

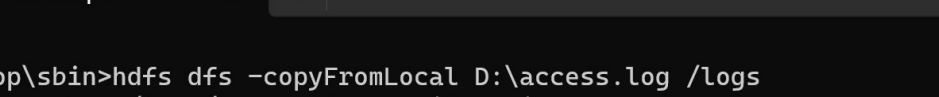
## PART – 2

The time taken to run code with combiner: 18.61sec

The time taken to run code without combiner: 20.17sec

The difference is about: 1.5sec

## Copying the access.log file from local to Hadoop cluster



The screenshot shows a Windows Command Prompt window with a dark background. The title bar at the top reads "Command Prompt" and includes standard window controls (close, maximize, and a dropdown arrow). The command prompt shows the following sequence of commands and outputs:

```
D:\hadoop\sbin>hdfs dfs -copyFromLocal D:\access.log /logs
copyFromLocal: `/logs/access.log': File exists

D:\hadoop\sbin>hdfs dfs -ls /logs
Found 1 items
-rw-r--r--    1 91833 supergroup        3497667 2023-06-22 15:26 /logs/access.log

D:\hadoop\sbin>
```

### Implementing the map\_reduce on access.log using .jar file of the Java Program.

```
D:\Hadoop>ls -lsh /usr/hdp-6.0-SNAPSHOT/jar/com.mycompany.hadoop.mapreduce.HadoopMapReduce /logs/access.log/output
ls: does not exist or is not a normal file: D:\Hadoop\MapReduce-1.8-SNAPSHOT

D:\Hadoop>ls -lsh /usr/hdp-6.0-SNAPSHOT/jar/com.mycompany.hadoop.mapreduce.HadoopMapReduce /logs/access.log/output
2023-06-22 17:12:61.794 INFO ClientMemoryPool Connecting to ResourceManager at /r.s.o.f.0:8032
2023-06-22 17:12:61.808 WARN ResourceControllerImpl Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-06-22 17:12:61.868 INFO org.apache.jobs.resourcemanager.EnableEraseCodingForPath: tmp/hadoop-yarn/staging/Y18331/staging/job_1687061736221_0001
2023-06-22 17:12:61.923 INFO FileContextMapperImpl Input loader files to process : 1
2023-06-22 17:12:61.941 INFO org.apache.jobs.submit: number of splits: 1
2023-06-22 17:12:61.943 INFO org.apache.jobs.submit: Submitting tokens for job: job_1687061736221_0001
2023-06-22 17:12:61.948 INFO org.apache.jobs.submit: Executing with tokens: []
2023-06-22 17:12:62.797 WARN conf.Configuration resource-types.xml not found
2023-06-22 17:12:62.797 WARN resource.ResourceUtil::Unable to find 'resource-types.xml'.
2023-06-22 17:12:63.262 INFO impl.YarnClientImpl: Submitted application application_1687061736221_0001
2023-06-22 17:12:63.328 INFO org.apache.job: The url to track the job: http://mapstar-laptop:8086/proxy/application_1687061736221_0001/
2023-06-22 17:12:63.328 INFO org.apache.job: Running Job: job_1687061736221_0001
2023-06-22 17:12:63.328 INFO org.apache.job: map 0% reduce 0%
2023-06-22 17:12:63.328 INFO org.apache.job: map 0% reduce 0% running in uber mode : false
2023-06-22 17:12:63.328 INFO org.apache.job: map 100% reduce 0%
2023-06-22 17:12:63.328 INFO org.apache.job: map 100% reduce 100%
2023-06-22 17:12:63.328 INFO org.apache.job: job_1687061736221_0001 completed successfully
2023-06-22 17:12:63.328 INFO org.apache.job: Counters: 36

File System Counters
FILE: Number of bytes read=1798213
FILE: Number of bytes written=6067601
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=907768
HDFS: Number of bytes written=931214
HDFS: Number of read operations=0
HDFS: Number of large read operations=0
HDFS: Number of write operations=0
HDFS: Number of bytes read erasure-coded=0
Job Counters
Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=4062
Total time spent by all reducers in occupied slots (ms)=2407
Total time spent by all map tasks (ms)=4062
Total time spent by all reduce tasks (ms)=2407
Total vcore-millisecods taken by all map tasks=4062
Total vcore-millisecods taken by all reduce tasks=2407
Total megabyte-millisecods taken by all map tasks=19088
Total megabyte-millisecods taken by all reduce tasks=3576068

Map-Reduce Framework
Map input records=31111
Map output records=31111
Map output bytes=172099
Map output materialized bytes=1798213
Input split bytes=182
Combine input records=0
Combine output records=0
Reduce input groups=1081
Reduce shuffle bytes=1798213
Reduce input records=31111
Reduce output records=1081
Spilled Records=78222
Shuffled Maps=51
Failed Shuffles=0
Averaged Map outputs=0
GC time elapsed (ms)=3
CPU time spent (%)=1264
Physical memory (bytes) snapshot=687358784
Virtual memory (bytes) snapshot=88007912
Total committed heap usage (bytes)=532676608
Peak Map Physical memory (bytes)=154595720
Peak Map Virtual memory (bytes)=mem6468
Peak Reduce Physical memory (bytes)=212760804
Peak Reduce Virtual memory (bytes)=450076752

Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
MR_IO_TIMEOUT=0
MR_MAP_FAILED=0
```