

LAB 02

Wordcount program in Hadoop using Cloudera platform.

Name: Muhamad Rafiq Iqbal B Samsudin

Matric: 17206926

Activity 1

Screenshot 1 – Text file

```
[cloudera@quickstart Desktop]$ cat > /home/cloudera/Processfile1.txt
Riyaz Ahamed
Riyaz Ahamed
Cloudera
Cloudera
Cloudera
Big Data
Big Data
Big Data
Cloudera
Riyaz Ahamed
^Z
[1]+  Stopped                  cat > /home/cloudera/Processfile1.txt
```

Screenshot 2 – transferred to hdfs

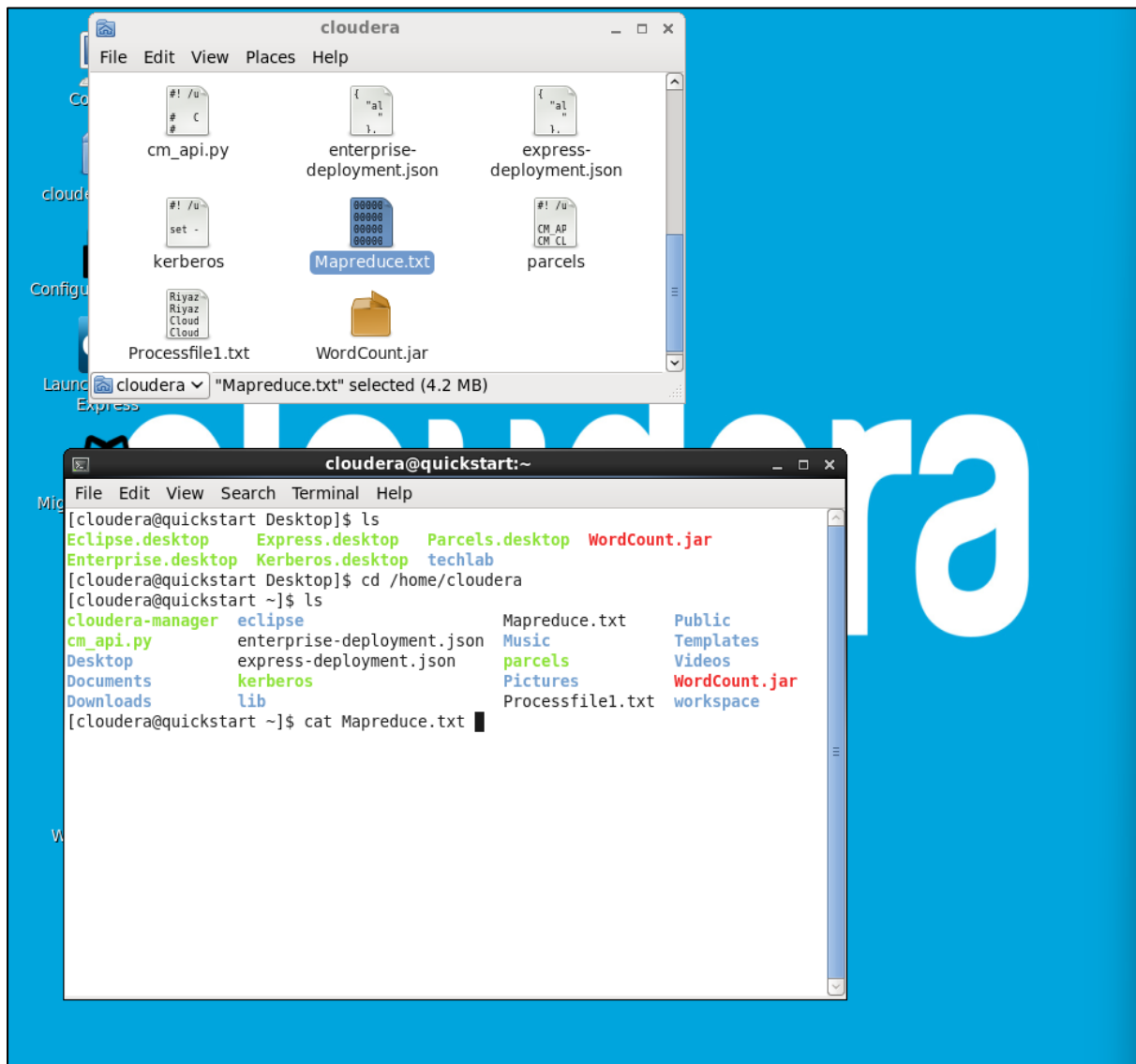
```
[cloudera@quickstart Desktop]$ hdfs dfs -mkdir /inputfolder1
[cloudera@quickstart Desktop]$ hdfs dfs -put /home/cloudera/Processfile1.txt /inputfolder1/
[cloudera@quickstart Desktop]$ hdfs dfs -cat /inputfolder1/Processfile1.txt
Riyaz Ahamed
Riyaz Ahamed
Cloudera
Cloudera
Cloudera
Big Data
Big Data
Big Data
Cloudera
```

Screenshot 3 – Output of map reduce.

```
[cloudera@quickstart Desktop]$ hdfs dfs -ls /out1
Found 2 items
-rw-r--r--  1 cloudera supergroup      0 2024-10-18 07:30 /out1/_SUCCESS
-rw-r--r--  1 cloudera supergroup    41 2024-10-18 07:30 /out1/part-r-00000
[cloudera@quickstart Desktop]$ hdfs dfs -cat /out1/part -r-00000
cat: `/out1/part': No such file or directory
cat: `-r-00000': No such file or directory
[cloudera@quickstart Desktop]$ hdfs dfs -cat /out1/part-r-00000
Ahamed  3
Big      3
Cloudera      4
Data     3
Riyaz    3
```

Activity 2

Screenshot 1 – Text file



Screenshot 2 – transferred to hdfs

```
cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
[cloudera@quickstart Desktop]$ hdfs dfs -mkdir /inputfolder2

Found 10 items
drwxrwxrwx - hdfs supergroup 0 2017-04-05 04:27 /benchmarks
drwxr-xr-x - hbase supergroup 0 2024-10-17 01:46 /hbase
drwxr-xr-x - cloudera supergroup 0 2024-10-18 07:27 /inputfolder1
drwxr-xr-x - cloudera supergroup 0 2024-10-21 00:21 /inputfolder2
drwxr-xr-x - cloudera supergroup 0 2024-10-18 07:30 /out1
drwxr-xr-x - cloudera supergroup 0 2024-10-21 00:23 /out2
drwxr-xr-x - solr solr 0 2017-04-05 04:29 /solr
drwxrwxrwt - hdfs supergroup 0 2024-10-17 01:46 /tmp
[cloudera@quickstart Desktop]$ hdfs dfs -put /home/cloudera/MApreduce.txt /inputfolder2

[cloudera@quickstart Desktop]$ hadoop jar /home/cloudera/Wordcount.jar Wordcount /input
folder2/MApreduce.txt /out2
```

Screenshot 3 – Output of map reduce.

```
[cloudera@quickstart ~]$ hdfs dfs -ls /out2
Found 2 items
-rw-r--r-- 1 cloudera supergroup 0 2024-10-21 00:23 /out2/_SUCCESS
-rw-r--r-- 1 cloudera supergroup 2789387 2024-10-21 00:23 /out2/part-r-000000
0
[cloudera@quickstart ~]$ hdfs dfs -cat /out2/part-r-000000
```

```
cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
Tables,Gresham,Oregon,credit 4
Tables,Hampton 2
Tables,Hartford,Connecticut,credit 2
Tables,Honolulu 4
Tables,Houston,Texas,cash 1
Tables,Houston,Texas,credit 3
Tables,Huntsville,Alabama,cash 1
Tables,Huntsville,Alabama,credit 2
Tables,Indianapolis 7
Tables,Irvine,California,cash 4
Tables,Irvine,California,credit 5
Tables,Irving,Texas,cash 1
Tables,Irving,Texas,credit 2
Tables,Jackson,Mississippi,credit 5
Tables,Jacksonville 11
Tables,Jersey 2
Tables,Kansas 4
Tables,Las 2
Tables,Lexington,Kentucky,credit 1
Tables,Lincoln,Nebraska,credit 2
Tables,Long 1
Tables,Los 2
Tables,Louisville,Kentucky,credit 4
Tables,Lowell,Massachusetts,cash 2
Tables,Lowell,Massachusetts,credit 1
Tables,Madison,Wisconsin,cash 1
Tables,Madison,Wisconsin,credit 11
Tables,McKinney,Texas,cash 1
Tables,McKinney,Texas,credit 1
Tables,Memphis,Tennessee,credit 4
Tables,Miami,Florida,credit 5
Tables,Midland,Texas,credit 5
Tables,Milwaukee,Wisconsin,cash 1
Tables,Milwaukee,Wisconsin,credit 1
Tables,Minneapolis,Minnesota,cash 1
Tables,Minneapolis,Minnesota,credit 2
Tables,Montgomery,Alabama,credit 1
Tables,Nashville 5
Tables,New 4
Tables,Newark,New 4
Tables,Oakland,California,cash 1
Tables,Oakland,California,credit 5
Tables,Oklahoma 4
Tables,Omaha,Nebraska,cash 2
Tables,Orange,California,credit 1
Tables,Orlando,Florida,cash 2
Tables,Orlando,Florida,credit 3
Tables,Overland 3
Tables,Pasadena,California,cash 2
Tables,Pasadena,California,credit 1
```

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
Weights,Montgomery,Alabama,credit 2  
Weights,Nashville 1  
Weights,New 9  
Weights,Newark,New 2  
Weights,Oakland,California,credit 1  
Weights,Oklahoma 4  
Weights,Omaha,Nebraska,credit 4  
Weights,Orange,California,cash 1  
Weights,Orange,California,credit 4  
Weights,Orlando,Florida,credit 4  
Weights,Overland 6  
Weights,Pasadena,California,credit 3  
Weights,Pasadena,Texas,credit 4  
Weights,Paterson,New 3  
Weights,Philadelphia,Pennsylvania,cash 1  
Weights,Philadelphia,Pennsylvania,credit 5  
Weights,Phoenix,Arizona,credit 3  
Weights,Pittsburgh,Pennsylvania,credit 5  
Weights,Plano,Texas,credit 3  
Weights,Portland,Oregon,credit 3  
Weights,Reno,Nevada,credit 4  
Weights,Richmond 3  
Weights,Rockford,Illinois,cash 1  
Weights,Rockford,Illinois,credit 3  
Weights,Sacramento,California,cash 2  
Weights,Sacramento,California,credit 3  
Weights,Saint 1  
Weights,Salem,Oregon,credit 3  
Weights,Salt 7  
Weights,San 11  
Weights,Santa 3  
Weights,Scottsdale,Arizona,credit 3  
Weights,Seattle,Washington,cash 1  
Weights,Seattle,Washington,credit 3  
Weights,Springfield,Illinois,credit 5  
Weights,St. 3  
Weights,Stamford,Connecticut,credit 1  
Weights,Sunnyvale,California,cash 1  
Weights,Tampa,Florida,cash 1  
Weights,Tampa,Florida,credit 2  
Weights,Vancouver,Washington,cash 1  
Weights,Vancouver,Washington,credit 4  
Weights,Washington,District 5  
Weights,West 6  
Weights,Westminster,Colorado,credit 3  
York,New 478  
York,cash 150  
York,credit 835  
of 479  
[cloudera@quickstart ~]$
```