**HADOOP 2.7.2 INSTALLATION**

**Step 1->extract the tar file of hadoop-2.7.2 by the following command and rename it to hadoop**

*sudo tar -xvf hadoop-2.7.2.tar.gz*

**Step 2->Now we move 'hadoop' folder to /usr/local**

*sudo mv hadoop-2.7.2 hadoop*

*sudo mv hadoop /usr/local/hadoop*

**Step 3-> give the necessary permissions to the /usr/local/ directory by the following command**

*sudo chmod 777 /usr/local/hadoop*

*sudo chown -R bdcoe /usr/local/hadoop*

**Step 4->Now install the java-8-oracle from the following commands**

1. *sudo add-apt-repository ppa:webupd8team/java*
2. *sudo apt-get update*
3. *sudo apt-get install oracle-java8-installer*

**step 5->Install Open SSH Server and generate key by the following command**

*1.sudo apt-get install openssh-server*

*2.ssh-keygen -t rsa -P ""*

**Step 6->Enable SSH access to your local machine with this newly created key by the following command.**

*$ cat $HOME/.ssh/id\_rsa.pub >> $HOME/.ssh/authorized\_keys*

*$ ssh bdcoe@localhost*

*$ exit*

**Step-> Create log file in hadoop folder**

*sudo mkdir –p /usr/local/hadoop/logs*

**Step-> Give permissions to logs file in Hadoop**

*sudo chown -R bdcoe /usr/local/hadoop/logs*

*sudo chmod 777 /usr/local/hadoop/logs*

**Step 7-> update .bashrc**

export JAVA\_HOME=/usr/lib/jvm/java-8-oracle

export HADOOP\_PREFIX=/usr/local/hadoop

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

export PATH=$PATH:$HADOOP\_PREFIX/sbin

export HADOOP\_YARN\_HOME=$HADOOP\_PREFIX

**Step 8-> Reload ~/.bashrc file:**

source .bashrc

**Step 9-> Create HDFS directories and give necessary permissions:**

sudo mkdir -p /hadoop\_data/namenode

sudo mkdir -p /hadoop\_data/datanode

sudo chown -R bdcoe /hadoop\_data/\*

**Step10-> Modify JAVA\_HOME in /usr/local/hadoop/etc/hadoop/hadoop-env.sh:**

export JAVA\_HOME=/usr/lib/jvm/java-8-oracle

**Step 12->Modify /usr/local/hadoop/etc/hadoop/core-site.xml to have something like:**

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

</property>

</configuration>

**Step-> 13 Modify /usr/local/hadoop/etc/hadoop/yarn-site.xml to have something like:**

<configuration>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

</configuration>

**Step->14 Create /usr/local/hadoop/etc/hadoop/mapred-site.xml from template:**

cp /usr/local/hadoop/etc/hadoop/mapred-site.xml.template /usr/local/hadoop/etc/hadoop/mapred-site.xml

**Step->15 Modify /usr/local/hadoop/etc/hadoop/mapred-site.xml to have something like:**

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>

**Step->16 Modify /usr/local/hadoop/etc/hadoop/hdfs-site.xml to have something like:**

<configuration>

<property>

<name>dfs.namenode.name.dir</name>

<value>file:/hadoop\_data/namenode</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>file:/hadoop\_data/datanode</value>

</property>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

</configuration>

**Step 17->Format file system: $HADOOP\_PREFIX/bin**

hdfs namenode -format

**Step->Start Hadoop:**

start-dfs.sh

start-yarn.sh