

## MapReduce Character Counter

- 1- Download eclipse software that helps me to convert the java source to the JAR file (<https://www.eclipse.org/downloads/>)
- 2- Use the java source founding at this link: [Java source](#)
- 3- Put this source into eclipse and extract the jar file as (CharCount.jar).
- 4- Remember: it is important to know the path for your jar file.
- 5- Open the terminal
- 6- Run Hadoop: `$start -all.sh`

```
hadoop@bsmh-VirtualBox:~$ jps
2899 NameNode
3542 ResourceManager
3078 DataNode
3354 SecondaryNameNode
3693 NodeManager
11263 Jps
```

- 7- Create a new .txt file by using `nano english.txt`
- 8- Then copy this file into HDFS with: `hdfs dfs -put /english.txt`
- 9- List all files in the HDFS: `hdfs dfs -l /`

```
hadoop@bsmh-VirtualBox:~$ hdfs dfs -ls /
2022-03-28 14:27:32,218 WARN util.NativeCodeLoader: Unable to load native-hadoop
Found 17 items
-rw-r--r-- 1 hadoop supergroup 4208 2022-03-22 15:55 /Average_Age.jar
drwxr-xr-x - hadoop supergroup 0 2022-03-22 18:09 /Titanic_Output
drwxr-xr-x - hadoop supergroup 0 2022-03-22 18:26 /Titanic_Output1
drwxr-xr-x - hadoop supergroup 0 2022-03-01 18:32 /aljava
-rw-r--r-- 1 hadoop supergroup 33 2022-03-22 19:14 /arb.txt
-rw-r--r-- 1 hadoop supergroup 9 2022-03-01 19:03 /b.txt
-rw-r--r-- 1 hadoop supergroup 33 2022-03-22 19:13 /english.txt
```

- 10- Display the txt file content: `hdfs dfs -cat /english.txt`

```
hadoop@bsmh-VirtualBox:~$ hdfs dfs -cat /english.txt
2022-03-28 15:08:01,161 WARN util.NativeCodeLoader: Unable to load
Foz 90
Mariam 50
Sara 87
Basma 70hadoop@bsmh-VirtualBox:~$
```

- 11- Start now with MapReduce job, first, enter to the jar file path: `cd /home/bsmh`

```
hadoop@bsmh-VirtualBox:~$ cd /home//bsmh
hadoop@bsmh-VirtualBox:/home/bsmh$ ls
Average_Age.jar CharCount.jar Desktop Documents Downloads Music Pictures Public score
.jar snap Templates titanic_data1.txt titanic_data.txt Videos
```

- 12- Run mapreduce job with: `hadoop jar CharCount.jar /english.txt /CharCountResult`

```

hadoop@bsmh-VirtualBox:~/home/bsmh$ hadoop jar CharCount.jar /english.txt /CharCountResult
2022-03-28 14:46:29,054 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2022-03-28 14:46:29,465 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2022-03-28 14:46:29,576 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2022-03-28 14:46:29,713 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2022-03-28 14:46:29,751 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1647975056431_0008
2022-03-28 14:46:29,911 INFO mapred.FileInputFormat: Total input files to process : 1
2022-03-28 14:46:29,975 INFO mapreduce.JobSubmitter: number of splits:2
2022-03-28 14:46:30,095 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1647975056431_0008
2022-03-28 14:46:30,095 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-03-28 14:46:30,252 INFO conf.Configuration: resource-types.xml not found
2022-03-28 14:46:30,252 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-03-28 14:46:30,433 INFO impl.YarnClientImpl: Submitted application application_1647975056431_0008
2022-03-28 14:46:30,476 INFO mapreduce.Job: The url to track the job: http://bsmh-VirtualBox:8088/proxy/application_1647975056431_0008/
2022-03-28 14:46:30,481 INFO mapreduce.Job: Running job: job_1647975056431_0008
2022-03-28 14:46:36,574 INFO mapreduce.Job: Job job_1647975056431_0008 running in uber mode : false
2022-03-28 14:46:36,576 INFO mapreduce.Job: map 0% reduce 0%
2022-03-28 14:46:41,653 INFO mapreduce.Job: map 100% reduce 0%
2022-03-28 14:46:44,685 INFO mapreduce.Job: map 100% reduce 100%
2022-03-28 14:46:45,735 INFO mapreduce.Job: Job job_1647975056431_0008 completed successfully
2022-03-28 14:46:45,804 INFO mapreduce.Job: Counters: 54
File System Counters:

```

Successfully Run...

- 13- Display the output folder: **hdfs dfs -ls /CharCountResult**

```

hadoop@bsmh-VirtualBox:~$ hdfs dfs -ls /CharCountResult
2022-03-28 14:47:39,377 WARN util.NativeCodeLoader: Unable to load native-hadoop library here applicable
Found 2 items
-rw-r--r-- 1 hadoop supergroup 0 2022-03-28 14:46 /CharCountResult/_SUCCESS
-rw-r--r-- 1 hadoop supergroup 68 2022-03-28 14:46 /CharCountResult/part-00000

```

- 14- Finally open the output file: **hdfs dfs -cat /CharCountResult/part\***

```

hadoop@bsmh-VirtualBox:~$ hdfs dfs -cat /CharCountResult/part*
2022-03-28 14:48:10,384 WARN util.NativeCodeLoader: Unable to load here applicable
4
0 3
5 1
7 2
8 1
9 1
B 1
F 1
M 1
S 1
a 6
i 1
m 2
o 1
r 2
s 1
z 1

```