

Expt 4 - Reference File

1. To start Hadoop

hdfs namenode -format

start-dfs.sh

start-yarn.sh

jps

<http://localhost:50070/>

```
compstaff2: ~  
dbit@compstaff2:~$ jps  
6145 NameNode  
6678 ResourceManager  
7848 Jps  
6809 NodeManager  
6283 DataNode  
6508 SecondaryNameNode  
dbit@compstaff2:~$
```

Hadoop, 2018.

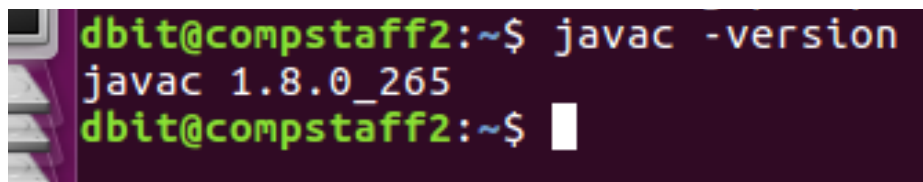
2. Make Sure Hadoop is installed and running and check hadoop version

hadoop version

```
dbit@compstaff2:~$ hadoop version  
Hadoop 2.7.7  
Subversion Unknown -r c1aad84bd27cd79c3d1a7dd58202a8c3ee1ed3ac  
Compiled by stevel on 2018-07-18T22:47Z  
Compiled with protoc 2.5.0  
From source with checksum 792e15d20b12c74bd6f19a1fb886490  
This command was run using /usr/local/hadoop/share/hadoop/common/hadoop-common-2.7.7.jar  
dbit@compstaff2:~$
```

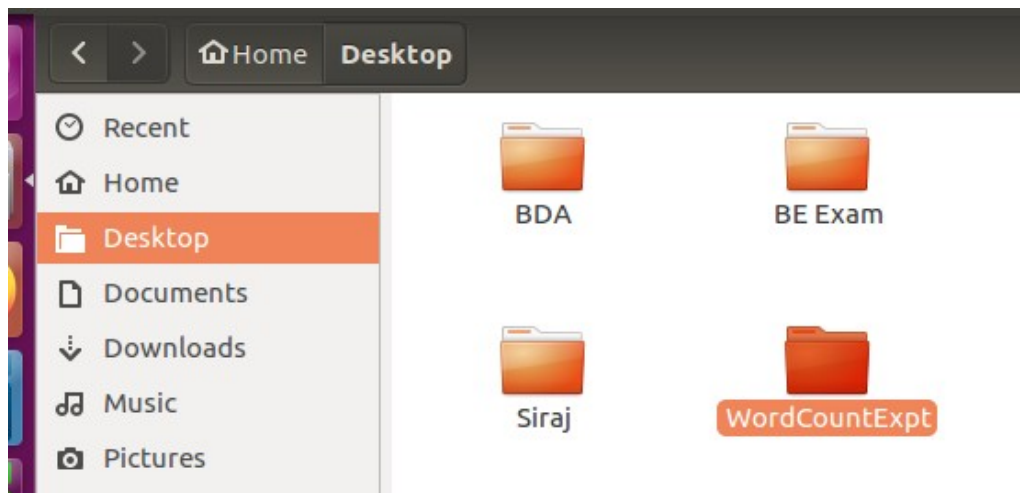
3. Make sure javac is running correctly

javac -version

A terminal window with a dark background. The prompt is 'dbit@compstaff2:~\$'. The command 'javac -version' has been entered. The output is 'javac 1.8.0_265'. The prompt is now 'dbit@compstaff2:~\$' with a cursor at the end.

```
dbit@compstaff2:~$ javac -version
javac 1.8.0_265
dbit@compstaff2:~$
```

4. Create one folder on Desktop -> WordCountExpt



5. You can download WordCount.java from {/usr/local/hadoop/share/hadoop/mapreduce/sources/hadoop-mapreduce-examples-2.7.7-sources.jar} or you can run your own version.

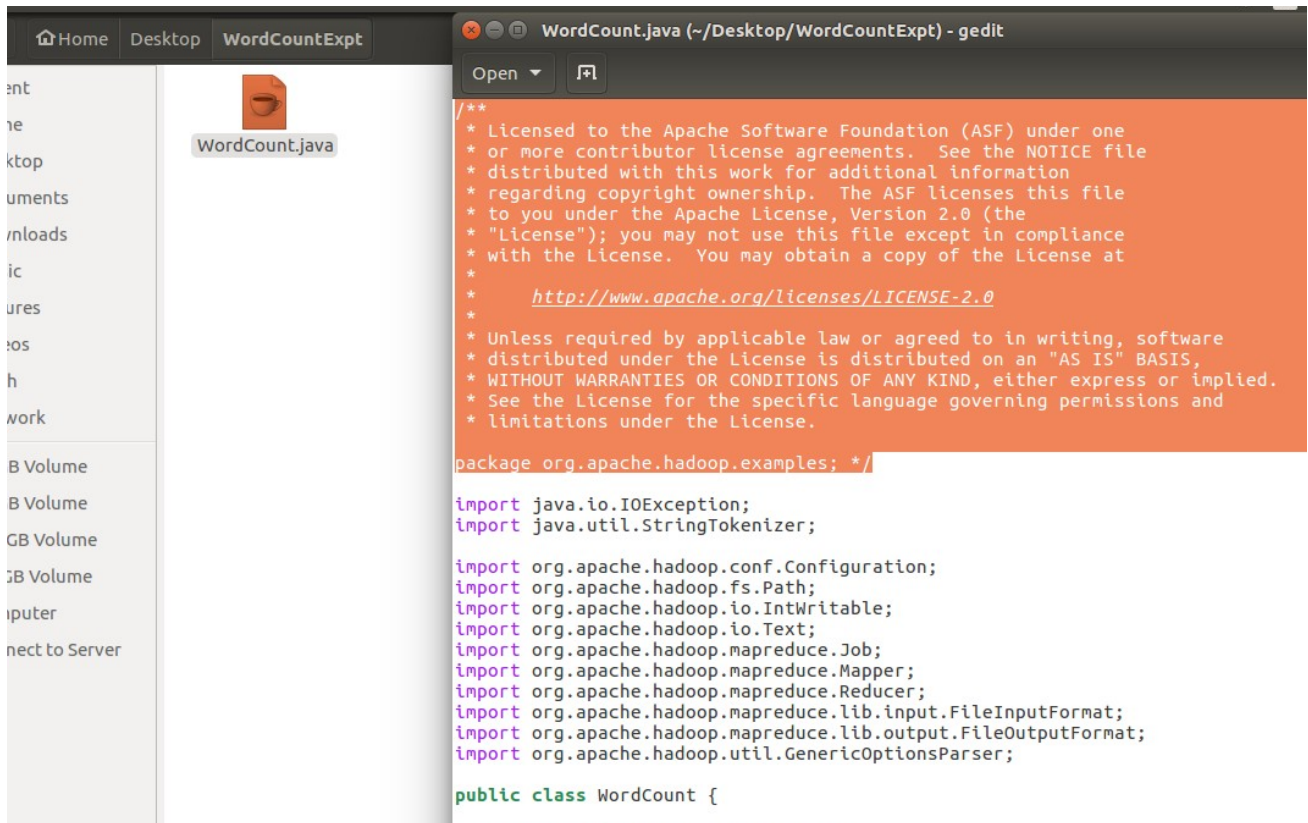
Note:

a. the above jar file copy on Desktop and Extract at Desktop

b. Copy the file WordCount.java from the path {/home/dbit/Desktop/hadoop-mapreduce-examples-2.7.7-sources/org/apache/hadoop/examples/WordCount.java} and paste it in the folder "WordCountExpt".

c. Open the file "WordCount.java" and comment the below line and save the file.

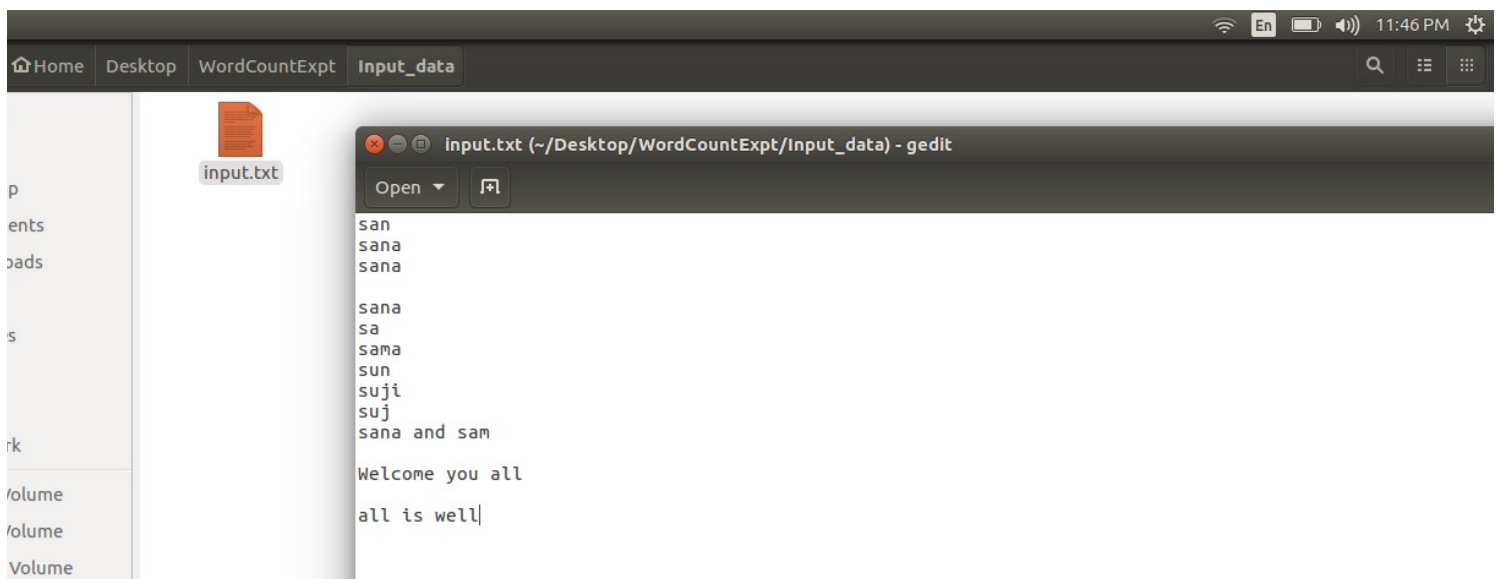
```
/* package org.apache.hadoop.examples; */
```



6. Create a folder for Input data

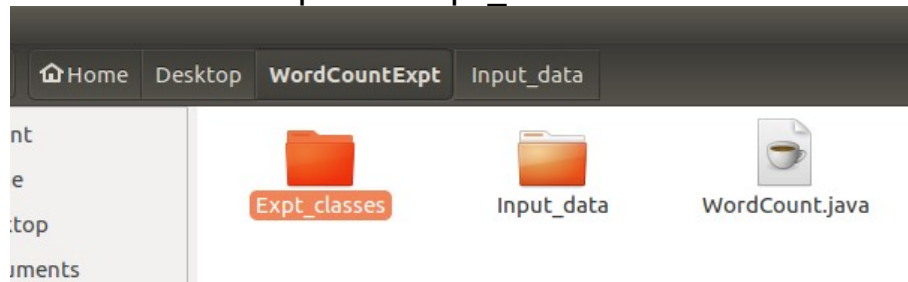
Desktop -> WordCountExpt -> Input_data

Inside Input_data folder, create an input file "input.txt" and add some text.



7. Create a new folder to hold the java class files

Desktop -> WordCountExpt -> Expt_classes



8.

```
export HADOOP_CLASSPATH=$(hadoop classpath)
echo $HADOOP_CLASSPATH
```

```
dbit@compstaff2:~$ export HADOOP_CLASSPATH=$(hadoop classpath)
dbit@compstaff2:~$ echo $HADOOP_CLASSPATH
/usr/local/hadoop/etc/hadoop:/usr/local/hadoop/share/hadoop/common/lib/*:/usr/local/hadoop/share/hadoop/hdfs/lib/*:/usr/local/hadoop/share/hadoop/hdfs/*:/usr/local/hadoop/share/hadoop/yarn/*:/usr/local/hadoop/share/hadoop/mapreduce/lib/*:/usr/local/hadoop/share/hadoop/common/lib/*:/usr/local/hadoop/share/hadoop/common/*:/usr/local/hadoop/share/hadoop/hdfs/lib/*:/usr/local/hadoop/share/hadoop/hdfs/*:/usr/local/hadoop/share/hadoop/yarn/lib/*:/usr/local/hadoop/share/hadoop/mapreduce/lib/*:/usr/local/hadoop/share/hadoop/mapreduce/*:/usr/local/hadoop/share/hadoop/common/lib/*:/usr/local/hadoop/share/hadoop/common/*:/usr/local/hadoop/share/hadoop/hdfs/*:/usr/local/hadoop/share/hadoop/hdfs/*:/usr/local/hadoop/share/hadoop/yarn/lib/*:/usr/local/hadoop/share/hadoop/yarn/*:/usr/local/hadoop/share/hadoop/mapreduce/lib/*:/usr/local/hadoop/share/hadoop/mapreduce/*:/contrib/capacity-scheduler/*.jar:/contrib/capacity-scheduler/*.jar
dbit@compstaff2:~$
```

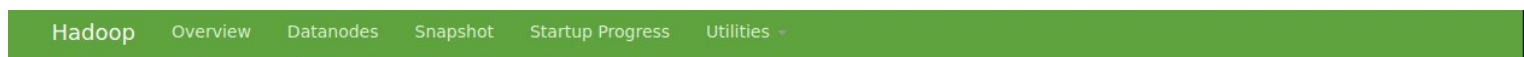
9. Create folder in HDFS:

/WordCountExpt

/WordCountExpt/Input

/WordCountExpt/Input/input.txt

```
dbit@compstaff2:~$ hadoop fs -mkdir /WordCountExpt
dbit@compstaff2:~$ hadoop fs -mkdir /WordCountExpt/Input
dbit@compstaff2:~$ hadoop fs -put /home/dbit/Desktop/WordCountExpt/Input_data/input.txt /WordCountExpt/Input
dbit@compstaff2:~$
```



Browse Directory

/WordCountExpt/Input							Go!
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	dbit	supergroup	85 B	10/28/2020, 11:54:38 PM	1	128 MB	input.txt

10. Change the current directory to the WordCount Expt directory

```
dbit@compstaff2:~$ cd /home/dbit/Desktop/WordCountExpt/  
dbit@compstaff2:~/Desktop/WordCountExpt$
```

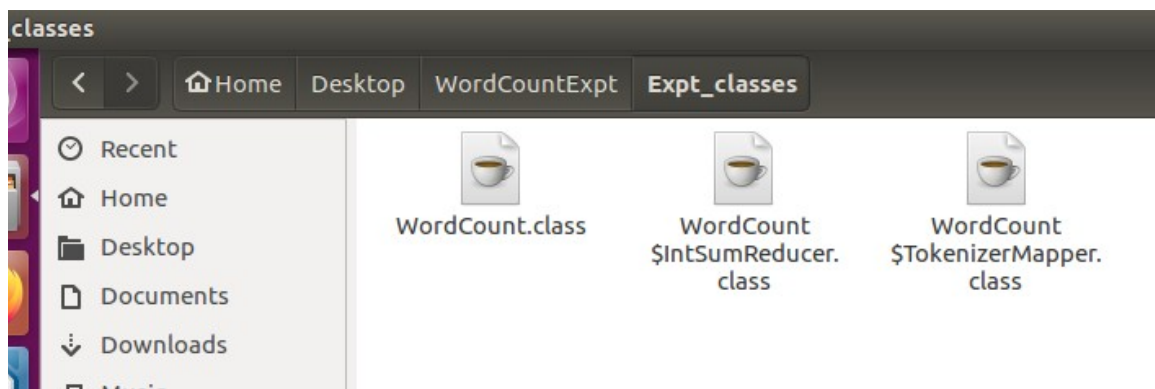
11. Compile the java code

```
javac -classpath ${HADOOP_CLASSPATH} -d <classes folder>  
<Expt Java file>
```

```
javac -classpath ${HADOOP_CLASSPATH} -d  
/home/dbit/Desktop/WordCountExpt/Expt_classes  
/home/dbit/Desktop/WordCountExpt/WordCount.java
```

```
compstaff2:~/Desktop/WordCountExpt$ javac -classpath ${HADOOP_CLASSPATH} -d '/home/dbit/Desktop/WordCountExpt/Expt_classes' '/home/dbit/Desktop/WordCountExpt/WordCount.java'  
compstaff2:~/Desktop/WordCountExpt$
```

Now check the files

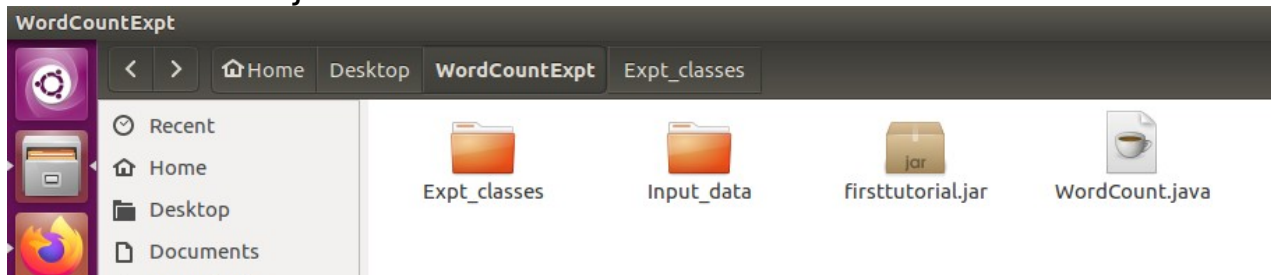


12. put the output files in one jar file :

```
jar -cvf <jar file name> -C <classes folder>
```

```
dbit@compstaff2:~/Desktop/WordCountExpt$ jar -cvf firsttutorial.jar -C /home/dbit/Desktop/WordCountExpt/Expt_classes/ .  
added manifest  
adding: WordCount.class(in = 1907) (out= 1039)(deflated 45%)  
adding: WordCount$IntSumReducer.class(in = 1739) (out= 740)(deflated 57%)  
adding: WordCount$TokenizerMapper.class(in = 1736) (out= 755)(deflated 56%)  
dbit@compstaff2:~/Desktop/WordCountExpt$
```


Now we have a jar file



13. Now run this jar file on hadoop

hadoop jar <jar file name> <class name> <hdfs Input folder>
<hdfs output folder>

hadoop jar '/home/dbit/Desktop/WordCountExpt/firstttutorial.jar'
WordCount /WordCountExpt/Input /WordCountExpt/Output

```
adding: WordCount$TokenizerMapper.class(in = 1736) (out= 755)(deflated 56%)
dbit@compstaff2:~/Desktop/WordCountExpt$ hadoop jar '/home/dbit/Desktop/WordCountExpt/firstttutorial.jar' WordCount /WordCountExpt/Input /WordCo
untExpt/Output
20/10/29 00:13:21 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
20/10/29 00:13:22 INFO input.FileInputFormat: Total input paths to process : 1
20/10/29 00:13:22 INFO mapreduce.JobSubmitter: number of splits:1
20/10/29 00:13:22 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1603905811233_0002
20/10/29 00:13:22 INFO impl.YarnClientImpl: Submitted application application_1603905811233_0002
20/10/29 00:13:23 INFO mapreduce.Job: The url to track the job: http://compstaff2:8088/proxy/application_1603905811233_0002/
20/10/29 00:13:23 INFO mapreduce.Job: Running job: job_1603905811233_0002
```

```
dbit@compstaff2: ~/Desktop/WordCountExpt
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=4991
Total time spent by all reduces in occupied slots (ms)=4382
Total time spent by all map tasks (ms)=4991
Total time spent by all reduce tasks (ms)=4382
Total vcore-milliseconds taken by all map tasks=4991
Total vcore-milliseconds taken by all reduce tasks=4382
Total megabyte-milliseconds taken by all map tasks=5110784
Total megabyte-milliseconds taken by all reduce tasks=4487168
Map-Reduce Framework
  Map input records=16
  Map output records=18
  Map output bytes=153
  Map output materialized bytes=152
  Input split bytes=116
  Combine input records=18
  Combine output records=14
  Reduce input groups=14
  Reduce shuffle bytes=152
  Reduce input records=14
  Reduce output records=14
  Spilled Records=28
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=175
  CPU time spent (ms)=1890
  Physical memory (bytes) snapshot=431947776
  Virtual memory (bytes) snapshot=3835953152
  Total committed heap usage (bytes)=295174144
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=85
File Output Format Counters
  Bytes Written=90
dbit@compstaff2:~/Desktop/WordCountExpt$
```

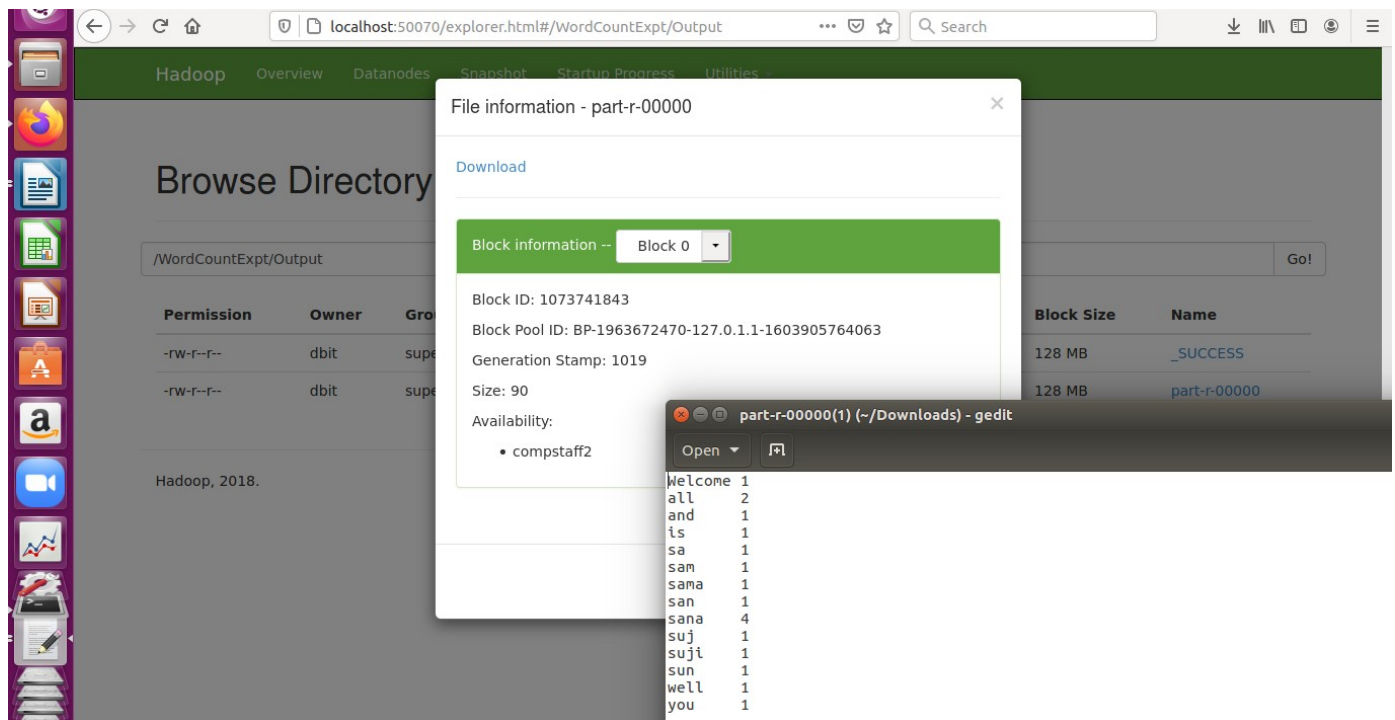
14. Check the output

`hadoop fs -cat /WordCountExpt/Output/part-r-00000`



A terminal window on a Linux system. The prompt is `dbit@compstaff2:~/Desktop/WordCountExpt$`. The command `hadoop fs -cat /WordCountExpt/Output/part-r-00000` has been executed. The output is a list of words and their counts, one per line. A 'System Settings' window is partially visible on the left side of the terminal.

```
dbit@compstaff2:~/Desktop/WordCountExpt$ hadoop fs -cat /WordCountExpt/Output/part-r-00000
Welcome 1
all 2
and 1
is 1
sa 1
sam 1
sama 1
san 1
su 1
suji 1
sun 1
well 1
you 1
dbit@compstaff2:~/Desktop/WordCountExpt$
```



The screenshot shows the Hadoop web interface in a browser. The address bar shows `localhost:50070/explorer.html#/WordCountExpt/Output`. The main content area is titled 'Browse Directory' and shows the directory `/WordCountExpt/Output`. A 'File information - part-r-00000' dialog box is open, displaying details for the file `part-r-00000`. The dialog includes a 'Download' button and a 'Block information' section. The block information shows: Block ID: 1073741843, Block Pool ID: BP-1963672470-127.0.1.1-1603905764063, Generation Stamp: 1019, Size: 90, and Availability: compstaff2. In the background, a table lists the contents of the directory, showing words and their counts. A 'gedit' window is also open, displaying the same word count data.

Block Size	Name
128 MB	_SUCCESS
128 MB	part-r-00000

```
part-r-00000(1) (~/.Downloads) - gedit
Welcome 1
all 2
and 1
is 1
sa 1
sam 1
sama 1
san 1
su 1
suji 1
sun 1
well 1
you 1
```