**Aim : To Implement CharCount problem using Hadoop MapReduce in Eclipse.**

Step 1 : Run your cloudera system. Open Eclipse.

Step 2 : Click on File > New > java project. Give Project Name (“CharCount”).

Step 3 : Click on Libraries tab. Then click on Add External JARs… Tab To add Hadoop Libraries.

Step 4 : Follow this steps :

Click on File System -> usr -> lib -> hadoop (Select all the libraries (JAR files)) -> Click OK.

Again Click on Add External JARs… -> client -> select all jar files -> ok -> Finish.

Step 5 : Now you will see project name “CharCount” on sidebar.

Right Click on Project name “CharCount” -> New -> class.

Create 3 new classes named :

**CharCountDriver**(having the main function)**,**

**CharCountMapper,**

**CharCountReducer.**

****

Step 6 : Write your code in java windows.

Step 7 : Right Click on the project name

CharCount -> Export -> Java -> JAR File -> Next -> For select the export destination for JAR file:

browse -> Name : CharCount.jar -> save in folder -> cloudera -> Finish -> OK



Step 8 : Now open terminal . follow this commands :

[cloudera@quickstart ~]$ cd Desktop

[cloudera@quickstart Desktop]$ ls

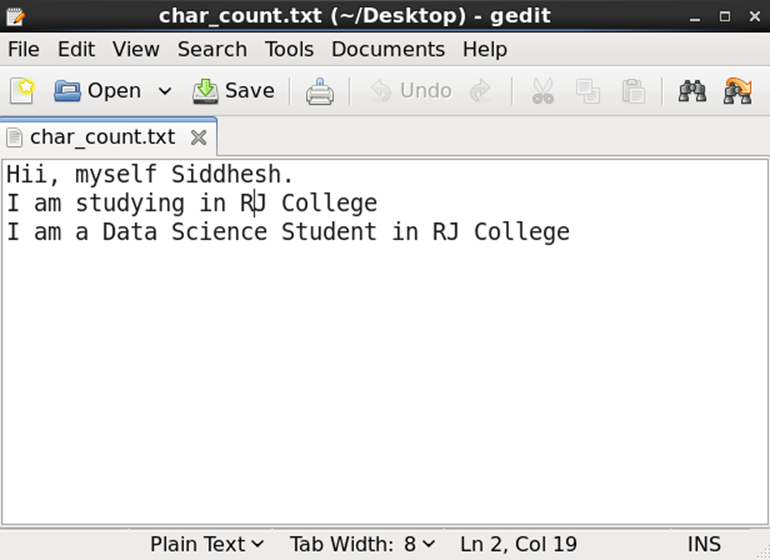
CharCount.jar Enterprise.desktop Kerberos.desktop Parcels.desktop

Eclipse.desktop Express.desktop matrix\_mul

[cloudera@quickstart Desktop]$ pwd

/home/cloudera/Desktop

[cloudera@quickstart Desktop]$ gedit char\_count.txt



[cloudera@quickstart Desktop]$ cat char\_count.txt

Hii, myself Siddhesh.

I am studying in RJ College

I am a Data Science Student in RJ College

[cloudera@quickstart Desktop]$ hdfs dfs -ls /

Found 6 items

drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks

drwxr-xr-x - hbase supergroup 0 2023-02-17 23:53 /hbase

drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr

drwxrwxrwt - hdfs supergroup 0 2023-02-03 02:48 /tmp

drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user

drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var

[cloudera@quickstart Desktop]$ hdfs dfs -mkdir /char\_inputdir

[cloudera@quickstart ~]$ hdfs dfs -ls /

Found 7 items

drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks

drwxr-xr-x - cloudera supergroup 0 2023-02-18 00:44 /char\_inputdir

drwxr-xr-x - hbase supergroup 0 2023-02-17 23:53 /hbase

drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr

drwxrwxrwt - hdfs supergroup 0 2023-02-03 02:48 /tmp

drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user

drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var

[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/char\_count.txt /char\_inputdir/

[cloudera@quickstart ~]$ hdfs dfs -ls /char\_inputdir

Found 1 items

-rw-r--r-- 1 cloudera supergroup 92 2023-02-18 00:53 /char\_inputdir/char\_count.txt

[cloudera@quickstart ~]$ hdfs dfs -cat /char\_inputdir/char\_count.txt

Hii, myself Siddhesh.

I am studying in RJ College

I am a Data Science Student in RJ College

[cloudera@quickstart ~]$ hadoop jar /home/cloudera/Desktop/CharCount.jar CharCountDriver

23/02/18 01:00:37 INFO mapreduce.Job: map 0% reduce 0%

23/02/18 01:00:58 INFO mapreduce.Job: map 50% reduce 0%

23/02/18 01:00:59 INFO mapreduce.Job: map 100% reduce 0%

23/02/18 01:01:15 INFO mapreduce.Job: map 100% reduce 100%

23/02/18 01:01:16 INFO mapreduce.Job: Job job\_1676706772828\_0001 completed successfully

23/02/18 01:01:16 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=364

FILE: Number of bytes written=431890

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=362

HDFS: Number of bytes written=108

HDFS: Number of read operations=9

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

Job Counters

Launched map tasks=2

Launched reduce tasks=1

Data-local map tasks=2

Total time spent by all maps in occupied slots (ms)=37871

Total time spent by all reduces in occupied slots (ms)=14386

Total time spent by all map tasks (ms)=37871

Total time spent by all reduce tasks (ms)=14386

Total vcore-milliseconds taken by all map tasks=37871

Total vcore-milliseconds taken by all reduce tasks=14386

Total megabyte-milliseconds taken by all map tasks=38779904

Total megabyte-milliseconds taken by all reduce tasks=14731264

Map-Reduce Framework

Map input records=3

Map output records=92

Map output bytes=549

Map output materialized bytes=370

Input split bytes=224

Combine input records=92

Combine output records=45

Reduce input groups=27

Reduce shuffle bytes=370

Reduce input records=45

Reduce output records=27

Spilled Records=90

Shuffled Maps =2

Failed Shuffles=0

Merged Map outputs=2

GC time elapsed (ms)=514

CPU time spent (ms)=2340

Physical memory (bytes) snapshot=640139264

Virtual memory (bytes) snapshot=4519460864

Total committed heap usage (bytes)=527638528

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=138

File Output Format Counters

Bytes Written=108

[cloudera@quickstart ~]$ hdfs dfs -ls /

Found 8 items

drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks

drwxr-xr-x - cloudera supergroup 0 2023-02-18 00:53 /char\_inputdir

drwxr-xr-x - cloudera supergroup 0 2023-02-18 01:01 /char\_outputdir

drwxr-xr-x - hbase supergroup 0 2023-02-17 23:53 /hbase

drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr

drwxrwxrwt - hdfs supergroup 0 2023-02-03 02:48 /tmp

drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user

drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var

[cloudera@quickstart ~]$ hdfs dfs -ls /char\_outputdir

Found 2 items

-rw-r--r-- 1 cloudera supergroup 0 2023-02-18 01:01 /char\_outputdir/\_SUCCESS

-rw-r--r-- 1 cloudera supergroup 108 2023-02-18 01:01 /char\_outputdir/part-00000

**Output:**

[cloudera@quickstart ~]$ hdfs dfs -cat /char\_outputdir/part-00000

**3**

**15**

**, 1**

**. 1**

**C 2**

**D 1**

**H 1**

**I 2**

**J 2**

**R 2**

**S 3**

**a 5**

**c 2**

**d 4**

**e 9**

**f 1**

**g 3**

**h 2**

**i 7**

**l 5**

**m 3**

**n 5**

**o 2**

**s 3**

**t 4**

**u 2**

**y 2**