

Date : 14/10/22

Name : SRINITHI B

EX:10

Roll.No:19UITE036

Reg.No:920419205042

MapReduce (WordCount Program)

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

```
Administrator: Windows PowerShell
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -mkdir /user
PS D:\hadoop-env\hadoop-3.2.2\sbin> █
```

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

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

```
Administrator: Windows PowerShell
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> █
```

Step 4: List the files or directories in hdfs

hdfs dfs -ls /user/

```
Administrator: Windows PowerShell
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 Administrator supergroup 50 2022-10-11 11:34 /user/data.txt
PS D:\hadoop-env\hadoop-3.2.2\sbin> █
```

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

hdfs dfs -cat /user/data.txt

```
Administrator: Windows PowerShell
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 Administrator supergroup 50 2022-10-11 11:34 /user/data.txt
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -cat /user/data.txt
Cloud and Grid Lab. Cloud and Grid Lab.
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

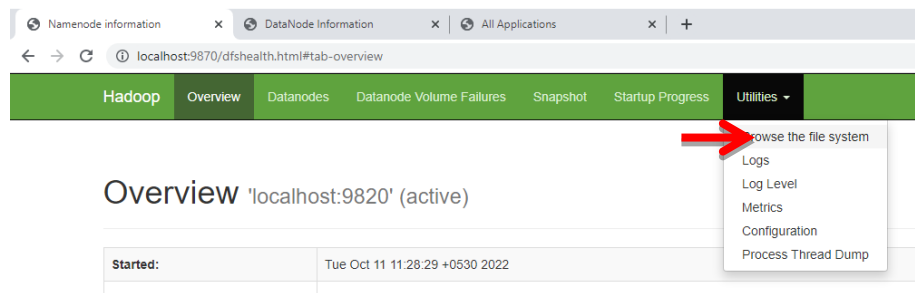
```
Administrator: Windows PowerShell
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -mkdir /user
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -put D:/data.txt /user
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -ls /user/
Found 1 items
-rw-r--r-- 1 Administrator supergroup 50 2022-10-11 11:34 /user/data.txt
PS D:\hadoop-env\hadoop-3.2.2\sbin> hdfs dfs -cat /user/data.txt
Cloud and Grid Lab. Cloud and Grid Lab.
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
2022-10-11 11:36:37,212 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2022-10-11 11:36:38,133 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/Administrator/.staging/job_1665467943708_0001
2022-10-11 11:36:38,559 INFO input.FileInputFormat: Total input files to process : 1
2022-10-11 11:36:38,864 INFO mapreduce.JobSubmitter: number of splits:1
2022-10-11 11:36:39,157 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1665467943708_0001
2022-10-11 11:36:39,159 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-10-11 11:36:39,488 INFO conf.Configuration: resource-types.xml not found
2022-10-11 11:36:39,484 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-10-11 11:36:40,208 INFO impl.YarnClientImpl: Submitted application application_1665467943708_0001
2022-10-11 11:36:40,320 INFO mapreduce.Job: The url to track the job: http://a4it55:8088/proxy/application_1665467943708_0001/
2022-10-11 11:36:40,322 INFO mapreduce.Job: Running job: job_1665467943708_0001
2022-10-11 11:36:54,677 INFO mapreduce.Job: Job job_1665467943708_0001 running in uber mode : false
2022-10-11 11:36:54,680 INFO mapreduce.Job: map 0% reduce 0%
2022-10-11 11:37:00,861 INFO mapreduce.Job: map 100% reduce 0%
2022-10-11 11:37:07,881 INFO mapreduce.Job: map 100% reduce 100%
2022-10-11 11:37:09,016 INFO mapreduce.Job: Job job_1665467943708_0001 completed successfully
2022-10-11 11:37:09,239 INFO mapreduce.Job: Counters: 54
```

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

hadoop fs -cat /out/*

```
Bytes Written=28
PS D:\hadoop-env\hadoop-3.2.2\sbin> hadoop fs -cat /out/*
Cloud 3
Grid 2
Lab. 3
and 2
PS D:\hadoop-env\hadoop-3.2.2\sbin>
```




Output



← → ↻ localhost:9870/explorer.html/#/Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities




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	Administrator	supergroup	0 B	Oct 11 11:37	0	0 B	out	
<input type="checkbox"/>	drwx-----	Administrator	supergroup	0 B	Oct 11 11:36	0	0 B	tmp	
<input type="checkbox"/>	drwxr-xr-x	Administrator	supergroup	0 B	Oct 11 11:34	0	0 B	user	

← → ↻ localhost:9870/explorer.html#/outHadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities



Browse Directory

Go!



Show 25 entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	Administrator	supergroup	0 B	Oct 11 11:37	1 <div>Replication</div>	128 MB	_SUCCESS	
<input type="checkbox"/>	-rw-r--r--	Administrator	supergroup	28 B	Oct 11 11:37	1	128 MB	part-r-00000	

Showing 1 to 2 of 2 entries

Previous 1 Next

File information - part-r-00000

Download

Head the file (first 32K)

Tail the file (last 32K)

Block information

Block 0

Block ID: 1073741832
Block Pool ID: BP-789191111-172.16.8.55-1665467846441
Generation Stamp: 1008
Size: 28
Availability:

- a4it55 KC.VNR

File contents

Cloud 3
Grid 2
Lab 3
and 2