Exp 1

```
sudo apt update
```

```
Su
```

```
password *******
```

```
visudo -f /etc/sudoers
```

Paste it👇

```
root ALL=(ALL:ALL) ALL
USERNAME ALL=(ALL) ALL
```

Control s and control x

```
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
```

```
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
```

```
wget https://dlcdn.apache.org/hadoop/common/hadoop-3.3.5/hadoop-3.3.5.tar.gz
```

```
tar xzf hadoop-3.3.5.tar.gz
```

```
su
```

```
Password *******
```

```
visudo -f /etc/sudoers
```

Paste it👇

```
root ALL=(ALL:ALL) ALL
hdoop ALL=(ALL) ALL
```

Control s and control x

```
sudo nano .bashrc
```

Paste it👇

```
#Hadoop Related Options
export HADOOP_HOME=/home/hdoop/hadoop-3.3.5
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"
```

Control s and control x

```
source ~/.bashrc

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

Paste it👇

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

Control s and control x

```
which javac

readlink -f /usr/bin/javac

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

Paste it👇

```
<configuration>
<property>
  <name>hadoop.tmp.dir</name>
  <value>/home/hdoop/tmpdata</value>
</property>
<property>
  <name>fs.default.name</name>
  <value>hdfs://127.0.0.1:9000</value>
</property>
</configuration>
```

Control s and control x

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

Paste it👇

```
<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>
</configuration>
```

Control s and control x

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

Paste it👇

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

Control s and control x

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

Paste it👇

```
<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,CLASSPA
TH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
</property>
</configuration>
```

Control s and control x

```
hdfs namenode -format
```

```
cd hadoop-3.3.5
```

```
Cd sbin
```

```
./start-dfs.sh
```

```
./start-yarn.sh
```

```
jps
```

http://localhost:9870

http://localhost:9864

http://localhost:8088

```
sbin/stop-yarn.sh
```

```
sbin/stop-dfs.sh
```

Command to start Hadoop

```
cd hadoop-3.3.5
```

```
cd sbin
```

```
./start-dfs.sh
```

```
./start-yarn.sh
```

Exp 2

Start Hadoop

Directory create

```
hdfs dfs -mkdir /new

hdfs dfs -ls /

hdfs dfs -mkdir /new/newone

hdfs dfs -ls /new

hdfs dfs -mkdir -p /test/test1/test2

hdfs dfs -ls /test
```

Directory delete

```
hdfs dfs -mkdir -p /one/two/three
hdfs dfs -ls /one/two
hdfs dfs -rmdir /one/two/three
hdfs dfs -ls /one/two

hdfs dfs -rm -R /one
```

Copy file

```
hdfs dfs -ls /
```

Create txt file

```
ls -ltra

hdfs dfs -mkdir /new

hdfs dfs -copyFromLocal test.txt /newone

hdfs dfs -ls /newone
```

Or

```
hdfs dfs -put test.txt /newone
hdfs dfs -ls /newone

hdfs dfs -rm /newone/*
```

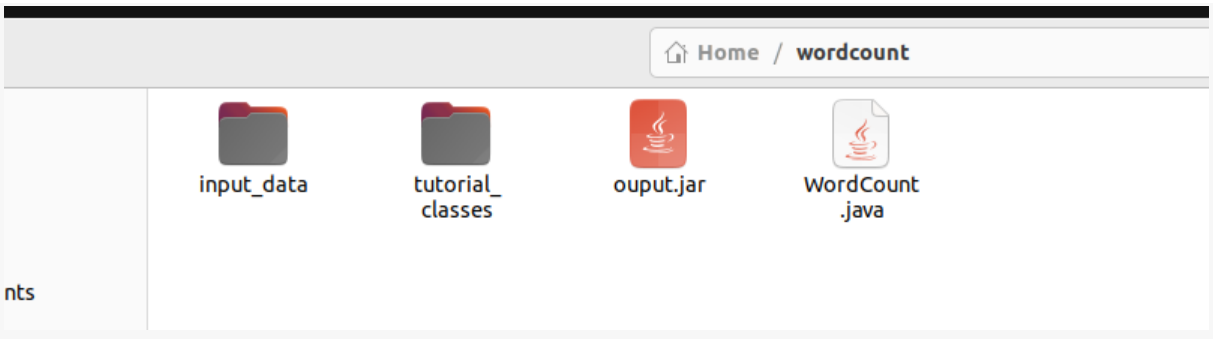
EXP 3

Create folder word count

Download file in folder
https://www.dropbox.com/s/yp9i7nwmgzr3nkx/WordCount.java?dl=0

Create folder input_data, tutorial_class inside word count folder

Create input.txt file inside input_data (words file)




```
hadoop verison

export HADOOP_CLASSPATH=$(hadoop classpath)

echo $HADOOP_CLASSPATH
```

Start Hadoop

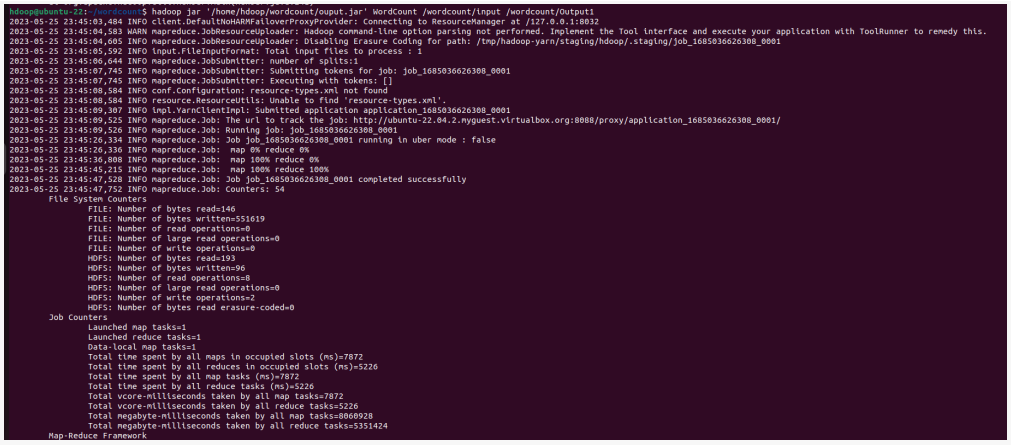
```
cd hadoop-3.3.5

cd sbin

./start-dfs.sh

./start-yarn.sh


hadoop fs -mkdir /wordcount
hadoop fs -mkdir /wordcount/input
hadoop fs -put '/home/hdoop/wordcount/input_data/input.txt' /wordcount/input

cd

ls
```

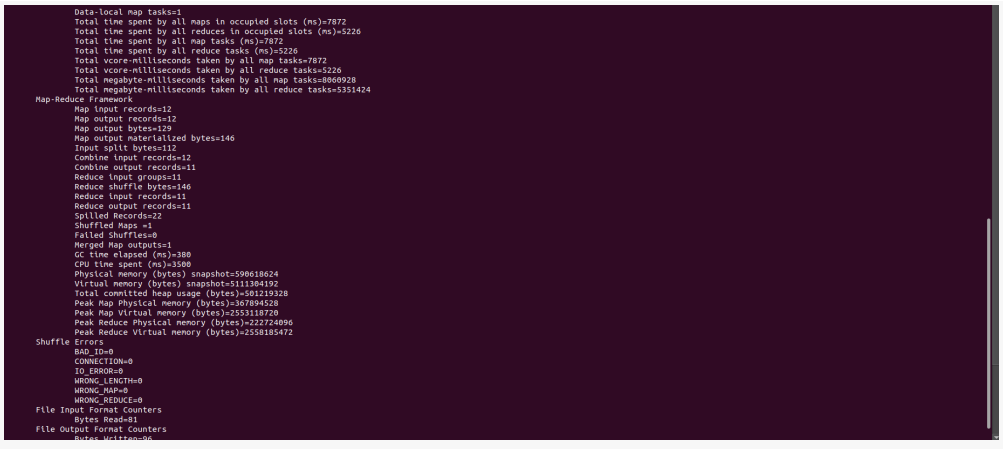
```
cd worcount/
```

```
javac -classpath ${HADOOP_CLASSPATH) -d
'/home/hdoop/wordcount/tutorial_classes'
'/home/hdoop/wordcount/WordCount.java'
```

```
jar -cvf ouput.jar -C tutorial_classes/ .
hadoop jar '/home/hdoop/wordcount/ouput.jar' WordCount /wordcount/input
/wordcount/Output1
```

```
hadoop dfs -cat /wordcount/Output1/*
```





EXP 4

```
wget https://dlcdn.apache.org/hive/hive-3.1.3/apache-hive-3.1.3-bin.tar.gz
tar xzf apache-hive-3.1.3-bin.tar.gz

sudo nano .bashrc
```

Paste it👇

```
export HIVE_HOME= "home/hdoop/apache-hive-3.1.3-bin"
export PATH=$PATH:$HIVE_HOME/bin
```

Control s and control x

```
source ~/.bashrc

sudo nano $HIVE_HOME/bin/hive-config.sh
```

Paste it👇

```
export HADOOP_HOME=/home/hdoop/hadoop-3.1.3
```

Control s and control x

START HADOOP

```
hdfs dfs -mkdir /tmp

hdfs dfs -chmod g+w /tmp

hdfs dfs -ls /

hdfs dfs -mkdir -p /user/hive/warehouse

hdfs dfs -chmod g+w /user/hive/warehouse

hdfs dfs -ls /user/hive

cd $HIVE_HOME/conf

cp hive-default.xml.template hive-site.xml

sudo nano hive-site.xml
```

Put the following at the beginning of hive-site.xml
Paste it👇

```
  <property>
    <name>system:java.io.tmpdir</name>
    <value>/tmp/hive/java</value>
  </property>
  <property>
    <name>system:user.name</name>
    <value>${user.name}</value>
  </property>
```

Edit line 3224 for and transactional.

Control s and control x

```
$HIVE_HOME/bin/schematool -dbType derby -initSchema

ls $HIVE_HOME/lib

ls $HADOOP_HOME/share/hadoop/hdfs/lib

cp $HADOOP_HOME/share/hadoop/hdfs/lib/guava-27.0-jre.jar $HIVE_HOME/lib/

$HIVE_HOME/bin/schematool -dbType derby -initSchema

cd $HIVE_HOME/bin

hive
```

EXP 5

Exp 4 required

```
CREATE DATABASE <database-name>;
```

Command:

```
CREATE DATABASE student_detail;    # this will create database student_detail

SHOW DATABASES;                    # list down all the available databases
```

To Create Table in Hive

```
CREATE TABLE IF NOT EXISTS student_data(
Student_Name STRING COMMENT 'This col. Store the name of student',
Student_Rollno INT COMMENT 'This col. Stores the rollno of student',
Student_Marks FLOAT)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ',';


SHOW TABLES IN student_detail;
```

EXP 6

https://phoenixnap.com/kb/install-spark-on-windows-10

EXP 7

https://www.edureka.co/blog/apache-pig-installation