Following are the steps for the execution of Practical 11---

i) Create a folder on the desktop named **WordCountTutorial**

ii) Inside the **WordCountTutorial** folder create a folder named **tutorial_classes**

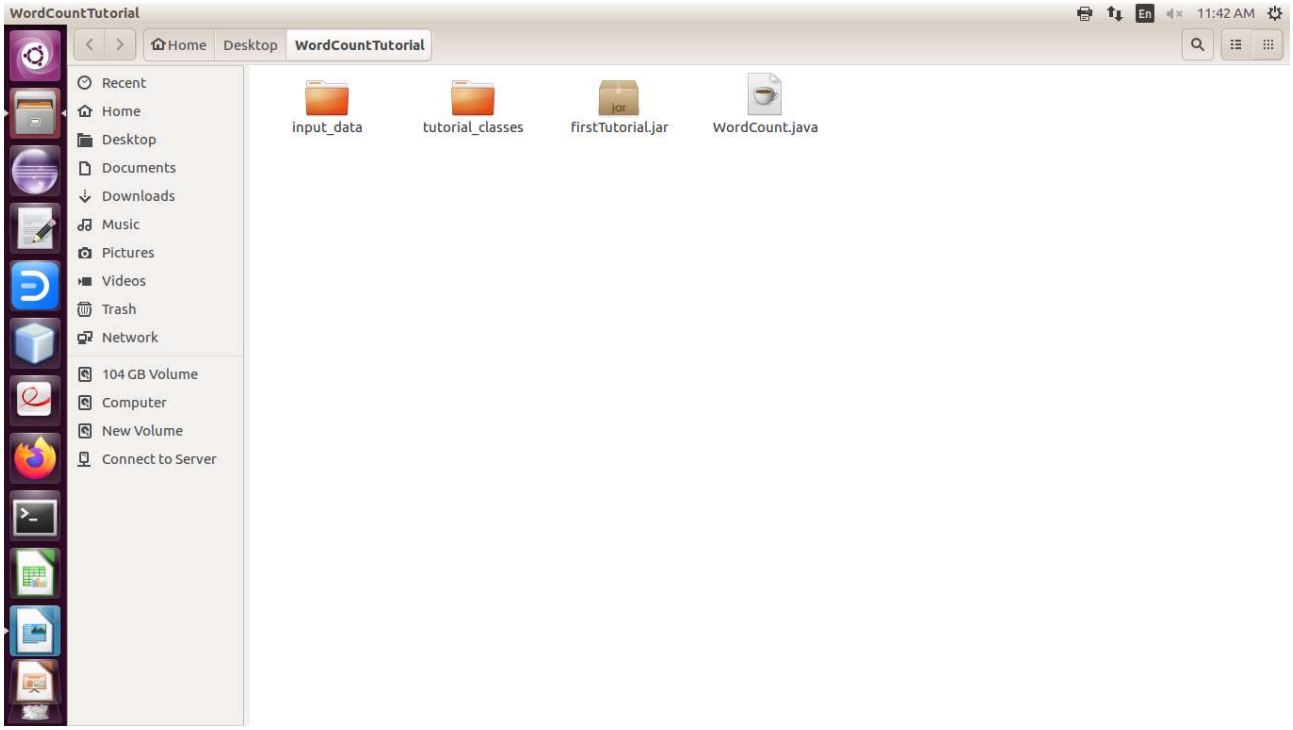
iii) Inside the **WordCountTutorial** folder create a folder named **input_data**

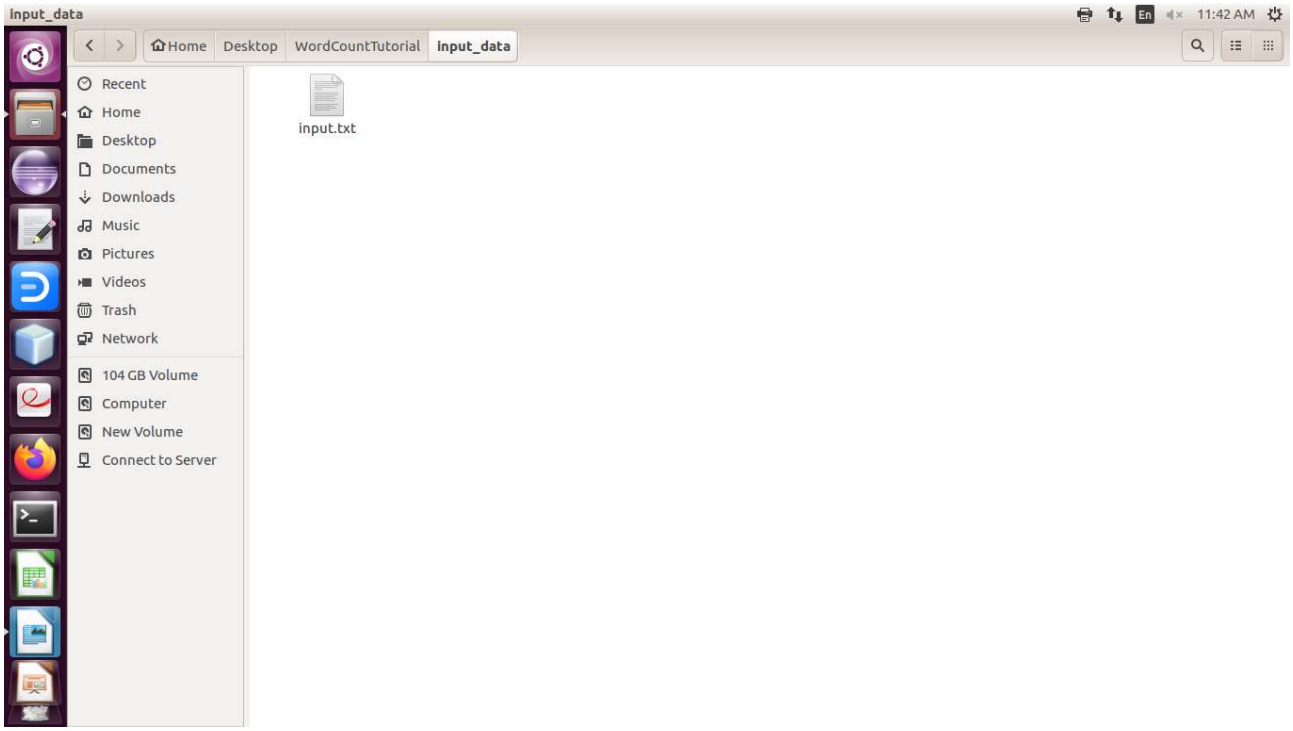
iv) Inside the **input_data** folder, create a **input.txt** file

v) Type some words which are repeated once or twice

vi) Close the file.

vii) Import WordCount.java file from the server.




input_data
En
11:42 AM
Home
Desktop
WordCountTutorial
input_data
Recent
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Network
104 GB Volume
Computer
New Volume
Connect to Server
input.txt

(base) student@student:~$ hadoop version
Hadoop 2.7.0
Subversion https://git-wip-us.apache.org/repos/asf/hadoop.git -r d4c8d4d4d203c934e8074b31289a28724c0842cf
Compiled by jenkins on 2015-04-10T18:40Z
Compiled with protoc 2.5.0
From source with checksum a9e90912c37a35c3195d23951fd18f
This command was run using /usr/local/hadoop/hadoop-2.7.0/share/hadoop/common/hadoop-common-2.7.0.jar

(base) student@student:~$ javac -version
javac 1.8.0_292

(base) student@student:~$ export HADOOP_CLASSPATH=$(hadoop classpath)

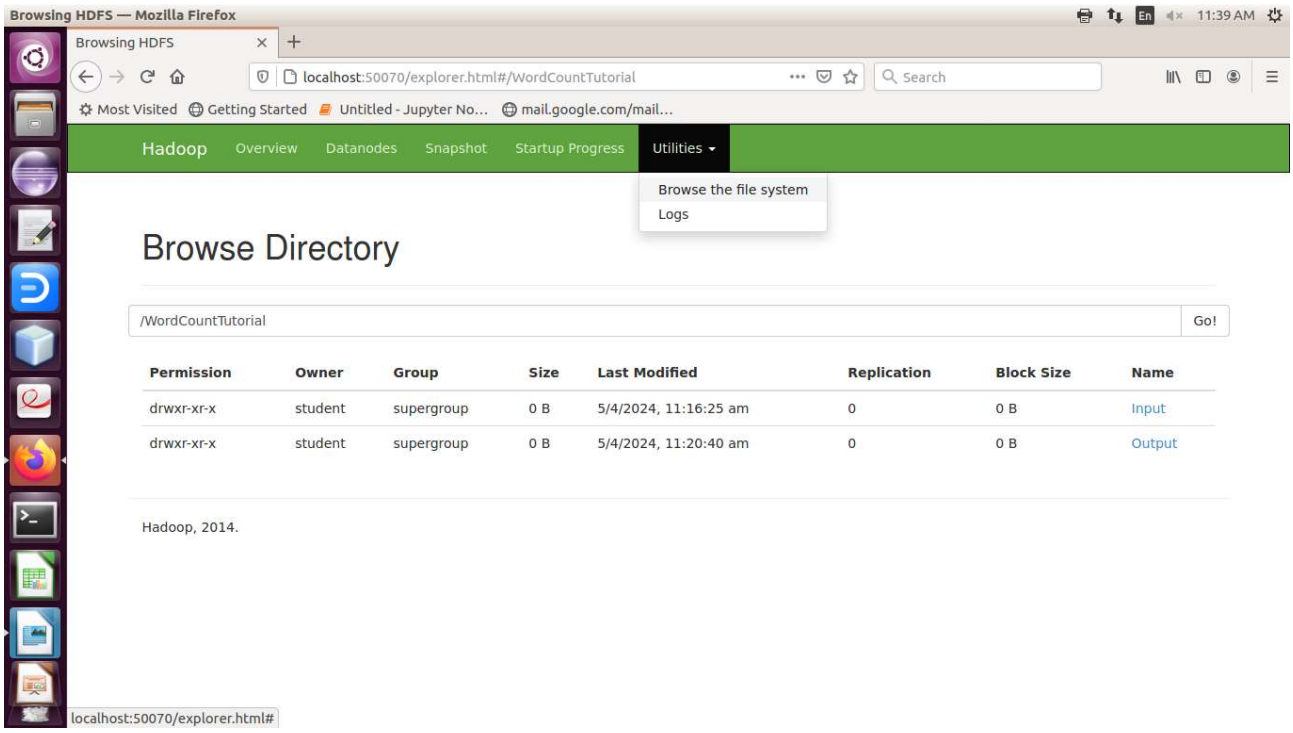
(base) student@student:~$ echo $HADOOP_CLASSPATH
/usr/local/hadoop/hadoop-2.7.0/etc/hadoop:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/common/lib/*:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/common/*:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/hdfs:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/hdfs/lib/*:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/hdfs/*:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/yarn/lib/*:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/yarn/*:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/mapreduce/lib/*:/usr/local/hadoop/hadoop-2.7.0/share/hadoop/mapreduce/*:/contrib/capacity-scheduler/*.jar

(base) student@student:~$ sudo apt-get update
[sudo] password for student:
Hit:1 http://in.archive.ubuntu.com/ubuntu xenial InRelease
Hit:2 http://in.archive.ubuntu.com/ubuntu xenial-updates InRelease

Hit:3 http://security.ubuntu.com/ubuntu xenial-security InRelease
Hit:4 http://in.archive.ubuntu.com/ubuntu xenial-backports InRelease
Reading package lists... Done

(base) student@student:~$ sudo apt-get install rsync
Reading package lists... Done
Building dependency tree
Reading state information... Done
rsync is already the newest version (3.1.1-3ubuntu1.3).
The following package was automatically installed and is no longer required:
  libqmi-glib1
Use 'sudo apt autoremove' to remove it.
0 upgraded, 0 newly installed, 0 to remove and 4 not upgraded.

Type in Web Browser "**localhost:50070"** and lcick on the Utilities from the option bar.
Click on the Browse into System Files and
Check for the WordCountTutorial folder



(base) student@student:~$ start-dfs.sh
Starting namenodes on [localhost]
student@localhost's password:
localhost: starting namenode, logging to /usr/local/hadoop/hadoop-2.7.0/logs/hadoop-student-namenode-student.out
student@localhost's password:
localhost: starting datanode, logging to /usr/local/hadoop/hadoop-2.7.0/logs/hadoop-student-datanode-student.out
Starting secondary namenodes [0.0.0.0]
student@0.0.0.0's password:

0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/hadoop-2.7.0/logs/hadoop-student-secondarynamenode-student.out

(base) student@student:~$ start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/hadoop-2.7.0/logs/yarn-student-resourcemanager-student.out
student@localhost's password:
localhost: starting nodemanager, logging to /usr/local/hadoop/hadoop-2.7.0/logs/yarn-student-nodemanager-student.out

(base) student@student:~$ hadoop fs -mkdir /WordCountTutorial

(base) student@student:~$ hadoop fs -mkdir /WordCountTutorial/Input

(base) student@student:~$ hadoop fs -put '/home/student/Desktop/WordCountTutorial/input_data/input.txt' /WordCountTutorial/Input

(base) student@student:~$ cd Desktop

(base) student@student:~/Desktop$ cd WordCountTutorial

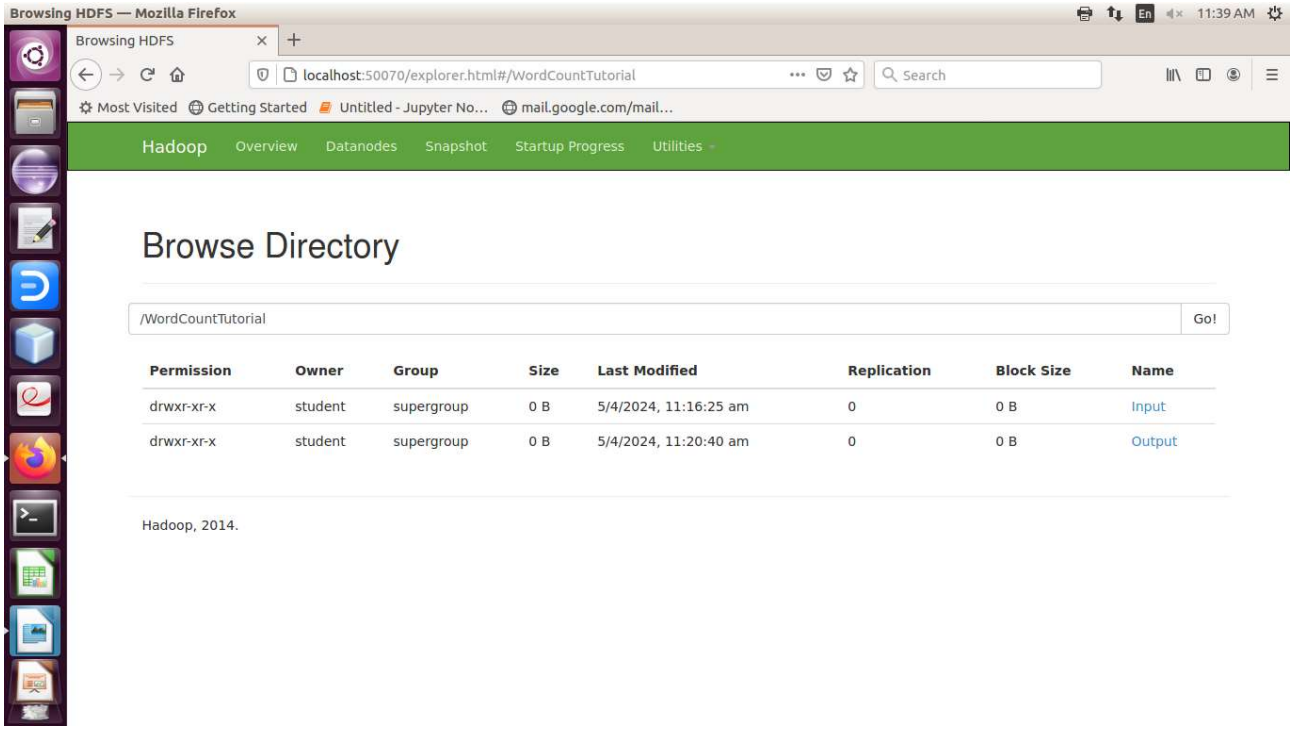
(base) student@student:~/Desktop/WordCountTutorial$ javac -classpath ${HADOOP_CLASSPATH} -d '/home/student/Desktop/WordCountTutorial/tutorial_classes' '/home/student/Desktop/WordCountTutorial/WordCount.java'

(base) student@student:~/Desktop/WordCountTutorial$ jar -cvf firstTutorial.jar -C tutorial_classes/ .
added manifest
adding: WordCount$IntSumReducer.class(in = 1739) (out= 739)(deflated 57%)
adding: WordCount$TokenizerMapper.class(in = 1736) (out= 754)(deflated 56%)
adding: WordCount.class(in = 1491) (out= 814)(deflated 45%)

(base) student@student:~/Desktop/WordCountTutorial$ hadoop jar '/home/student/Desktop/WordCountTutorial/firstTutorial.jar' WordCount /WordCountTutorial/Input /WordCountTutorial/Output
24/04/05 11:20:27 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
24/04/05 11:20:27 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
24/04/05 11:20:27 INFO input.FileInputFormat: Total input paths to process : 1
24/04/05 11:20:28 INFO mapreduce.JobSubmitter: number of splits:1
24/04/05 11:20:28 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1712295860590_0001
24/04/05 11:20:29 INFO impl.YarnClientImpl: Submitted application application_1712295860590_0001
24/04/05 11:20:29 INFO mapreduce.Job: The url to track the job: http://student:8088/proxy/application_1712295860590_0001/
24/04/05 11:20:29 INFO mapreduce.Job: Running job: job_1712295860590_0001
24/04/05 11:20:34 INFO mapreduce.Job: Job job_1712295860590_0001 running in uber mode : false

```
24/04/05 11:20:34 INFO mapreduce.Job:  map 0% reduce 0%
24/04/05 11:20:38 INFO mapreduce.Job:  map 100% reduce 0%
24/04/05 11:20:42 INFO mapreduce.Job:  map 100% reduce 100%
24/04/05 11:20:42 INFO mapreduce.Job: Job job_1712295860590_0001 completed successfully
24/04/05 11:20:42 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=115
                FILE: Number of bytes written=229509
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=222
                HDFS: Number of bytes written=77
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        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)=1765
                Total time spent by all reduces in occupied slots (ms)=1967
                Total time spent by all map tasks (ms)=1765
                Total time spent by all reduce tasks (ms)=1967
                Total vcore-seconds taken by all map tasks=1765
                Total vcore-seconds taken by all reduce tasks=1967
                Total megabyte-seconds taken by all map tasks=1807360
                Total megabyte-seconds taken by all reduce tasks=2014208
        Map-Reduce Framework
                Map input records=13
                Map output records=13
                Map output bytes=154
                Map output materialized bytes=115
                Input split bytes=120
                Combine input records=13
                Combine output records=8
                Reduce input groups=8
                Reduce shuffle bytes=115
                Reduce input records=8
                Reduce output records=8
                Spilled Records=16
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=86
                CPU time spent (ms)=930
                Physical memory (bytes) snapshot=425881600
                Virtual memory (bytes) snapshot=3832934400
                Total committed heap usage (bytes)=293076992
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
```

IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=102
File Output Format Counters
Bytes Written=77



(base) student@student:~/Desktop/WordCountTutorial$ hadoop dfs -cat /WordCountTutorial/Output/*
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

Harshvardhan 2
Himanshu 1
Sahil 2
Sandeep 3
Saqib 1
Shiva 2
Shubham 1
Yash 1