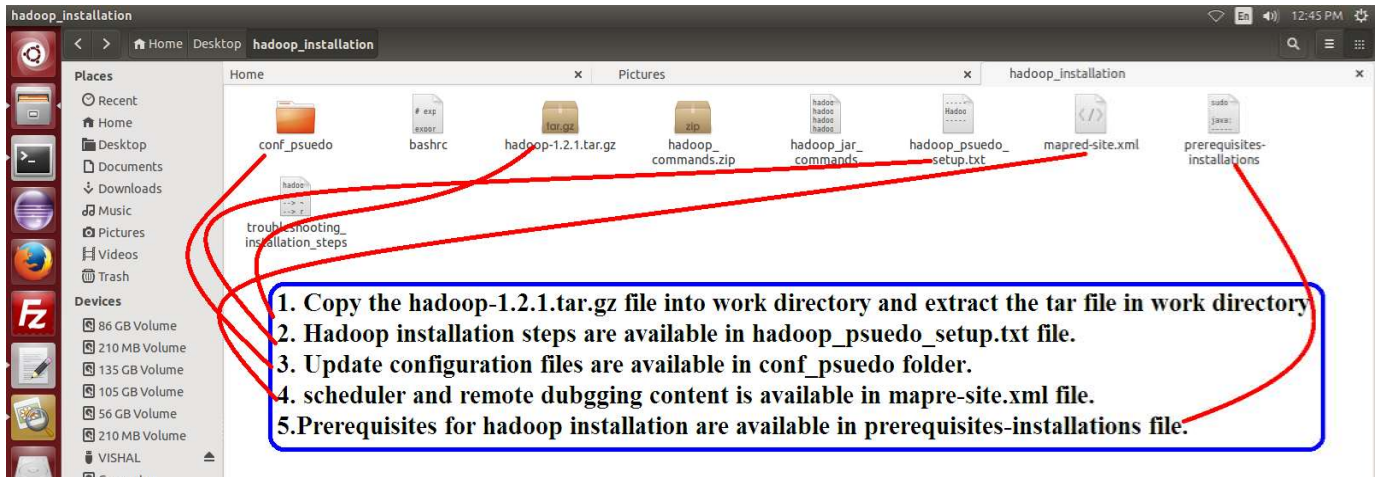
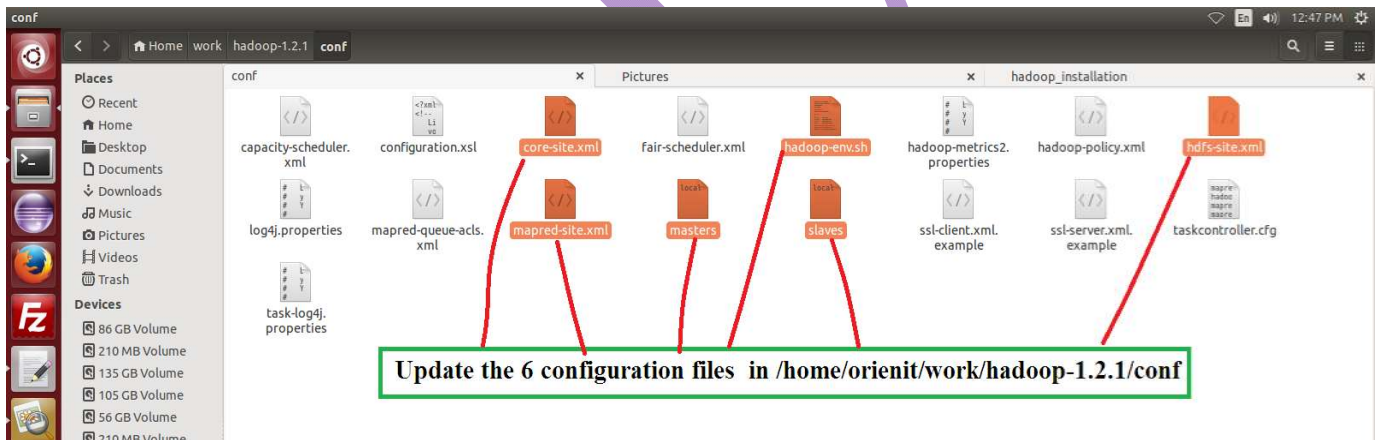


# HADOOP-1.2.1 Version Installation Steps

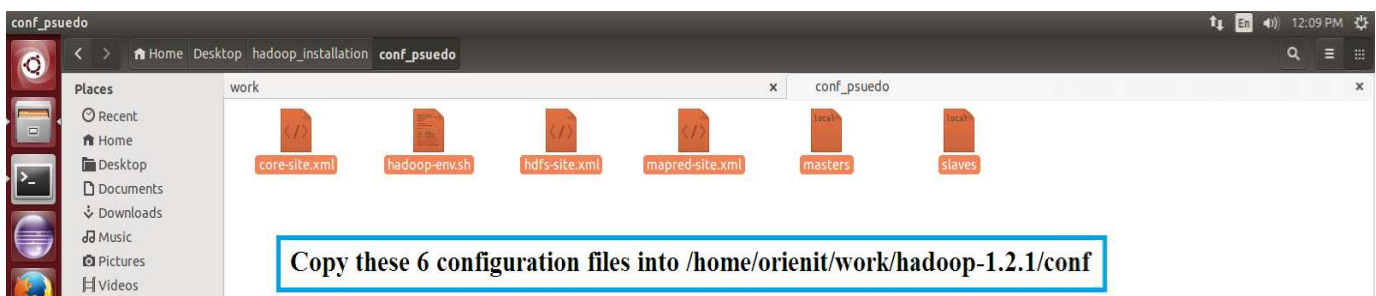
1. Download the **hadoop-1.2.1** version from [Apache Mirrors](#)
2. Create the **work** directory in **orientit** user, path is ``/home/orientit``
3. Please follow the below screen shot steps.



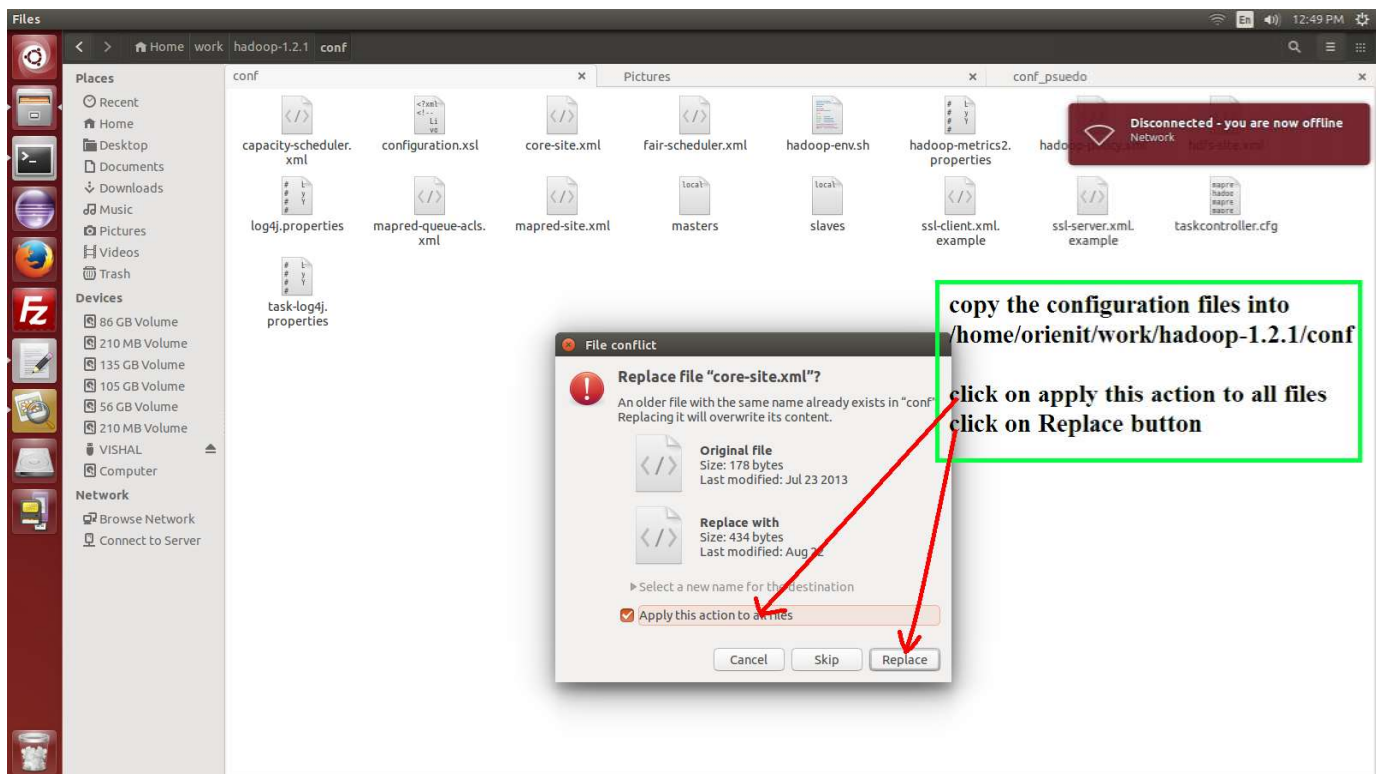
4. By default hadoop configuration files are empty.  
As part of hadoop installation we need to update the configurations files as per our requirements.



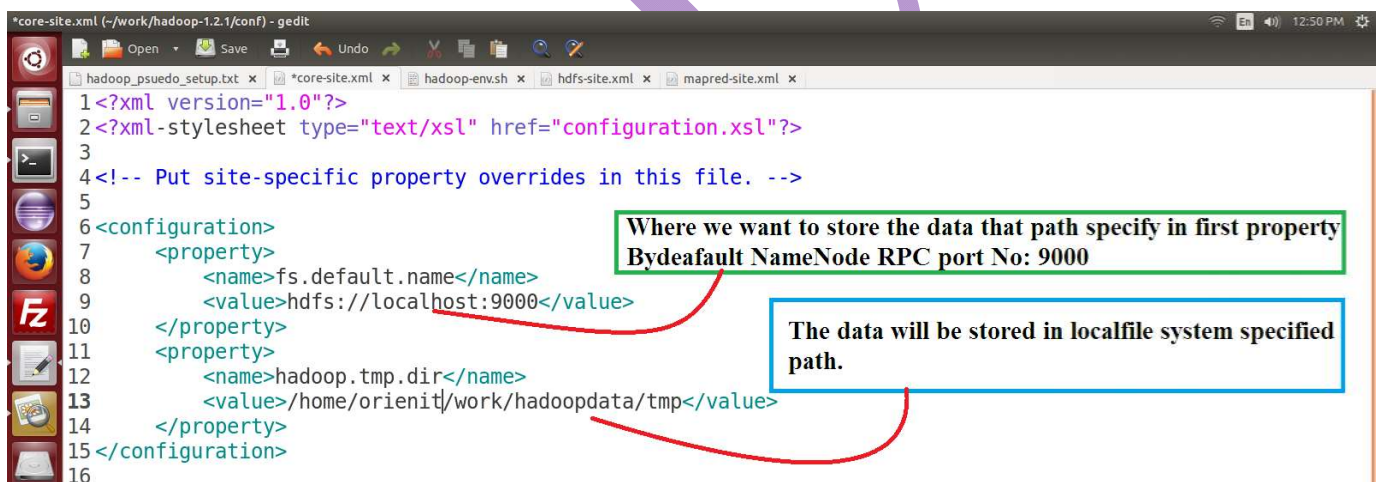
5. The below 6 configuration files having **installation properties**, we need to copy these 6 files into `/home/orientit/work/hadoop-1.2.1/conf`



## 6. Steps is available on screen shots, we can follow that actions

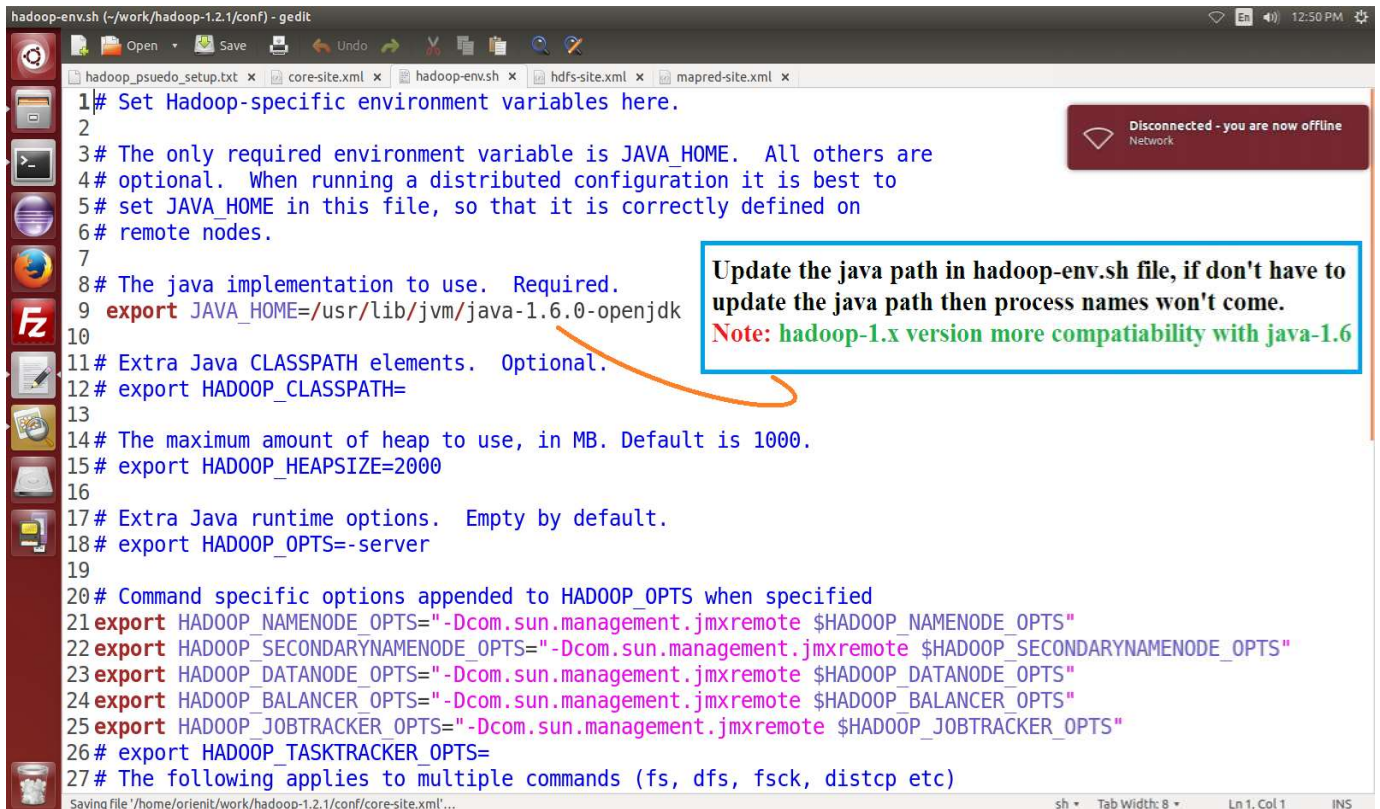


## 7. Open **core-site.xml** file, verify the properties and understanding the properties description.





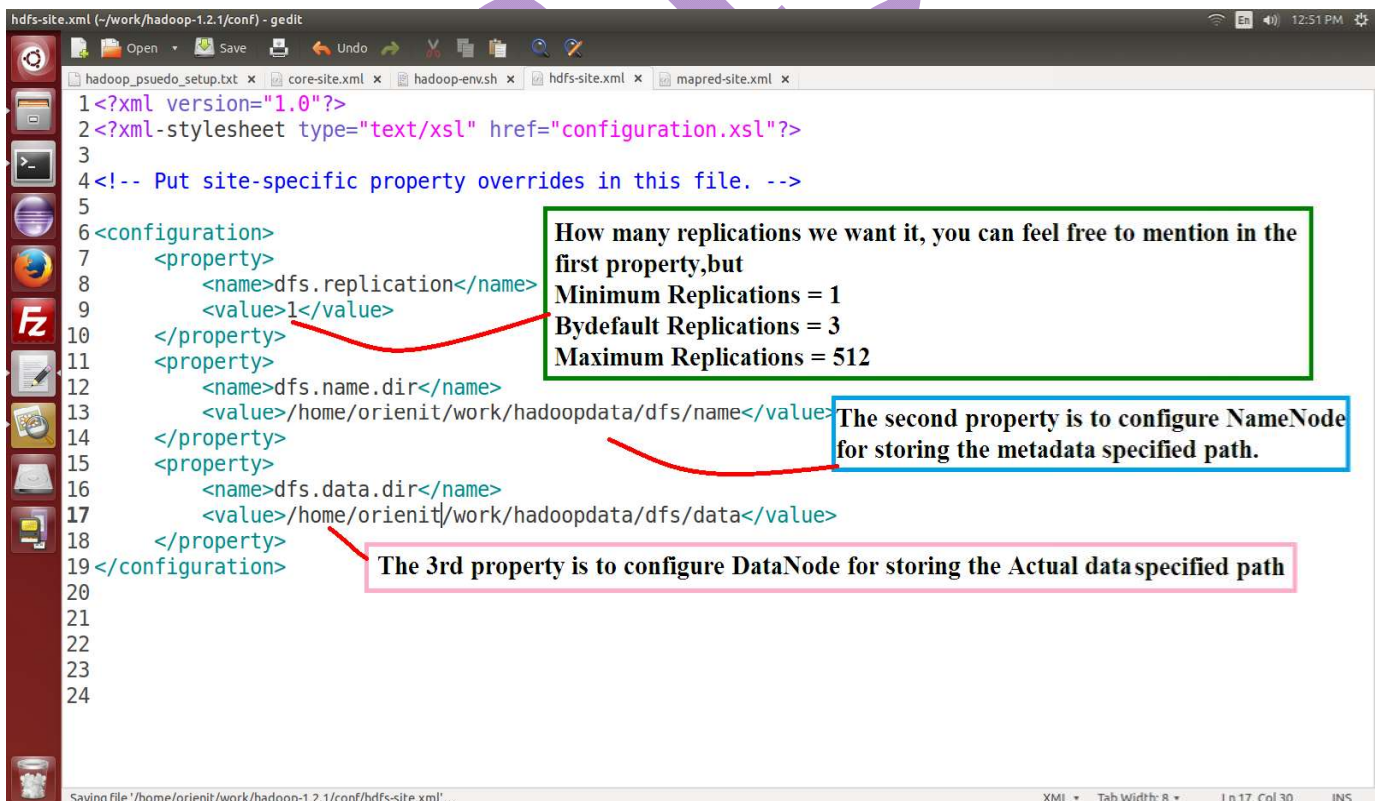
## 8. Open the **hadoop-env.sh** file and follow the below screen shot information.



```
hadoop-env.sh (~/.work/hadoop-1.2.1/conf) - gedit
1# Set Hadoop-specific environment variables here.
2
3# The only required environment variable is JAVA_HOME. All others are
4# optional. When running a distributed configuration it is best to
5# set JAVA_HOME in this file, so that it is correctly defined on
6# remote nodes.
7
8# The java implementation to use. Required.
9 export JAVA_HOME=/usr/lib/jvm/java-1.6.0-openjdk
10
11# Extra Java CLASSPATH elements. Optional.
12# export HADOOP_CLASSPATH=
13
14# The maximum amount of heap to use, in MB. Default is 1000.
15# export HADOOP_HEAPSIZE=2000
16
17# Extra Java runtime options. Empty by default.
18# export HADOOP_OPTS=-server
19
20# Command specific options appended to HADOOP_OPTS when specified
21 export HADOOP_NAMENODE_OPTS="-Dcom.sun.management.jmxremote $HADOOP_NAMENODE_OPTS"
22 export HADOOP_SECONDARYNAMENODE_OPTS="-Dcom.sun.management.jmxremote $HADOOP_SECONDARYNAMENODE_OPTS"
23 export HADOOP_DATANODE_OPTS="-Dcom.sun.management.jmxremote $HADOOP_DATANODE_OPTS"
24 export HADOOP_BALANCER_OPTS="-Dcom.sun.management.jmxremote $HADOOP_BALANCER_OPTS"
25 export HADOOP_JOBTRACKER_OPTS="-Dcom.sun.management.jmxremote $HADOOP_JOBTRACKER_OPTS"
26# export HADOOP_TASKTRACKER_OPTS=
27# The following applies to multiple commands (fs, dfs, fsck, distcp etc)
```

Update the java path in hadoop-env.sh file, if don't have to update the java path then process names won't come.  
**Note: hadoop-1.x version more compatiability with java-1.6**

## 9. Open the **hdfs-site.xml** file and follow the below properties.



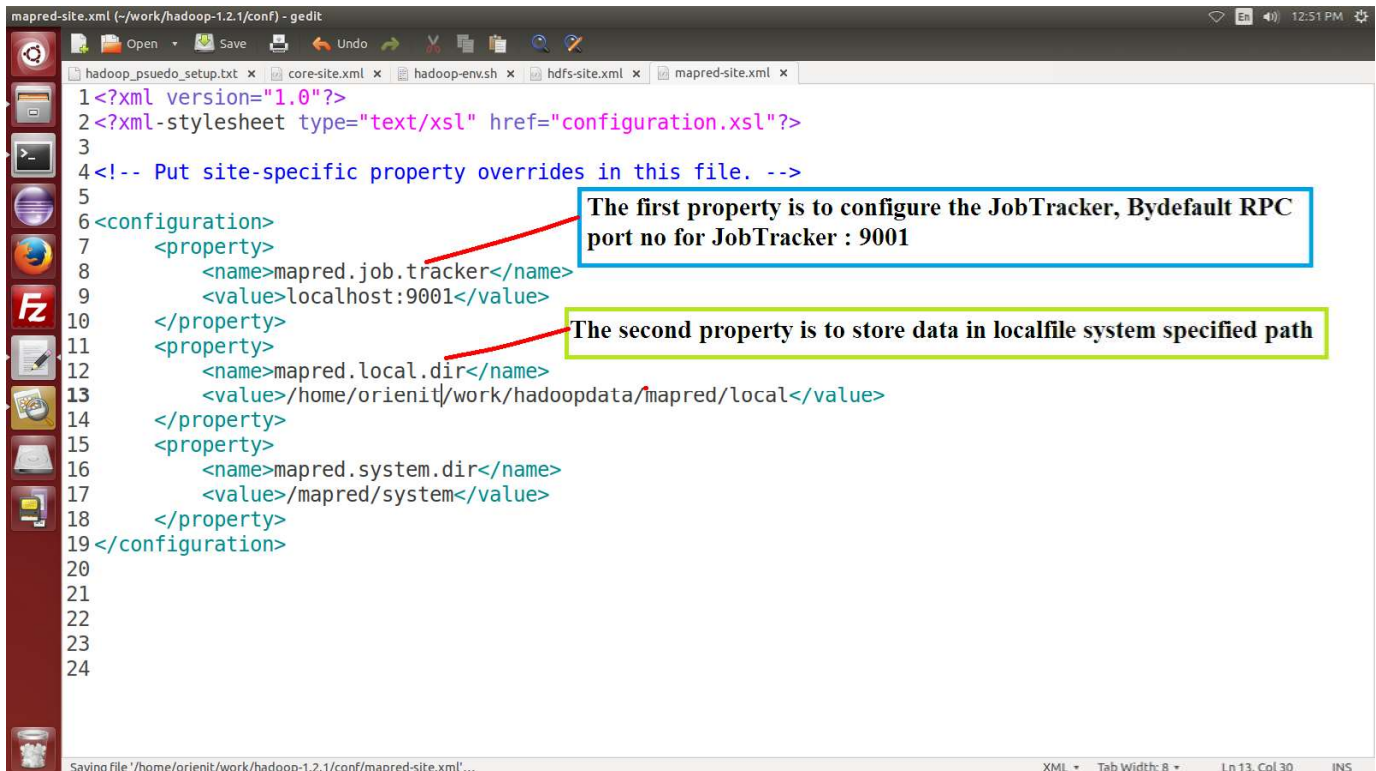
```
hdfs-site.xml (~/.work/hadoop-1.2.1/conf) - gedit
1<?xml version="1.0"?>
2<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
3
4<!-- Put site-specific property overrides in this file. -->
5
6<configuration>
7  <property>
8    <name>dfs.replication</name>
9    <value>1</value>
10  </property>
11  <property>
12    <name>dfs.name.dir</name>
13    <value>/home/orienit/work/hadoopdata/dfs/name</value>
14  </property>
15  <property>
16    <name>dfs.data.dir</name>
17    <value>/home/orienit/work/hadoopdata/dfs/data</value>
18  </property>
19</configuration>
20
21
22
23
24
```

How many replications we want it, you can feel free to mention in the first property, but  
Minimum Replications = 1  
By default Replications = 3  
Maximum Replications = 512

The second property is to configure NameNode for storing the metadata specified path.

The 3rd property is to configure DataNode for storing the Actual data specified path

10. Open the **mapred-site.xml** file and understand the screen shot information

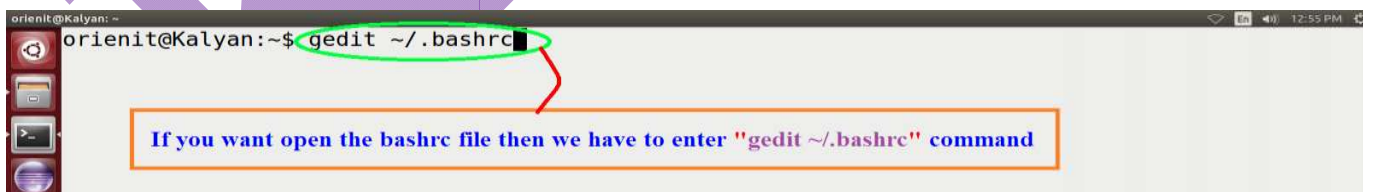


```
1 <?xml version="1.0"?>
2 <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
3
4 <!-- Put site-specific property overrides in this file. -->
5
6 <configuration>
7   <property>
8     <name>mapred.job.tracker</name>
9     <value>localhost:9001</value>
10  </property>
11  <property>
12    <name>mapred.local.dir</name>
13    <value>/home/orienit/work/hadoopdata/mapred/local</value>
14  </property>
15  <property>
16    <name>mapred.system.dir</name>
17    <value>/mapred/system</value>
18  </property>
19 </configuration>
20
21
22
23
24
```

The first property is to configure the JobTracker, By default RPC port no for JobTracker : 9001

The second property is to store data in local file system specified path

11. If we want to install **hadoop-1.2.1**, we need to update the **~/.bashrc** file. To open **~/.bashrc** file using below command.

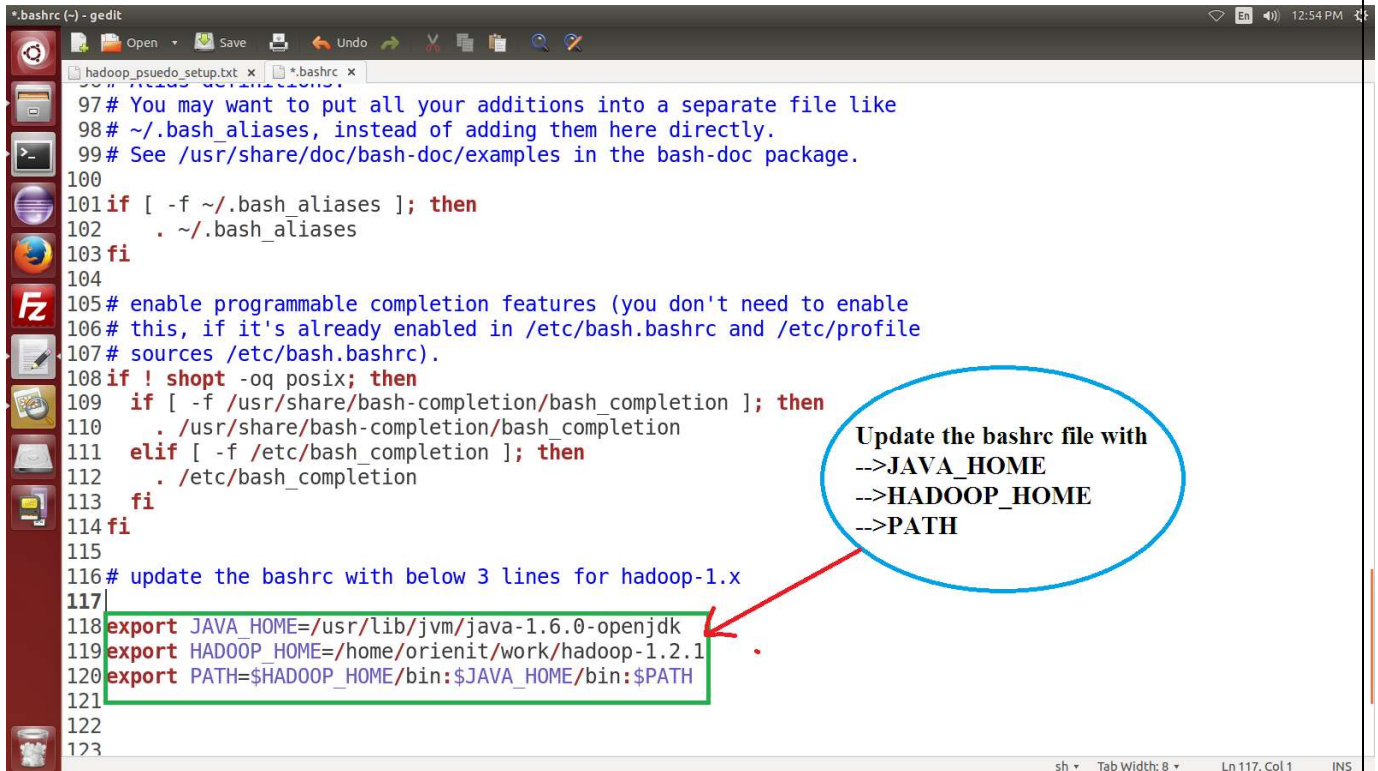


```
orienit@Kalyan: ~$ gedit ~/.bashrc
```

If you want to open the bashrc file then we have to enter "gedit ~/.bashrc" command

12. Update the **hadoop-1.2.1** related information in **~/.bashrc** file

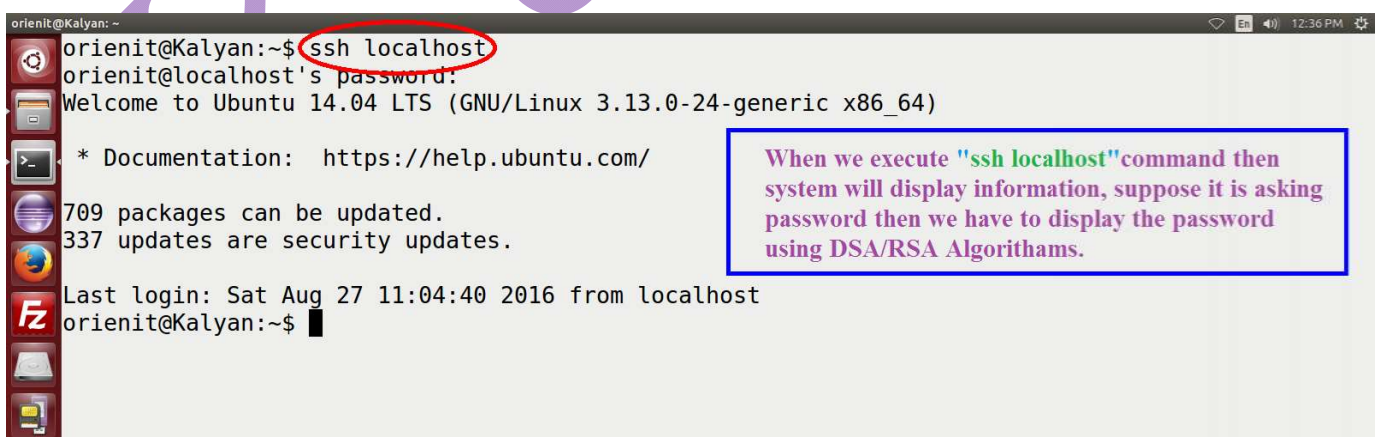
```
export JAVA_HOME=/usr/lib/jvm/java-1.6.0-openjdk
export HADOOP_HOME=/home/orienit/work/hadoop-1.2.1
export PATH=$HADOOP_HOME/bin:$JAVA_HOME/bin:$PATH
```



```
*.bashrc (-) - gedit
97# You may want to put all your additions into a separate file like
98# ~/.bash_aliases, instead of adding them here directly.
99# See /usr/share/doc/bash-doc/examples in the bash-doc package.
100
101if [ -f ~/.bash_aliases ]; then
102    . ~/.bash_aliases
103fi
104
105# enable programmable completion features (you don't need to enable
106# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
107# sources /etc/bash.bashrc).
108if ! shopt -oq posix; then
109    if [ -f /usr/share/bash-completion/bash_completion ]; then
110        . /usr/share/bash-completion/bash_completion
111    elif [ -f /etc/bash_completion ]; then
112        . /etc/bash_completion
113    fi
114fi
115
116# update the bashrc with below 3 lines for hadoop-1.x
117
118export JAVA_HOME=/usr/lib/jvm/java-1.6.0-openjdk
119export HADOOP_HOME=/home/orienit/work/hadoop-1.2.1
120export PATH=$HADOOP_HOME/bin:$JAVA_HOME/bin:$PATH
121
122
123
```

Update the bashrc file with  
-->JAVA\_HOME  
-->HADOOP\_HOME  
-->PATH

13. Ssh is secure shell, follow the screen shot information.  
(ssh localhost)



```
orienit@Kalyan: ~
orienit@Kalyan:~$ ssh localhost
orienit@localhost's password:
Welcome to Ubuntu 14.04 LTS (GNU/Linux 3.13.0-24-generic x86_64)

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

709 packages can be updated.
337 updates are security updates.


Last login: Sat Aug 27 11:04:40 2016 from localhost
orienit@Kalyan:~$
```

When we execute "ssh localhost" command then system will display information, suppose it is asking password then we have to display the password using DSA/RSA Algorithms.



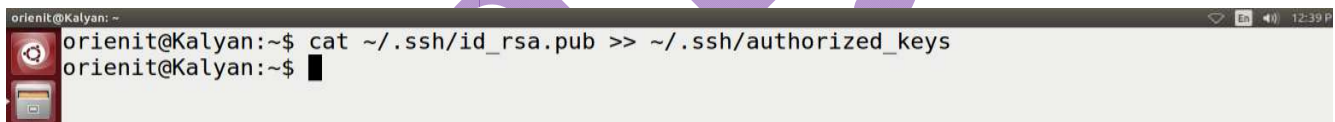
#### 14. Follow the below commands to Disable the password securely

```
ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa  
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys  
chmod 600 ~/.ssh/authorized_keys
```



```
orienit@Kalyan: ~  
orienit@Kalyan:~$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa  
Generating public/private rsa key pair.  
/home/orienit/.ssh/id_rsa already exists.  
Overwrite (y/n)? y  
Your identification has been saved in /home/orienit/.ssh/id_rsa.  
Your public key has been saved in /home/orienit/.ssh/id_rsa.pub.  
The key fingerprint is:  
a4:23:19:7c:fa:13:18:c6:c5:b6:7b:3c:b6:6f:4f:36 orienit@Kalyan  
The key's randomart image is:  
+---[ RSA 2048]-----+  
  ..  
  o .o  
  *....  
  . B.o  
  = +oS  
  o.o=  
  oo o E  
  .. .o .  
  o...  
+-----+  
orienit@Kalyan:~$
```

These command is used to disable the password using RSA/DSA Algorithms, when we execute these command then system won't ask the password multiple times.



```
orienit@Kalyan: ~  
orienit@Kalyan:~$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys  
orienit@Kalyan:~$
```

## 15. Format the Name node using below command (hadoop namenode -format)

```
orientit@Kalyan:~$ hadoop namenode -format
Warning: $HADOOP_HOME is deprecated.

16/09/07 12:56:57 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = Kalyan/127.0.0.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 1.2.1
STARTUP_MSG:   build = https://svn.apache.org/repos/asf/hadoop/common/branches/branch-1.2 -r 15
03152; compiled by 'mattf' on Mon Jul 22 15:23:09 PDT 2013
STARTUP_MSG:   java = 1.6.0_39
*****/

16/09/07 12:56:57 INFO util.GSet: Computing capacity for map BlocksMap
16/09/07 12:56:57 INFO util.GSet: VM type           = 64-bit
16/09/07 12:56:57 INFO util.GSet: 2.0% max memory = 932118528
16/09/07 12:56:57 INFO util.GSet: capacity        = 2^21 = 2097152 entries
16/09/07 12:56:57 INFO util.GSet: recommended=2097152, actual=2097152
16/09/07 12:56:57 INFO namenode.FSNamesystem: fsOwner=orientit
16/09/07 12:56:58 INFO namenode.FSNamesystem: supergroup=supergroup
16/09/07 12:56:58 INFO namenode.FSNamesystem: isPermissionEnabled=true
16/09/07 12:56:58 INFO namenode.FSNamesystem: dfs.block.invalidate.limit=100
16/09/07 12:56:58 INFO namenode.FSNamesystem: isAccessTokenEnabled=false accessKeyUpdateInterva
l=0 min(s), accessTokenLifetime=0 min(s)
16/09/07 12:56:58 INFO namenode.FSEditLog: dfs.namenode.edits.toleration.length = 0
16/09/07 12:56:58 INFO namenode.NameNode: Caching file names occuring more than 10 times
16/09/07 12:56:58 INFO common.Storage: Image file /home/orientit/work/hadoopdata/dfs/name/curren

orientit@Kalyan:~$
*****/
16/09/07 12:56:57 INFO util.GSet: Computing capacity for map BlocksMap
16/09/07 12:56:57 INFO util.GSet: VM type           = 64-bit
16/09/07 12:56:57 INFO util.GSet: 2.0% max memory = 932118528
16/09/07 12:56:57 INFO util.GSet: capacity        = 2^21 = 2097152 entries
16/09/07 12:56:57 INFO util.GSet: recommended=2097152, actual=2097152
16/09/07 12:56:57 INFO namenode.FSNamesystem: fsOwner=orientit
16/09/07 12:56:58 INFO namenode.FSNamesystem: supergroup=supergroup
16/09/07 12:56:58 INFO namenode.FSNamesystem: isPermissionEnabled=true
16/09/07 12:56:58 INFO namenode.FSNamesystem: dfs.block.invalidate.limit=100
16/09/07 12:56:58 INFO namenode.FSNamesystem: isAccessTokenEnabled=false accessKeyUpdateInterva
l=0 min(s), accessTokenLifetime=0 min(s)
16/09/07 12:56:58 INFO namenode.FSEditLog: dfs.namenode.edits.toleration.length = 0
16/09/07 12:56:58 INFO namenode.NameNode: Caching file names occuring more than 10 times
16/09/07 12:56:58 INFO common.Storage: Image file /home/orientit/work/hadoopdata/dfs/name/curren
t/fsimage of size 113 bytes saved in 0 seconds.
16/09/07 12:56:58 INFO namenode.FSEditLog: closing edit log: position=4, editlog=/home/orientit/
work/hadoopdata/dfs/name/current/edits
16/09/07 12:56:58 INFO namenode.FSEditLog: close success: truncate to 4, editlog=/home/orientit/
work/hadoopdata/dfs/name/current/edits
16/09/07 12:56:58 INFO common.Storage: Storage directory /home/orientit/work/hadoopdata/dfs/name
has been successfully formatted.
16/09/07 12:56:58 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at Kalyan/127.0.0.1
*****/
orientit@Kalyan:~$
```

Before start the hadoop we must and should format the NameNode otherwise NameNode won't come, After executing NameNode command then namenode information display on console



16. Start the Hadoop using below command (start-all.sh) and then observe the screen shot information.

```
orienit@Kalyan:~$ start-all.sh
Warning: $HADOOP_HOME is deprecated.

starting namenode, logging to /home/orienit/work/hadoop-1.2.1/libexec/../logs/hadoop-namenode-Kalyan.out
localhost: starting datanode, logging to /home/orienit/work/hadoop-1.2.1/libexec/../logs/hadoop-orienit-datanode-Kalyan.out
localhost: starting secondarynamenode, logging to /home/orienit/work/hadoop-1.2.1/libexec/../logs/hadoop-orienit-secondarynamenode-Kalyan.out
starting jobtracker, logging to /home/orienit/work/hadoop-1.2.1/libexec/../logs/hadoop-orienit-jobtracker-Kalyan.out
localhost: starting tasktracker, logging to /home/orienit/work/hadoop-1.2.1/libexec/../logs/hadoop-orienit-tasktracker-Kalyan.out
orienit@Kalyan:~$ jps
6891 SecondaryNameNode
7135 TaskTracker
6588 NameNode
6726 DataNode
6987 JobTracker
7236 Jps
orienit@Kalyan:~$
```

When we enter "start-all.sh" then all process Names are run, if we want see that process name just enter "jps" command.

**HDFS ProcessNames** : NameNode  
: SecondaryNameNode  
: DataNode  
**MapReduceProcessNames** : JobTracker  
: TaskTracker

**Note:** ProcessName order is not a mandatory.

17. Open the browser with url is (http://localhost:50070) and then click on **browse the file system** link to access hadoop data.

Contents of directory /

Goto : / go

| Name                  | Type | Size | Replication | Block Size | Modification Time | Permission | Owner  | Group      |
|-----------------------|------|------|-------------|------------|-------------------|------------|--------|------------|
| <a href="#">hbase</a> | dir  |      |             |            | 2016-09-12 14:50  | rw-r--r--  | hadoop | supergroup |
| <a href="#">home</a>  | dir  |      |             |            | 2016-09-12 14:35  | rw-r--r--  | hadoop | supergroup |
| <a href="#">tmp</a>   | dir  |      |             |            | 2016-09-12 14:45  | rw-r--r--  | hadoop | supergroup |

[Go back to DFS home](#)

**Local logs**

[Log directory](#)

[Hadoop, 2016.](#)