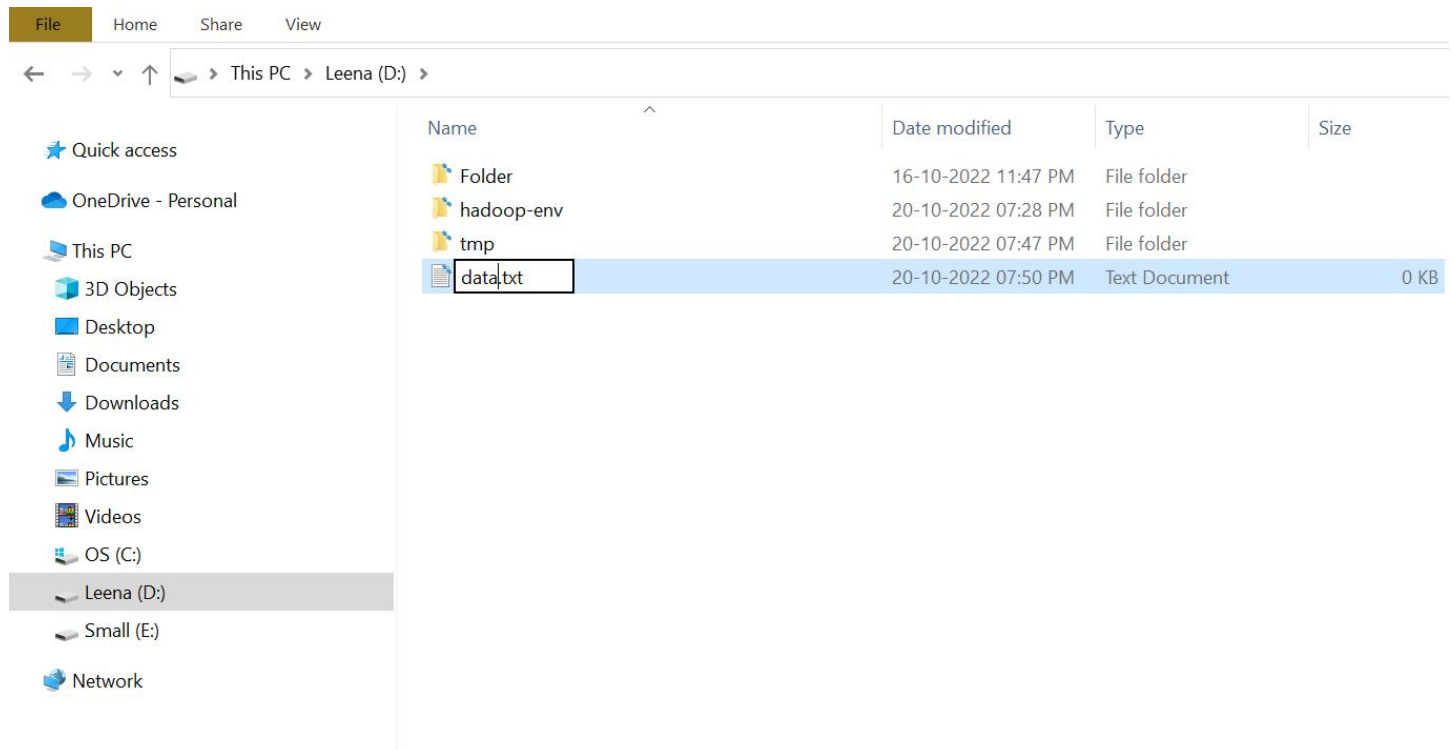


## EXERCISE: 10

### Word Count Program using MapReduce

#### Step 1: Create a text file “D:/data.txt”



**Step 2:** Create a directory in HDFS, where to kept text file.

```
hdfs dfs -mkdir /user
```

**Step 3:** Upload the data.txt file on HDFS in the specific directory

```
hdfs dfs -put D:/data.txt /user
```

**Step 4:** List the files or directories in hdfs

```
hdfs dfs -ls /user/
```

**Step 5:** To view the content of the file “/user/data.txt”

```
hdfs dfs -cat /user/data.txt
```

```

Administrator: Windows PowerShell
*****/
PS C:\WINDOWS\system32> cd D:\hadoop-env\hadoop-3.2.2\sbin
PS D:\hadoop-env\hadoop-3.2.2\sbin> .\start-dfs.cmd
PS D:\hadoop-env\hadoop-3.2.2\sbin> ./start-yarn.cmd
starting yarn daemons
PS D:\hadoop-env\hadoop-3.2.2\sbin> jps
1280 NodeManager
6496 ResourceManager
3860 Jps
6388 NameNode
312 DataNode
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -mkdir /user
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -put D:/data.txt /user
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -ls /user/
Found 1 items
-rw-r--r--  1 Leena supergroup          99 2022-10-20 19:53 /user/data.txt
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -cat /user/data.txt
Hadoop
MapReduce
Leena
Malar Sagay
Tharini
Malar Sagay
Leena
Cloud Computing
Leena
Hadoop
PS D:\hadoop-env\hadoop-3.2.2\sbin>

```

**Step 6: Run the jar file**

**hadoop jar D:/hadoop-env/hadoop-3.2.2/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.2.jar wordcount /user /out**

```

Administrator: Windows PowerShell
PS D:\hadoop-env\hadoop-3.2.2\sbin> hadoop jar D:/hadoop-env/hadoop-3.2.2/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.2.2.jar wordcount /user /out
2022-10-20 19:55:01,629 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2022-10-20 19:55:03,127 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/Leena/.staging/job_1666275438730_0001
2022-10-20 19:55:03,769 INFO input.FileInputFormat: Total input files to process : 1
2022-10-20 19:55:03,910 INFO mapreduce.JobSubmitter: number of splits:1
2022-10-20 19:55:04,182 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1666275438730_0001
2022-10-20 19:55:04,186 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-10-20 19:55:04,614 INFO conf.Configuration: resource-types.xml not found
2022-10-20 19:55:04,615 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-10-20 19:55:05,333 INFO impl.YarnClientImpl: Submitted application application_1666275438730_0001
2022-10-20 19:55:05,440 INFO mapreduce.Job: The url to track the job: http://LAPTOP-UDDUJ6SQ:8088/proxy/application_1666275438730_0001
2022-10-20 19:55:05,444 INFO mapreduce.Job: Running job: job_1666275438730_0001
2022-10-20 19:55:25,967 INFO mapreduce.Job: Job job_1666275438730_0001 running in uber mode : false
2022-10-20 19:55:25,971 INFO mapreduce.Job:  map 0% reduce 0%
2022-10-20 19:55:36,223 INFO mapreduce.Job:  map 100% reduce 0%

```

**Step 7:** To view the output in “/out/\*”

**hadoop fs -cat /out/\***

```
Administrator: Windows PowerShell

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=99
File Output Format Counters
    Bytes Written=75
PS D:\hadoop-env\hadoop-3.2.2\sbin> hadoop fs -cat /out/*
Cloud      1
Computing   1
Hadoop      2
Leena       3
Malar       2
MapReduce   1
Sagay       2
Tharini     1
PS D:\hadoop-env\hadoop-3.2.2\sbin>
```

## Output

Overview 'localhost:9820' (active)

Started:	Thu Oct 20 19:46:15 +0530 2022
Version:	3.2.2, r7a3bc90b05f257c8ace2f76d74264906f0f7a932
Compiled:	Sun Jan 03 14:56:00 +0530 2021 by hexiaoqiao from branch-3.2.2
Cluster ID:	CID-578272e3-8c70-44ca-b6ba-a4f9d4a34a71
Block Pool ID:	BP-1271832836-192.168.137.1-1666275217495

- Browse the file system
- Logs
- Log Level
- Metrics
- Configuration
- Process Thread Dump

Browse Directory

/

Go!

📁

📄

📄

Show

25

entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	drwxr-xr-x	Leena	supergroup	0 B	Oct 20 19:55	0	0 B	<a href="#">out</a>	🗑
<input type="checkbox"/>	drwx-----	Leena	supergroup	0 B	Oct 20 19:55	0	0 B	<a href="#">tmp</a>	🗑
<input type="checkbox"/>	drwxr-xr-x	Leena	supergroup	0 B	Oct 20 19:53	0	0 B	<a href="#">user</a>	🗑

Showing 1 to 3 of 3 entries

Previous

1

Next

Browsing HDFS

DataNode Information

All Applications

+

localhost:9870/explorer.html#/out

Gmail

SkillRack.com

Leena Velni | Linked...

PRIEE

GitHub

Tutorials List - Javat...

(2) PBLChat

Hadoop Architectur...

Copy of

Hadoop

Overview

Datanodes

Browse Directory

/out

Go!

📁

📄

📄

Show

25

entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	Leena	supergroup	0 B	Oct 20 19:55	0	0 B	<a href="#">out</a>	🗑
<input type="checkbox"/>	-rw-r--r--	Leena	supergroup	0 B	Oct 20 19:53	0	0 B	<a href="#">user</a>	🗑

Showing 1 to 2 of 2 entries

Previous

1

Next

Hadoop, 2021.

Close

Block ID: 1073741832

Block Pool ID: BP-1271832836-192.168.137.1-1666275217495

Generation Stamp: 1008

Size: 75

Availability:

LAPTOP-UDDUJ6SQ.mshome.net

File contents

Cloud 1

Computing 1

Hadoop 2

Leena 3

Malar 2

MapReduce 1

Sagay 2

Tharini 1

Close