Hadoop Installation

Java is the primary requirement for run hadoop on any system, so make sure you have Java installed on your system using following command.

```
# java -version
```

If you don't have Java installed on your system, follow the step to install java.

Java download

```
# cd /opt/
```

```
# tar xzf jdk-8u91-linux-x64.tar.gz
```

Download tar file from google ORACLE official site (8.91 version)

After extracting Java archive file, we just need to set up to use newer version of Java using alternatives. Use the following commands to do it.

```
# cd /opt/jdk1.8.0_91/
```

```
# sudo update alternatives --install /usr/bin/java java /opt/jdk1.8.0_91/bin/java 2
```

```
# sudo update alternatives --config java
```

It will provide the list of java version available. Select the latest version and press enter.

Now you may also require to set up javac and jar commands path using alternatives command.

```
# sudo update alternatives --install /usr/bin/jar jar /opt/jdk1.8.0_91/bin/jar 2
```

```
# sudo update alternatives --install /usr/bin/javac javac /opt/jdk1.8.0_91/bin/javac 2

# sudo update alternatives --set jar /opt/jdk1.8.0_91/bin/jar

# alternatives --set javac /opt/jdk1.8.0_91/bin/javac
```

Check again the Java installation using this command

```
# java -version

java version "1.8.0_60"

Java(TM) SE Runtime Environment (build 1.8.0_60-b27)

Java HotSpot(TM) 64-Bit Server VM (build 25.60-b23, mixed mode)
```

Configuring Environment Variables for Java

Most of Java based application's uses environment variables to work. Use following commands to set up these variable properly. It's also good to add following commands to any start-up script like ~/.bashrc or ~/.bash_profile.

Setup JAVA_HOME Variable

```
# export JAVA_HOME=/opt/jdk1.8.0_91
```

Setup JRE_HOME Variable

```
# export JRE_HOME=/opt/jdk1.8.0_91/jre
```

Setup PATH Variable

```
# export PATH=$PATH:/opt/jdk1.8.0_91/bin:/opt/jdk1.8.0_91/jre/bin
```

Once java installation completed, you can start hadoop installation using following steps

I recommend to create a normal (nor root) account for hadoop working. So create a system account using following command.

```
#adduser hadoop  // from main user or root
# passwd hadoop
```

After that Switch to user created like Hadoop here and perform following step

After creating account, it also required to set up key based ssh to its own account. To do this use execute following commands.

```
$ ssh-keygen -t rsa
$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
$ chmod 0600 ~/.ssh/authorized_keys
```

Lets verify key based login. Below command should not ask for password but first time it will prompt for adding RSA to the list of known hosts.

```
$ ssh localhost
$ exit
```

Let's unpack the hadoop gz which I shared you in pen drive

```
$ cd ~
$ tar xzf hadoop-2.7.1.tar.gz
$ mv hadoop-2.7.1 hadoop
```

Configure Hadoop Pseudo-Distributed Mode

Setup Environment Variables

First we need to set environment variable uses by hadoop. Edit ~/.bashrc file and append following values at end of file.

```
export HADOOP_HOME=/home/hadoop/hadoop //path where install
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
```

Now apply the changes in current running environment

```
$ source ~/.bashrc
```

Now edit $HADOOP_HOME/etc/hadoop/hadoop-env.sh file and set JAVA_HOME environment variable

```
export JAVA_HOME=/opt/jdk1.8.0_91/
```

Edit Configuration Files

Hadoop has many of configuration files, which need to configure as per requirements of your hadoop infrastructure. Lets start with the configuration with basic hadoop single node cluster setup. first navigate to below location.

Edit hadoop/etc/hadoop/core-site.xml:

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

Edit hadoop/etc/hadoop/hdfs-site.xml:

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

Edit mapred-site.xml

```
<configuration>
```

```
<property>
  <name>mapreduce.framework.name</name>
   <value>yarn</value>
</property>
</configuration>
```

Edit yarn-site.xml

```
<configuration>
<property>
  <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
</property>
</configuration>
```

Format Namenode

Now format the namenode using following command, make sure that Storage directory is

```
$ hdfs namenode -format
```

Start Hadoop Cluster

Lets start your hadoop cluster using the scripts provides by hadoop. Just navigate to your hadoop sbin directory and execute scripts one by one.

```
$ cd $HADOOP_HOME/sbin/
```

Now run start-dfs.sh script.

```
$ start-dfs.sh
```

Now run start-yarn.sh script.

```
$ start-yarn.sh
```

Go to the browser and open http://127.0.0.1:50070

Thankyou ☺