Step 1 - Update the package

sudo apt-get update

Step 2 - Install SSH server only if not installed during Ubuntu Server 12.04 installation

sudo apt-get install openssh-server

Step 3 - Test ssh

ssh localhost

Step 4 - Create a password-less keys in ssh

ssh-keygen –t rsa -P “”

cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys

Step 5 – add python related properties

sudo apt-get install python-software-properties

Step 6 - Install Java (Oracle Jdk)

Add repository in local machine to down load the required packages

sudo add-apt-repository ppa:webupd8team/java

sudo apt-get update

sudo apt-get install oracle-java7-installer

Verify Java installation

java -version

Step 7 - Download Hadoop stable version by finding the exact mirror from www.hadoop.apache.org

wget http://www.bizdirusa.com/mirrors/apache/hadoop/common/hadoop-1.2.1/hadoop-1.2.1-bin.tar.gz

Step 8 - Extract the tar file

tar -xvzf hadoop-1.2.1-bin.tar.gz

Step 9 - Copy the extracted folder in /usr/local/hadoop

sudo cp -r hadoop-1.2.1-bin.tar.gz /usr/local/hadoop

Give full permissions to this folder

sudo chmod –R 777 /usr/local/hadoop

Step 10 - Export the parameters to bashrc

sudo vi $HOME/.bashrc

At the end of the line type the following 3 lines

export JAVA\_HOME=/usr/lib/jvm/java-7-oracle

export HADOOP\_HOME=/usr/local/hadoop

export PATH=$PATH:$HADOOP\_HOME/bin

Update Bash

Exec bash

Step 11 - We use hadoop.tmp.dir as the base temporary directory

sudo mkdir -p /app/hadoop/tmp

sudo chmod –R 777 /app/hadoop/tmp

Step 12 - Configure hadoop environment settings

sudo vi /usr/local/hadoop/conf/hadoop-env.sh

Uncomment the following lines and add path wherever necessary

export JAVA\_HOME=/usr/lib/jvm/java-7-oracle

Step13 - Update core-site.xml

sudo vi /usr/local/hadoop/conf/core-site.xml

<property>

<name>fs.default.name</name>

<value>hdfs://localhost:54310</value>

</property>

<property>

<name>hadoop.tmp.dir</name>

<value>/app/hadoop/tmp</value>

</property>

Step 14 - Update mapred-site.xml

sudo vi /usr/local/hadoop/conf/mapred-site.xml

<property>

<name>mapred.job.tracker</name>

<value> localhost:54311</value>

</property>

Step 15 - Update HDFS-site.xml

<property>

<name>dfs.replication</name>

<value>1</value>

<description>Default block replication.

The actual number of replications can be specified when the file is created.

The default is used if replication is not specified in create time.

</description>

</property>

Step 16 - Format the hadoop name node

hadoop namenode -format

Step 17 - Start all the services

start-all.sh

Step 18 - Check for services are up or not

jps

Step19 – stop all services

Stop-all.sh