

For Join Data:

Step 1:

To Count the number of rows in JOIN data, Follow the commands in Achyutha_NagavetiBhavaniSanthoshi_README_2

Step 2:

To run LineCount jar file

```
anagave@hadoop-nn001:~$ hadoop jar lc.jar LineCount /user/anagave/output/part00003 /user/anagave/LineCount_output_part4
```

```
2021-03-12 19:49:43,144 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2021-03-12 19:49:44,044 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at hadoop-nn001.cs.okstate.edu/192.168.122.2:8032
2021-03-12 19:49:44,272 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at hadoop-nn001.cs.okstate.edu/192.168.122.2:8032
2021-03-12 19:49:44,542 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2021-03-12 19:49:44,566 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/anagave/.staging/job_1615570959549_0095
2021-03-12 19:49:44,893 INFO mapred.FileInputFormat: Total input files to process : 1
2021-03-12 19:49:44,997 INFO mapreduce.JobSubmitter: number of splits:2
2021-03-12 19:49:45,157 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1615570959549_0095
2021-03-12 19:49:45,158 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-03-12 19:49:45,377 INFO conf.Configuration: resource-types.xml not found
2021-03-12 19:49:45,378 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-03-12 19:49:45,503 INFO impl.YarnClientImpl: Submitted application application_1615570959549_0095
2021-03-12 19:49:45,570 INFO mapreduce.Job: The url to track the job: http://hadoop-nn001.cs.okstate.edu:8088/proxy/application_1615570959549_0095/
2021-03-12 19:49:45,573 INFO mapreduce.Job: Running job: job_1615570959549_0095
2021-03-12 19:49:50,701 INFO mapreduce.Job: Job job_1615570959549_0095 running in uber mode : false
2021-03-12 19:49:50,704 INFO mapreduce.Job: map 0% reduce 0%
2021-03-12 19:49:54,823 INFO mapreduce.Job: map 100% reduce 0%
2021-03-12 19:49:59,893 INFO mapreduce.Job: map 100% reduce 80%
2021-03-12 19:50:00,907 INFO mapreduce.Job: map 100% reduce 100%
2021-03-12 19:50:00,924 INFO mapreduce.Job: Job job_1615570959549_0095 completed successfully
2021-03-12 19:50:01,067 INFO mapreduce.Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=66
        FILE: Number of bytes written=1853623
        FILE: Number of read operations=0
```

FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=199712
HDFS: Number of bytes written=15
HDFS: Number of read operations=31
HDFS: Number of large read operations=0
HDFS: Number of write operations=10
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=2
Launched reduce tasks=5
Rack-local map tasks=2
Total time spent by all maps in occupied slots (ms)=21905
Total time spent by all reduces in occupied slots (ms)=56380
Total time spent by all map tasks (ms)=4381
Total time spent by all reduce tasks (ms)=11276
Total vcore-milliseconds taken by all map tasks=4381
Total vcore-milliseconds taken by all reduce tasks=11276
Total megabyte-milliseconds taken by all map tasks=22430720
Total megabyte-milliseconds taken by all reduce tasks=57733120

Map-Reduce Framework

Map input records=39
Map output records=39
Map output bytes=624
Map output materialized bytes=96
Input split bytes=306
Combine input records=39
Combine output records=2
Reduce input groups=1
Reduce shuffle bytes=96
Reduce input records=2
Reduce output records=1
Spilled Records=4
Shuffled Maps =10
Failed Shuffles=0
Merged Map outputs=10
GC time elapsed (ms)=110
CPU time spent (ms)=4970
Physical memory (bytes) snapshot=1975676928
Virtual memory (bytes) snapshot=44709908480
Total committed heap usage (bytes)=5517606912
Peak Map Physical memory (bytes)=360275968
Peak Map Virtual memory (bytes)=6387335168
Peak Reduce Physical memory (bytes)=259465216
Peak Reduce Virtual memory (bytes)=6391468032

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=199406
File Output Format Counters
Bytes Written=15

Step 3:

To view the contents in output file

```
anagave@hadoop-nn001:~$ hdfs dfs -cat /user/anagave/  
LineCount_output_part4/part*  
2021-03-12 19:51:30,307 WARN util.NativeCodeLoader: Unable to load  
native-hadoop library for your platform... using builtin-java  
classes where applicable  
Total Lines 0
```