

## 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/hadoop/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/hadoop/dfsdata/namenode</value>
</property>
<property>
  <name>dfs.data.dir</name>
  <value>/home/hadoop/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,CLASSPATH_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

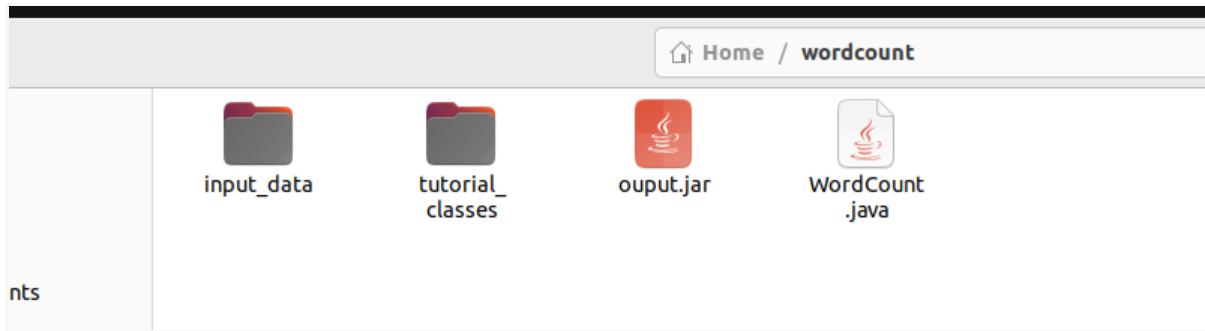
Create folder word count

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 wordcount/
```

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

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

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

```
hadoop@ubuntu-22:~/wordcount$ hadoop jar '/home/hadoop/wordcount/ouput.jar' WordCount /wordcount/input /wordcount/Output1  
2023-05-25 23:45:03,484 INFO client.DefaultHadoopFileOverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032  
2023-05-25 23:45:04,489 WARN Mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.  
2023-05-25 23:45:05,005 INFO Mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/staging/job_1685936626388_0001  
2023-05-25 23:45:05,292 INFO Input.FileInputFormat: Total input files to process : 1  
2023-05-25 23:45:06,044 INFO Mapreduce.JobSubmitter: Number of splits:1  
2023-05-25 23:45:07,745 INFO Mapreduce.JobSubmitter: Submitting tokens for job: job_1685936626388_0001  
2023-05-25 23:45:07,745 INFO Mapreduce.JobSubmitter: Executing with tokens: []  
2023-05-25 23:45:08,584 INFO conf.Configuration: resource-types.xml not found  
2023-05-25 23:45:08,584 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.  
2023-05-25 23:45:09,407 INFO InputClientImpl: Submitted application application_1685936626388_0001  
2023-05-25 23:45:09,525 INFO Mapreduce.Job: The url to track the job: http://ubuntu-22.04.2.myquest.virtualbox.org:8088/proxy/application_1685936626388_0001/  
2023-05-25 23:45:09,526 INFO Mapreduce.Job: Running job: job_1685936626388_0001  
2023-05-25 23:45:16,314 INFO Mapreduce.Job: Job job_1685936626388_0001 running in uber mode : false  
2023-05-25 23:45:26,330 INFO Mapreduce.Job: map 0% reduce 0%  
2023-05-25 23:45:26,408 INFO Mapreduce.Job: map 100% reduce 0%  
2023-05-25 23:45:45,215 INFO Mapreduce.Job: map 100% reduce 100%  
2023-05-25 23:45:47,528 INFO Mapreduce.Job: Job job_1685936626388_0001 completed successfully  
2023-05-25 23:45:47,752 INFO Mapreduce.Job: Counters: 54  
  
File System Counters  
  FILE: Number of bytes read:140  
  FILE: Number of bytes written:551619  
  FILE: Number of read operations=0  
  FILE: Number of large read operations=0  
  FILE: Number of write operations=0  
  HDFS: Number of bytes read:193  
  HDFS: Number of bytes written:46  
  HDFS: Number of read operations=8  
  HDFS: Number of large read operations=0  
  HDFS: Number of write operations=2  
  HDFS: Number of bytes read erasure-coded=0  
  
Job Counters  
  Launched map tasks=1  
    Launched reduce tasks=1  
    Data-local map tasks=1  
    Total time spent by all maps in occupied slots (ms)=7872  
    Total time spent by all reduces in occupied slots (ms)=3226  
    Total time spent by all map tasks (ms)=7872  
    Total time spent by all reduce tasks (ms)=5226  
    Total vcore-milliseconds taken by all map tasks=7872  
    Total vcore-milliseconds taken by all reduce tasks=5226  
    Total megabyte-milliseconds taken by all map tasks=8060928  
    Total megabyte-milliseconds taken by all reduce tasks=5351424  
  
Map-Reduce Framework  
  Map Input Records=12  
  Map output records=12  
  Map output bytes=129  
  Map output materialized bytes=146  
  Input split bytes=12  
  Combine input records=12  
  Combine output records=11  
  Reduce input groups=11  
  Reduce shuffle bytes=146  
  Reduce input records=11  
  Reduce output records=11  
  Spilled Records=20  
  Shuffled Maps =1  
  Failed Shuffles=0  
  Merged Map outputs=1  
  GC time elapsed (ms)=380  
  CPU time spent (ms)=3500  
  Physical memory (bytes) snapshot=599618024  
  Virtual memory (bytes) snapshot=511304192  
  Total committed heap usage (bytes)=501219328  
  Peak Map Physical memory (bytes)=367894528  
  Peak Map Virtual memory (bytes)=255318720  
  Peak Reduce Physical memory (bytes)=222724096  
  Peak Reduce Virtual memory (bytes)=2558185472  
  
Shuffle Errors  
  BAD_ID=0  
  CONNECTION=0  
  IO_ERROR=0  
  WRONG_CLUSTER=0  
  WRONG_MAP=0  
  WRONG_REDUCE=0  
  
File Input Format Counters  
  Bytes Read=81  
  File Output Format Counters  
  Bytes Written=80
```

```
  Data-local map tasks=1  
  Total time spent by all maps in occupied slots (ms)=7872  
  Total time spent by all reduces in occupied slots (ms)=5226  
  Total time spent by all map tasks (ms)=7872  
  Total time spent by all reduce tasks (ms)=5226  
  Total vcore-milliseconds taken by all map tasks=7872  
  Total vcore-milliseconds taken by all reduce tasks=5226  
  Total megabyte-milliseconds taken by all map tasks=8060928  
  Total megabyte-milliseconds taken by all reduce tasks=5351424  
  
Map-Reduce Framework  
  Map Input Records=12  
  Map output records=12  
  Map output bytes=129  
  Map output materialized bytes=146  
  Input split bytes=12  
  Combine input records=12  
  Combine output records=11  
  Reduce input groups=11  
  Reduce shuffle bytes=146  
  Reduce input records=11  
  Reduce output records=11  
  Spilled Records=20  
  Shuffled Maps =1  
  Failed Shuffles=0  
  Merged Map outputs=1  
  GC time elapsed (ms)=380  
  CPU time spent (ms)=3500  
  Physical memory (bytes) snapshot=599618024  
  Virtual memory (bytes) snapshot=511304192  
  Total committed heap usage (bytes)=501219328  
  Peak Map Physical memory (bytes)=367894528  
  Peak Map Virtual memory (bytes)=255318720  
  Peak Reduce Physical memory (bytes)=222724096  
  Peak Reduce Virtual memory (bytes)=2558185472  
  
Shuffle Errors  
  BAD_ID=0  
  CONNECTION=0  
  IO_ERROR=0  
  WRONG_CLUSTER=0  
  WRONG_MAP=0  
  WRONG_REDUCE=0  
  
File Input Format Counters  
  Bytes Read=81  
  File Output Format Counters  
  Bytes Written=80
```

```
hadoop@ubuntu-22:~/wordcount$ hadoop dfs -cat /wordcount/Output1/*  
WARNING: Use of this script to execute dfs is deprecated.  
WARNING: Attempting to execute replacement "hdfs dfs" instead.
```

```
baluffd 1  
fgsgg 1  
isata 1  
jerusalem 1  
master 1  
moham 1  
pari 1  
pratik 1  
prosola 1  
ram 1  
sheeha 2  
hadoop@ubuntu-22:~/wordcount$ S
```



## 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/hadoop/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/hadoop/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>