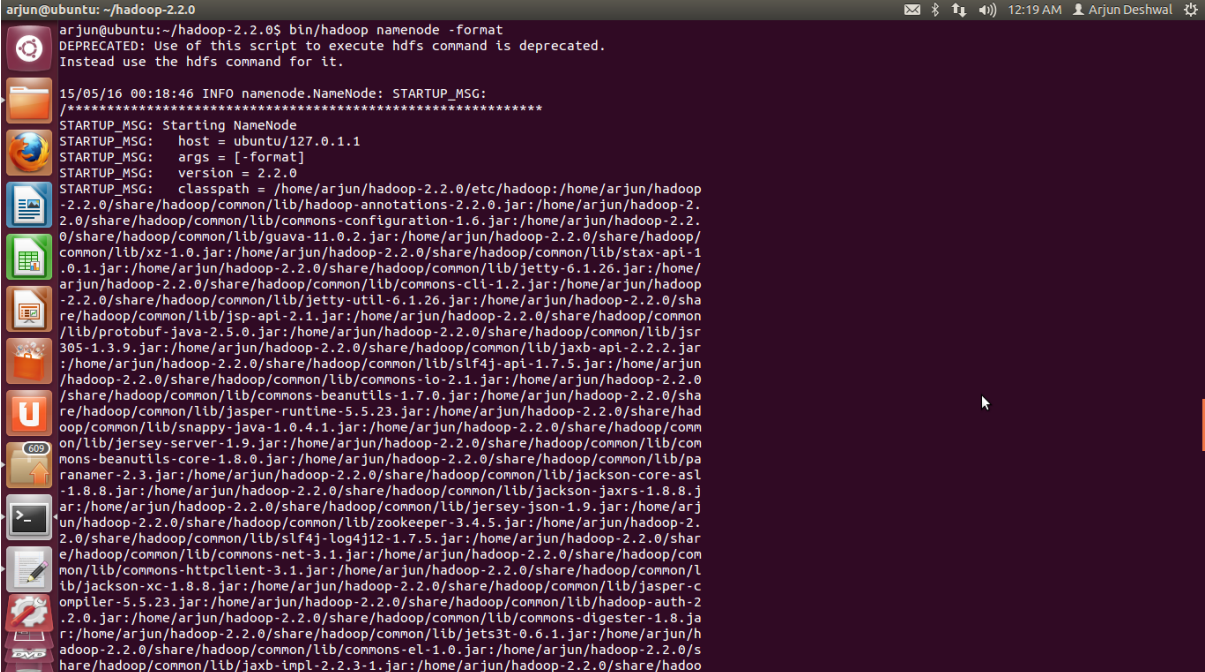


Fig 4

7. Start Hadoop cluster

7.1) Format Namenode(IMP :- this is done once when first time installing hadoop otherwise it will delete the data).

command:- bin/hadoop namenode -format



```
arjun@ubuntu: ~/hadoop-2.2.0
arjun@ubuntu:~/hadoop-2.2.0$ bin/hadoop namenode -format
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

15/05/16 00:18:46 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = ubuntu/127.0.1.1
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 2.2.0
STARTUP_MSG: classpath = /home/arjun/hadoop-2.2.0/etc/hadoop:/home/arjun/hadoop
-2.2.0/share/hadoop/common/lib/hadoop-annotations-2.2.0.jar:/home/arjun/hadoop-2.
2.0/share/hadoop/common/lib/commons-configuration-1.6.jar:/home/arjun/hadoop-2.2.
0/share/hadoop/common/lib/guava-11.0.2.jar:/home/arjun/hadoop-2.2.0/share/hadoop
/common/lib/xz-1.0.jar:/home/arjun/hadoop-2.2.0/share/hadoop/common/lib/stax-api-1
.0.1.jar:/home/arjun/hadoop-2.2.0/share/hadoop/common/lib/jetty-6.1.26.jar:/home
arjun/hadoop-2.2.0/share/hadoop/common/lib/commons-cli-1.2.jar:/home/arjun/hadoop
-2.2.0/share/hadoop/common/lib/jetty-util-6.1.26.jar:/home/arjun/hadoop-2.2.0/sha
re/hadoop/common/lib/jsp-api-2.1.jar:/home/arjun/hadoop-2.2.0/share/hadoop/common
/lib/protobuf-java-2.5.0.jar:/home/arjun/hadoop-2.2.0/share/hadoop/common/lib/jsr
305-1.3.9.jar:/home/arjun/hadoop-2.2.0/share/hadoop/common/lib/jaxb-api-2.2.2.jar
:/home/arjun/hadoop-2.2.0/share/hadoop/common/lib/slf4j-api-1.7.5.jar:/home/arjun
/hadoop-2.2.0/share/hadoop/common/lib/commons-io-2.1.jar:/home/arjun/hadoop-2.2.0
/share/hadoop/common/lib/commons-beanutils-1.7.0.jar:/home/arjun/hadoop-2.2.0/sha
re/hadoop/common/lib/jasper-runtime-5.5.23.jar:/home/arjun/hadoop-2.2.0/share/had
oop/common/lib/snappy-java-1.0.4.1.jar:/home/arjun/hadoop-2.2.0/share/hadoop/comm
on/lib/jersey-server-1.9.jar:/home/arjun/hadoop-2.2.0/share/hadoop/common/lib/com
mons-beanutils-core-1.8.0.jar:/home/arjun/hadoop-2.2.0/share/hadoop/common/lib/pa
ranamer-2.3.jar:/home/arjun/hadoop-2.2.0/share/hadoop/common/lib/jackson-core-asl
-1.8.8.jar:/home/arjun/hadoop-2.2.0/share/hadoop/common/lib/jackson-jaxrs-1.8.8.j
ar:/home/arjun/hadoop-2.2.0/share/hadoop/common/lib/jersey-json-1.9.jar:/home/arj
un/hadoop-2.2.0/share/hadoop/common/lib/zookeeper-3.4.5.jar:/home/arjun/hadoop-2.
2.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar:/home/arjun/hadoop-2.2.0/sha
re/hadoop/common/lib/commons-net-3.1.jar:/home/arjun/hadoop-2.2.0/share/hadoop/com
mon/lib/commons-httpclient-3.1.jar:/home/arjun/hadoop-2.2.0/share/hadoop/common/l
ib/jackson-xc-1.8.8.jar:/home/arjun/hadoop-2.2.0/share/hadoop/common/lib/jasper-c
ompiler-5.5.23.jar:/home/arjun/hadoop-2.2.0/share/hadoop/common/lib/hadoop-auth-2
.2.0.jar:/home/arjun/hadoop-2.2.0/share/hadoop/common/lib/commons-digester-1.8.ja
r:/home/arjun/hadoop-2.2.0/share/hadoop/common/lib/jets3t-0.6.1.jar:/home/arjun/h
adoop-2.2.0/share/hadoop/common/lib/commons-el-1.0.jar:/home/arjun/hadoop-2.2.0/s
hare/hadoop/common/lib/jaxb-impl-2.2.3-1.jar:/home/arjun/hadoop-2.2.0/share/hadoo
```

7.2) start Hadoop

command:-sbin/start-all.sh

8. Check hadoop cluster running

command :-jps

9. Stop hadoop cluster

command:-sbin/stop-all.sh

```
arjun@ubuntu: ~/hadoop-2.2.0
arjun@ubuntu:~/hadoop-2.2.0$ sbin/start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [localhost]
localhost: starting namenode, logging to /home/arjun/hadoop-2.2.0/logs/hadoop-arjun-namenode-ubuntu.out
localhost: starting datanode, logging to /home/arjun/hadoop-2.2.0/logs/hadoop-arjun-datanode-ubuntu.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/arjun/hadoop-2.2.0/logs/hadoop-arjun-secondarynamenode-ubuntu.out
starting yarn daemons
localhost: starting resourcemanager, logging to /home/arjun/hadoop-2.2.0/logs/yarn-arjun-resourcemanager-ubuntu.out
localhost: starting nodemanager, logging to /home/arjun/hadoop-2.2.0/logs/yarn-arjun-nodemanager-ubuntu.out
arjun@ubuntu:~/hadoop-2.2.0$
arjun@ubuntu:~/hadoop-2.2.0$
arjun@ubuntu:~/hadoop-2.2.0$
arjun@ubuntu:~/hadoop-2.2.0$
arjun@ubuntu:~/hadoop-2.2.0$
arjun@ubuntu:~/hadoop-2.2.0$
arjun@ubuntu:~/hadoop-2.2.0$ jps
6828 ResourceManager
7212 Jps
7036 NodeManager
6445 DataNode
6250 NameNode
6696 SecondaryNameNode
arjun@ubuntu:~/hadoop-2.2.0$
arjun@ubuntu:~/hadoop-2.2.0$
arjun@ubuntu:~/hadoop-2.2.0$
arjun@ubuntu:~/hadoop-2.2.0$
arjun@ubuntu:~/hadoop-2.2.0$
arjun@ubuntu:~/hadoop-2.2.0$ sbin/stop-all.sh
This script is Deprecated. Instead use stop-dfs.sh and stop-yarn.sh
Stopping namenodes on [localhost]
localhost: stopping namenode
localhost: stopping datanode
Stopping secondary namenodes [0.0.0.0]
0.0.0.0: stopping secondarynamenode
stopping yarn daemons
stopping resourcemanager
localhost: stopping nodemanager
no proxyserver to stop
arjun@ubuntu:~/hadoop-2.2.0$
```

WEB CONSOLE VIEW

Open Browser and type in the URL:-<http://localhost:50070>

Hadoop NameNode localhost:9000 - Mozilla Firefox

localhost:50070/dfshealth.jsp

NameNode 'localhost:9000' (active)

Started:	Sat May 16 00:27:14 PDT 2015
Version:	2.2.0, 1529768
Compiled:	2013-10-07T06:28Z by hortonmu from branch-2.2.0
Cluster ID:	CID-a9bde77f-fdd1-40c1-9de4-0de3ae0045fd
Block Pool ID:	BP-622258529-127.0.1.1-1431760728635

[Browse the filesystem](#)
[NameNode Logs](#)

Cluster Summary

Security is OFF
1 files and directories, 0 blocks = 1 total.
Heap Memory used 39.12 MB is 86% of Committed Heap Memory 45.43 MB. Max Heap Memory is 966.69 MB.
Non Heap Memory used 26.13 MB is 73% of Committed Non Heap Memory 35.72 MB. Max Non Heap Memory is 118 MB.

Configured Capacity	:	18.94 GB			
DFS Used	:	24 KB			
Non DFS Used	:	5.34 GB			
DFS Remaining	:	13.60 GB			
DFS Used%	:	0.00%			
DFS Remaining%	:	71.82%			
Block Pool Used	:	24 KB			
Block Pool Used%	:	0.00%			
DataNodes usages	:	Min %	Median %	Max %	stdev %
	:	0.00%	0.00%	0.00%	0.00%

[Live Nodes](#) : 1 (Decommissioned: 0)