⚙ Execution Log Summary:

```
python MRTotalRevenueByStore.py \
> -r hadoop \
> hdfs:///user/hadoop/liquor_sales/Liquor_Sales_Final.csv \
> --output-dir hdfs:///user/hadoop/output/total_revenue_by_store/
No configs found; falling back on auto-configuration
No configs specified for hadoop runner
Looking for hadoop binary in $PATH...
Found hadoop binary: /usr/bin/hadoop
Using Hadoop version 2.8.5
Looking for Hadoop streaming jar in /home/hadoop/contrib...
Looking for Hadoop streaming jar in /usr/lib/hadoop-mapreduce...
Found Hadoop streaming jar: /usr/lib/hadoop-mapreduce/hadoop-streaming.jar
Creating temp directory
/tmp/MRTotalRevenueByStore.hadoop.20250428.160953.141935
uploading working dir files to
hdfs:///user/hadoop/tmp/mrjob/MRTotalRevenueByStore.hadoop.20250428.160953.14
1935/files
/wd...
Copying other local files to
hdfs:///user/hadoop/tmp/mrjob/MRTotalRevenueByStore.hadoop.20250428.160953.14
1935/files
/
Running step 1 of 1...
packageJobJar: [] [/usr/lib/hadoop/hadoop-streaming-2.8.5-amzn-6.jar]
/tmp/streamjob8813397535975255142.jar tmpDir=null
Connecting to ResourceManager at ip-172-31-99-99.ec2.internal/172.31.99.99:8032
Connecting to ResourceManager at ip-172-31-99-99.ec2.internal/172.31.99.99:8032
Loaded native gpl library
Successfully loaded & initialized native-lzo library [hadoop-lzo rev
ff8f5709577defb6b78cdc1f98cfe129c4b6fe46]
Total input files to process : 1
number of splits:36
Submitting tokens for job: job_1745854108616_0001
Submitted application application_1745854108616_0001
The url to track the job:
http://ip-172-31-99-99.ec2.internal:20888/proxy/application_1745854108616_0001/
Running job: job_1745854108616_0001
Job job_1745854108616_0001 running in uber mode : false
map 0% reduce 0%
map 4% reduce 0%
map 8% reduce 0%
```

map 10% reduce 0%
map 13% reduce 0%
map 17% reduce 0%
map 20% reduce 0%
map 21% reduce 0%
map 23% reduce 0%
map 25% reduce 0%
map 26% reduce 0%
map 28% reduce 0%
map 30% reduce 0%
map 31% reduce 0%
map 32% reduce 0%
map 33% reduce 0%
map 34% reduce 0%
map 36% reduce 0%
map 38% reduce 0%
map 39% reduce 0%
map 40% reduce 0%
map 42% reduce 0%
map 43% reduce 0%
map 44% reduce 2%
map 45% reduce 2%
map 46% reduce 2%
map 49% reduce 2%
map 51% reduce 2%
map 52% reduce 2%
map 54% reduce 2%
map 55% reduce 2%
map 56% reduce 2%
map 58% reduce 2%
map 59% reduce 2%
map 61% reduce 2%
map 61% reduce 5%
map 67% reduce 6%
map 68% reduce 6%
map 69% reduce 6%
map 73% reduce 6%
map 76% reduce 6%
map 77% reduce 6%
map 77% reduce 7%
map 78% reduce 7%
map 79% reduce 7%
map 80% reduce 7%
map 81% reduce 7%
map 83% reduce 7%

map 83% reduce 8%
map 89% reduce 8%
map 92% reduce 8%
map 92% reduce 9%
map 94% reduce 9%
map 97% reduce 9%
map 100% reduce 9%
map 100% reduce 14%
map 100% reduce 19%
map 100% reduce 30%
map 100% reduce 35%
map 100% reduce 42%
map 100% reduce 44%
map 100% reduce 46%
map 100% reduce 48%
map 100% reduce 50%
map 100% reduce 51%
map 100% reduce 54%
map 100% reduce 66%
map 100% reduce 68%
map 100% reduce 69%
map 100% reduce 71%
map 100% reduce 85%
map 100% reduce 99%
map 100% reduce 100%
Job job_1745854108616_0001 completed successfully
Output directory: hdfs:///user/hadoop/output/total_revenue_by_store/
Counters: 51
File Input Format Counters
Bytes Read=4724934886
File Output Format Counters
Bytes Written=36760
File System Counters
FILE: Number of bytes read=49280181
FILE: Number of bytes written=107288009
FILE: Number of large read operations=0
FILE: Number of read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=4724939998
HDFS: Number of bytes written=36760
HDFS: Number of large read operations=0
HDFS: Number of read operations=129
HDFS: Number of write operations=14
Job Counters
Data-local map tasks=36

Killed map tasks=1
Killed reduce tasks=1
Launched map tasks=36
Launched reduce tasks=8
Total megabyte-milliseconds taken by all map tasks=1448712192
Total megabyte-milliseconds taken by all reduce tasks=991340544
Total time spent by all map tasks (ms)=943172
Total time spent by all maps in occupied slots (ms)=1448712192
Total time spent by all reduce tasks (ms)=322702
Total time spent by all reduces in occupied slots (ms)=991340544
Total vcore-milliseconds taken by all map tasks=943172
Total vcore-milliseconds taken by all reduce tasks=322702
Map-Reduce Framework
CPU time spent (ms)=631410
Combine input records=0
Combine output records=0
Failed Shuffles=0
GC time elapsed (ms)=7104
Input split bytes=5112
Map input records=19485373
Map output bytes=221833789
Map output materialized bytes=50500449
Map output records=19485372
Merged Map outputs=252
Physical memory (bytes) snapshot=18596872192
Reduce input groups=2480
Reduce input records=19485372
Reduce output records=2480
Reduce shuffle bytes=50500449
Shuffled Maps =252
Spilled Records=38970744
Total committed heap usage (bytes)=18113626112
Virtual memory (bytes) snapshot=151015665664
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
job output is in hdfs:///user/hadoop/output/total_revenue_by_store/
Removing HDFS temp directory
hdfs:///user/hadoop/tmp/mrjob/MRTotalRevenueByStore.hadoop.20250428.160953.14
1935...

```
Removing temp directory
/tmp/MRTotalRevenueByStore.hadoop.20250428.160953.141935...
[hadoop@ip-172-31-99-99 ~]$
```