

Prerequisite Test

=====

```
sudo apt update
```

```
sudo apt install openjdk-8-jdk -y
```

```
java -version; javac -version
```

```
sudo apt install openssh-server openssh-client -y
```

```
sudo adduser hdoop
```

```
sudo adduser hdoop sudo
```

```
su - hdoop
```

```
ssh-keygen -t rsa -P "" -f ~/.ssh/id_rsa
```

```
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

```
chmod 0600 ~/.ssh/authorized_keys
```

```
ssh localhost
```

Downloading Hadoop (Please note link is updated to new version of hadoop here on 6th May 2022)

=====

```
wget https://downloads.apache.org/hadoop/common/hadoop-3.2.3/hadoop-3.2.3.tar.gz
```

```
tar xzf hadoop-3.2.3.tar.gz
```

Editng 6 important files

=====

1st file

=====

```
sudo nano .bashrc
```

#Add below lines in this file

#Hadoop Related Options

```
export HADOOP_HOME=/home/hdoop/hadoop-3.2.3
```

```
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:$HADOOP_HOME/bin
export HADOOP_OPTS"-Djava.library.path=$HADOOP_HOME/lib/nativ"
```

```
source ~/.bashrc
```

2nd File

```
=====
```

```
sudo nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh
```

#Add below line in this file in the end

```
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

3rd File

```
=====
```

```
sudo nano $HADOOP_HOME/etc/hadoop/core-site.xml
```

#Add below lines in this file(between "<configuration>" and "</configuration>")

```
<property>
  <name>hadoop.tmp.dir</name>
  <value>/home/hadoop/tmpdata</value>
  <description>A base for other temporary directories.</description>
</property>
<property>
```

```
<name>fs.default.name</name>
<value>hdfs://localhost:9000</value>
<description>The name of the default file system</description>
</property>
```

4th File

=====

```
sudo nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml
```

#Add below lines in this file(between "<configuration>" and "</configuration>")

```
<property>
  <name>dfs.data.dir</name>
  <value>/home/hdoop/dfsdata/namenode</value>
</property>
<property>
  <name>dfs.data.dir</name>
  <value>/home/hdoop/dfsdata/datanode</value>
</property>
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
```

5th File

=====

```
sudo nano $HADOOP_HOME/etc/hadoop/mapred-site.xml
```

#Add below lines in this file(between "<configuration>" and "</configuration>")

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

6th File

=====

sudo nano \$HADOOP_HOME/etc/hadoop/yarn-site.xml

#Add below lines in this file(between "<configuration>" and "</configuration>")

```
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
<property>
  <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
  <value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>127.0.0.1</value>
</property>
<property>
  <name>yarn.acl.enable</name>
  <value>0</value>
</property>
<property>
```

```
<name>yarn.nodemanager.env-whitelist</name>
```

```
<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
```

```
</property>
```

Launching Hadoop

=====

hdfs namenode -format

cd \$HADOOP_HOME/sbin

start-all.sh

go to browser -> localhost:8088

localhost:9870