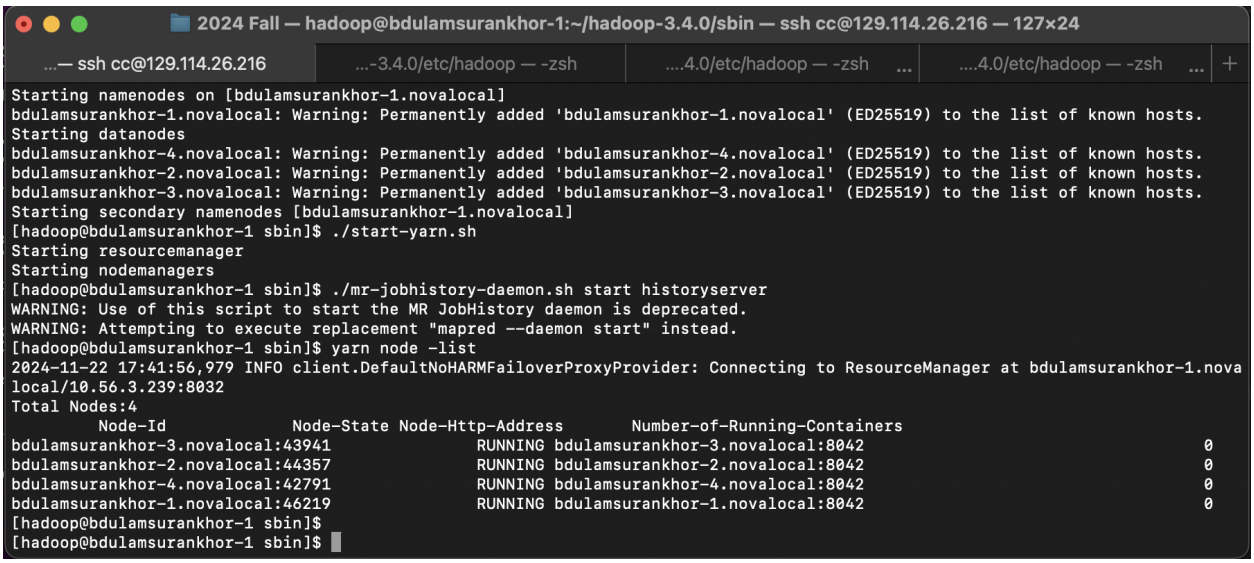
# CS451 Introduction to Parallel and Distributed Computing - Assignment 5

Batkhishig Dulamsurankhor - A20543498

November 22, 2024

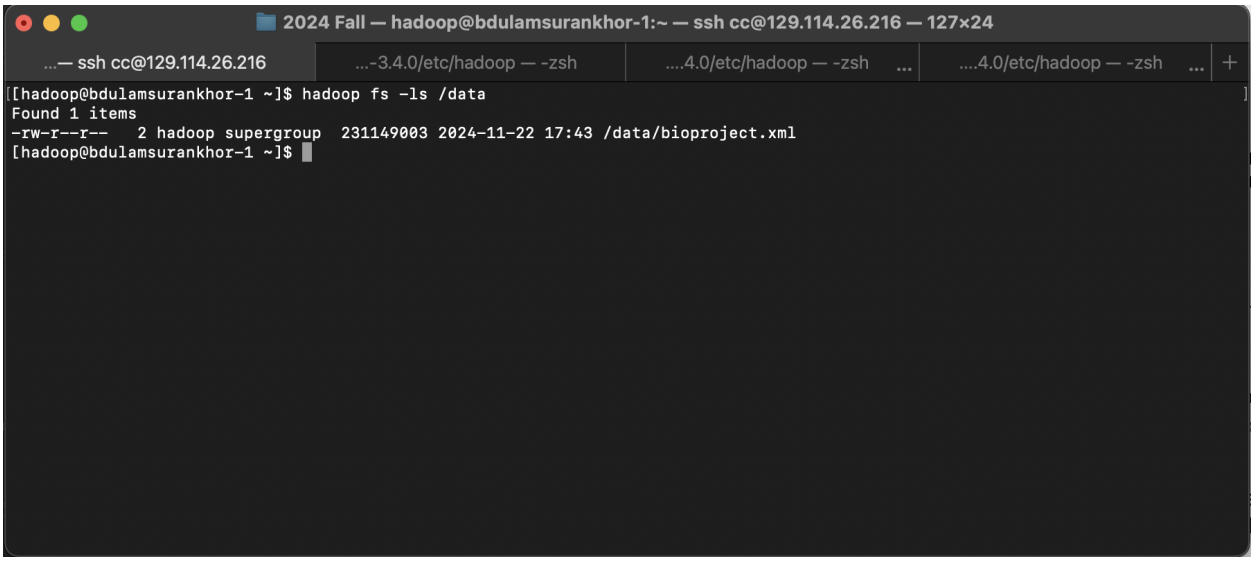
Part 1

14) a. i. yarn node -list

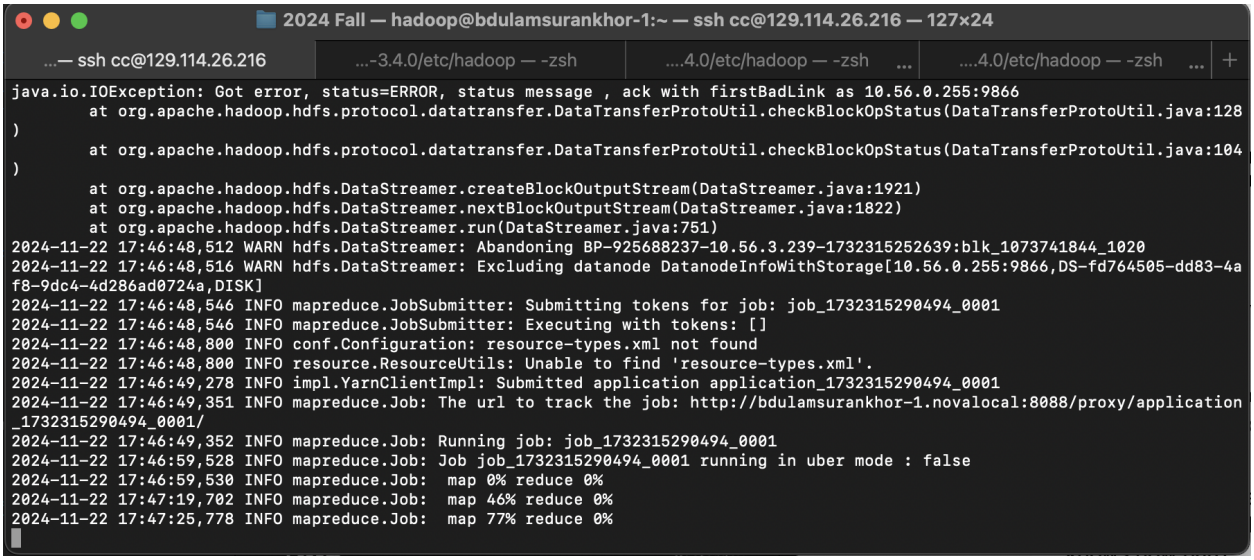



Part 2

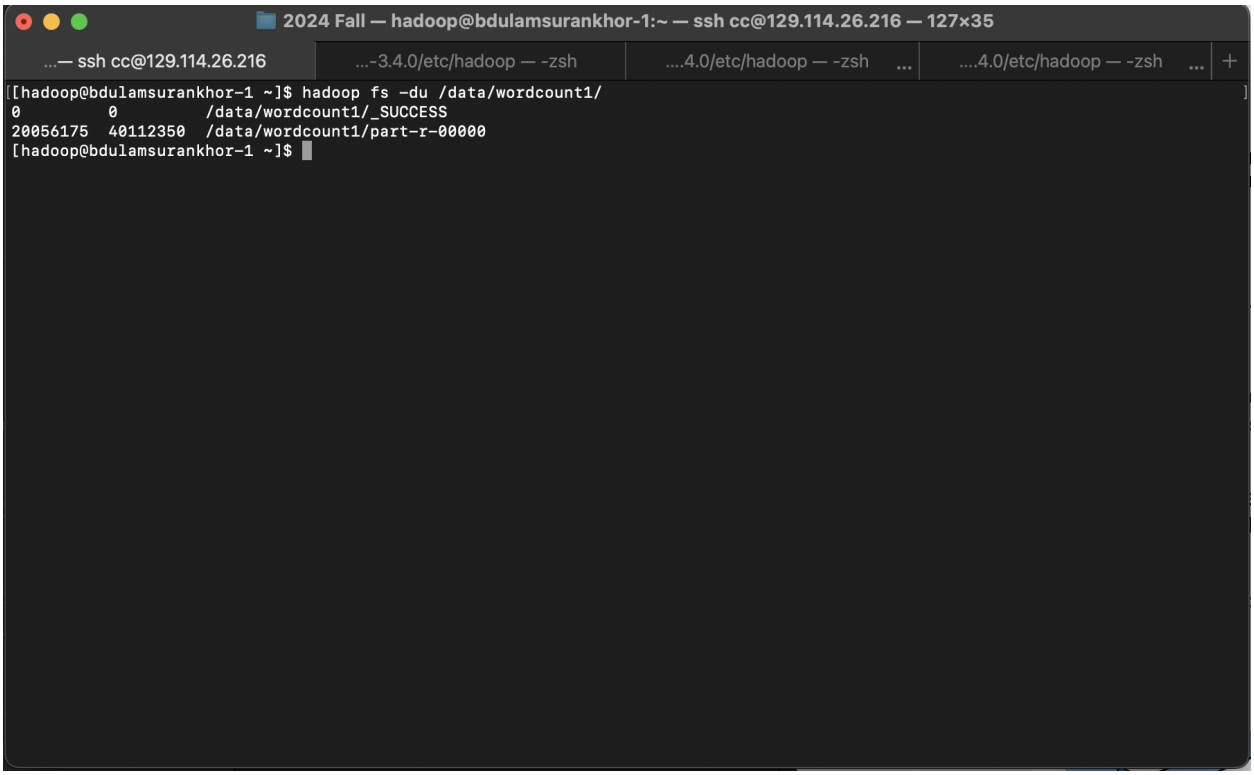
4) Verify that the file was uploaded to HDFS by hadoop fs -ls /data



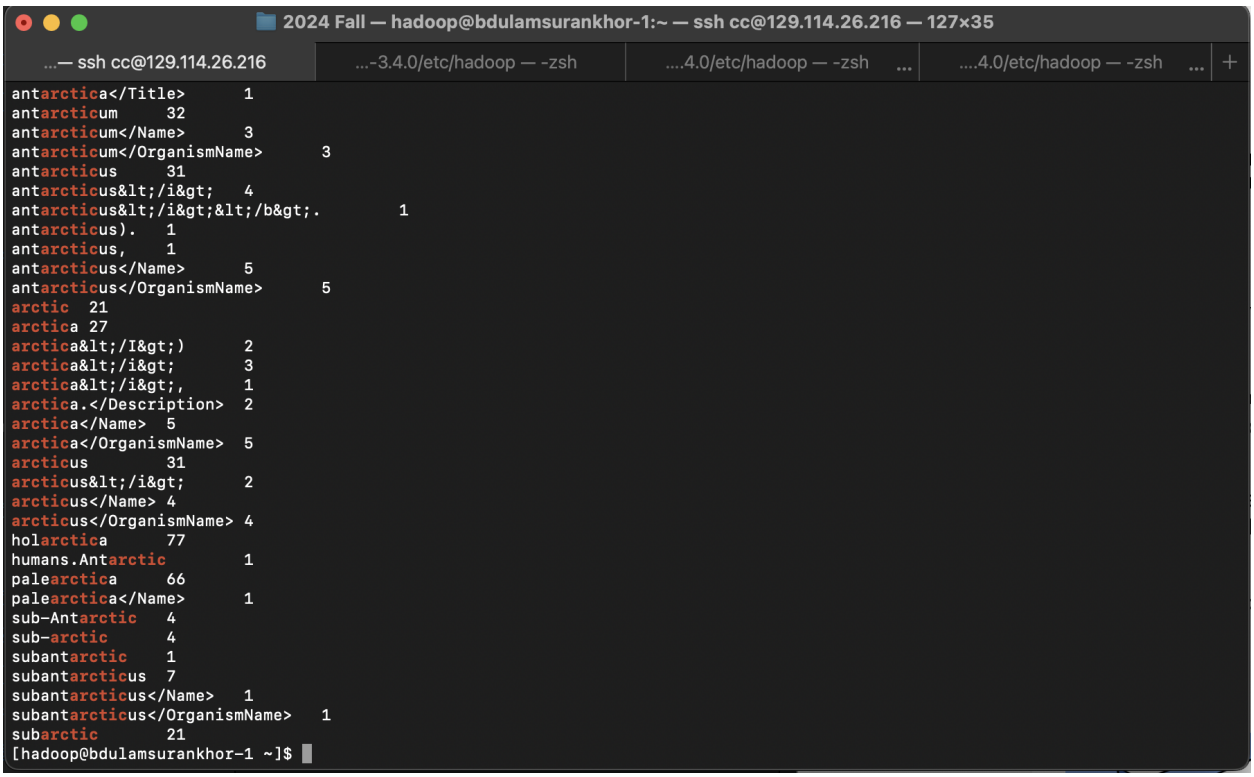

5) time hadoop jar hadoop-3.4.0/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.4.0.jar wordcount /data/bioproject.xml /data/wordcount1



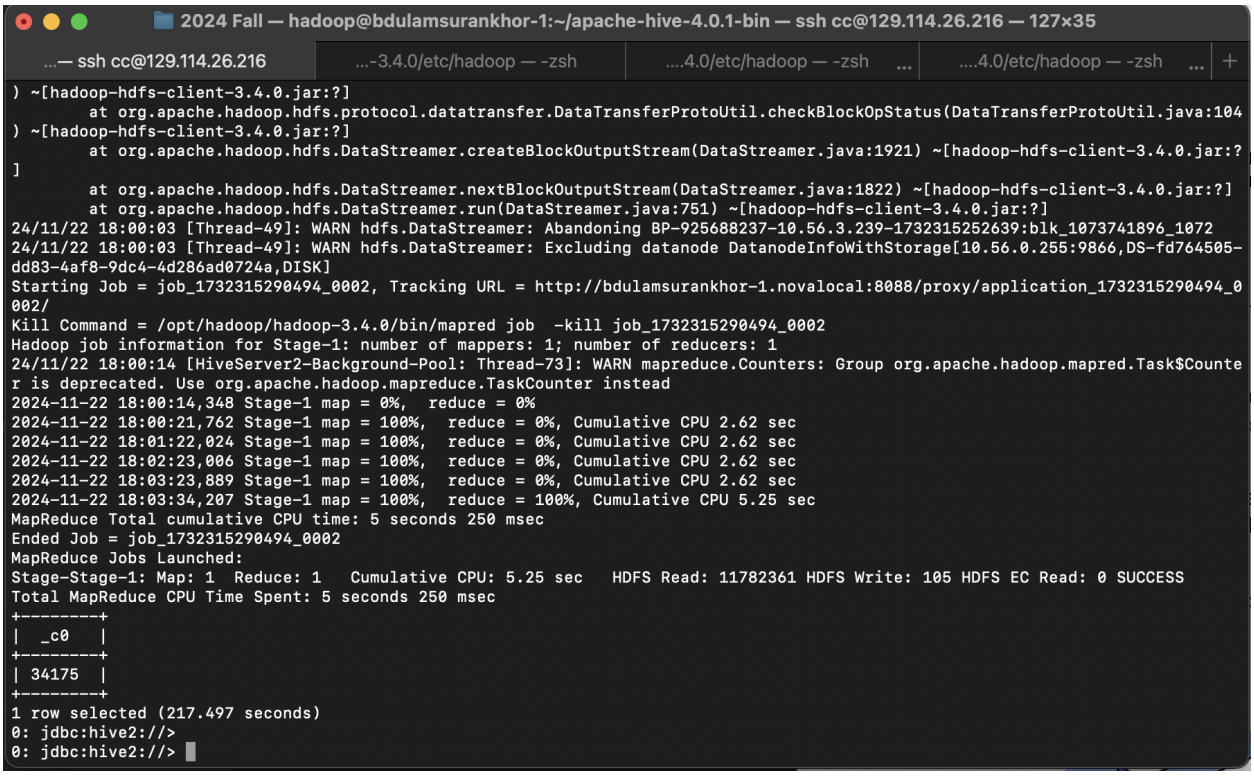

hadoop fs -du /data/wordcount1/



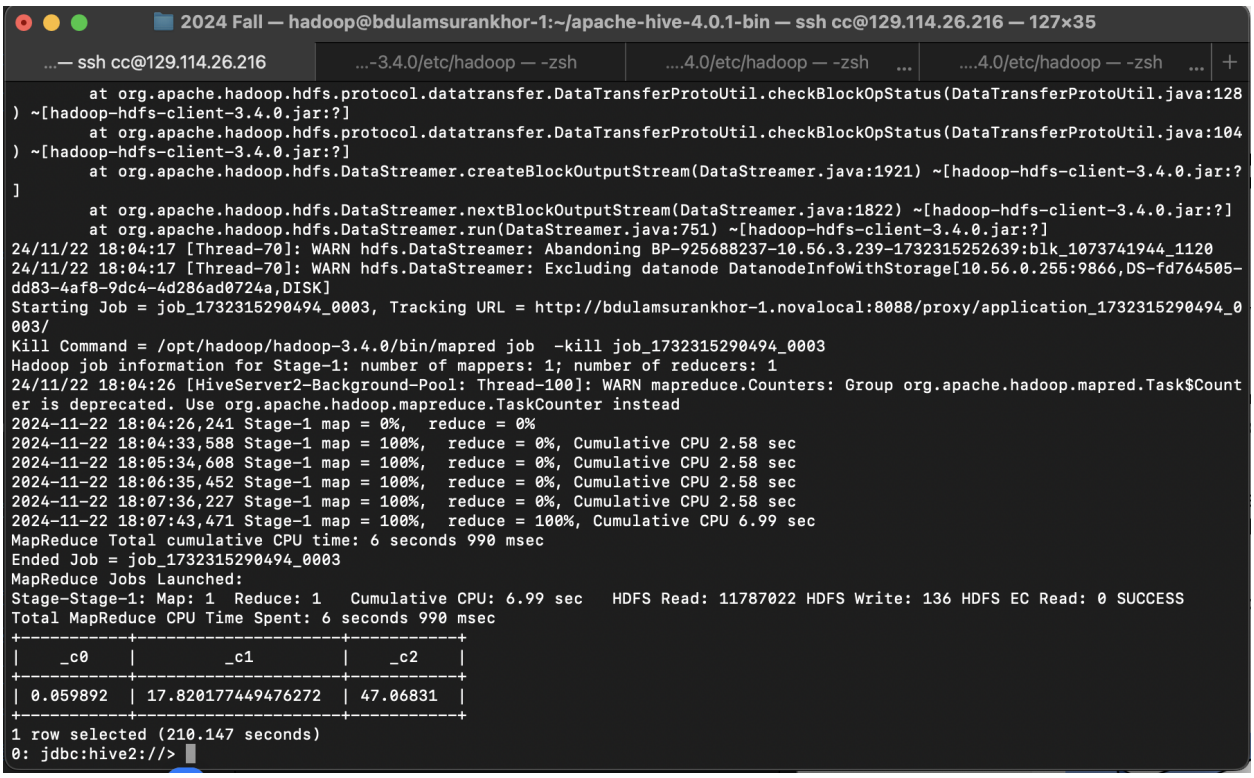

hadoop fs -cat /data/wordcount1/part-r-00000 — grep arctic



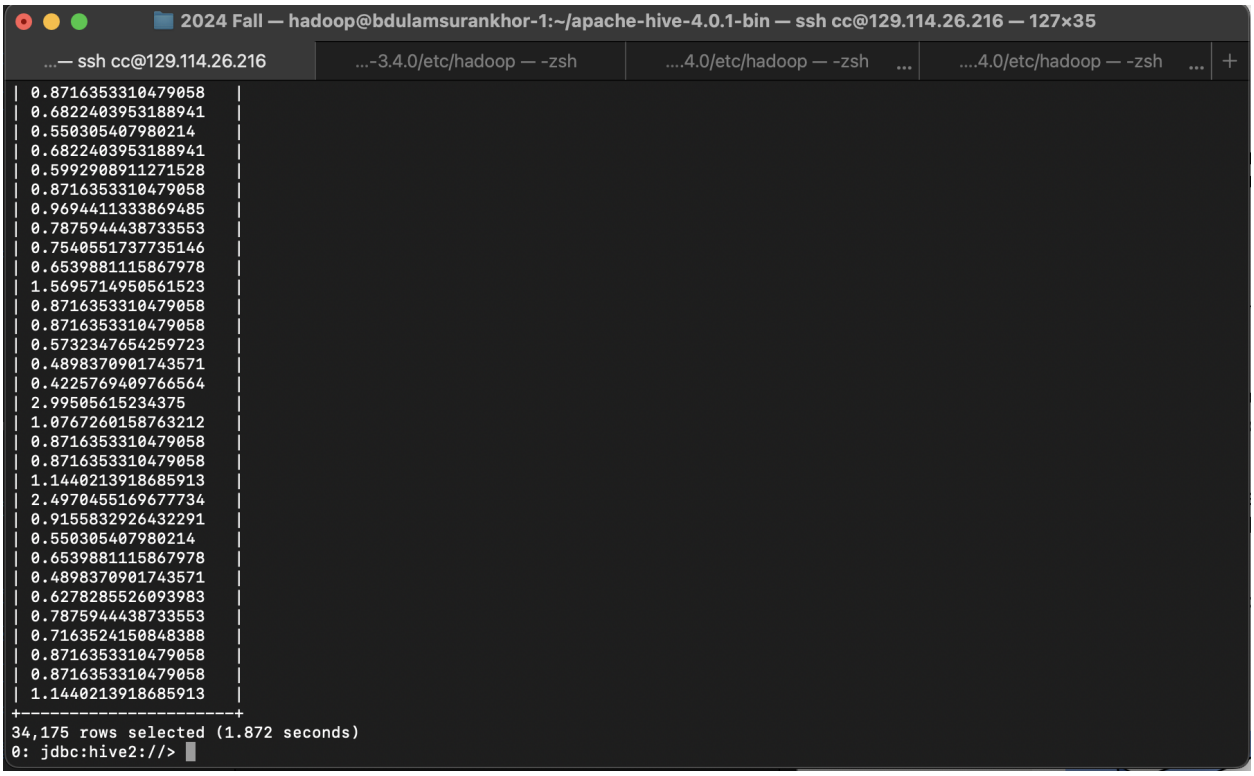

6) SELECT COUNT(*) FROM VehicleData;



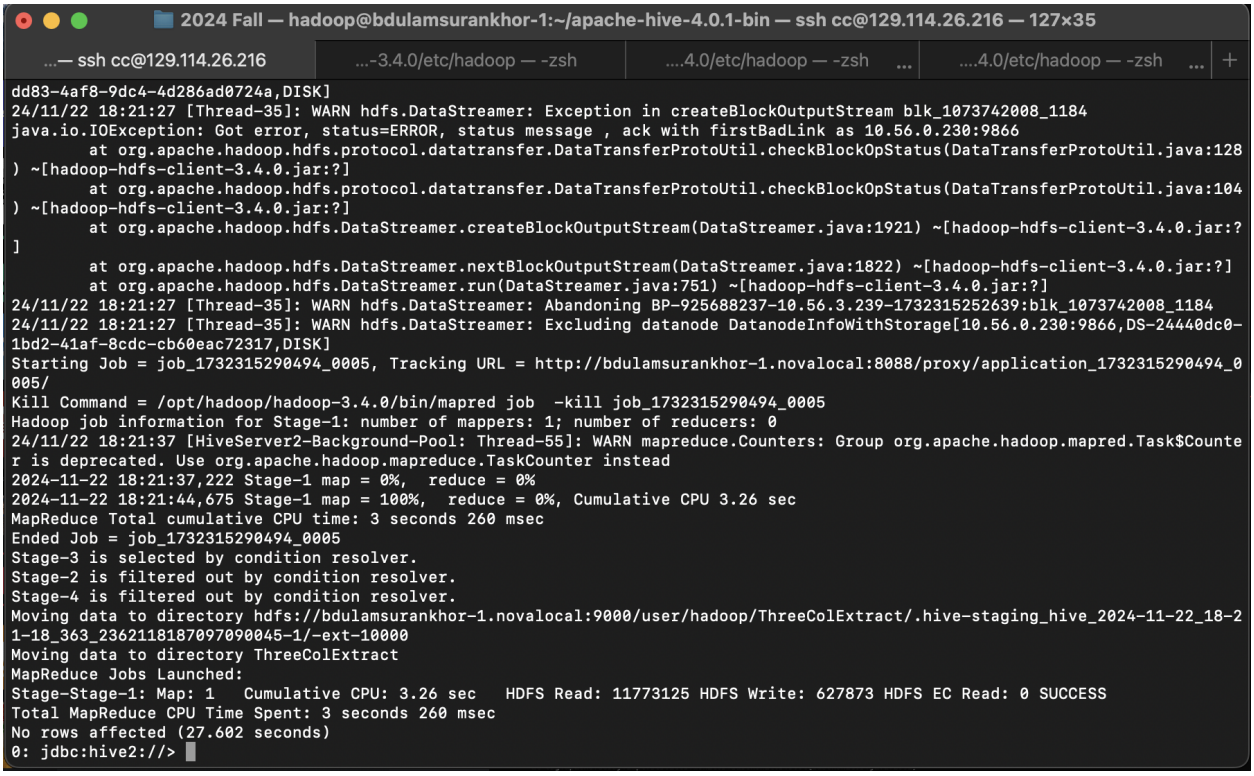

SELECT MIN(barrels08), AVG(barrels08), MAX(barrels08) FROM VehicleData;



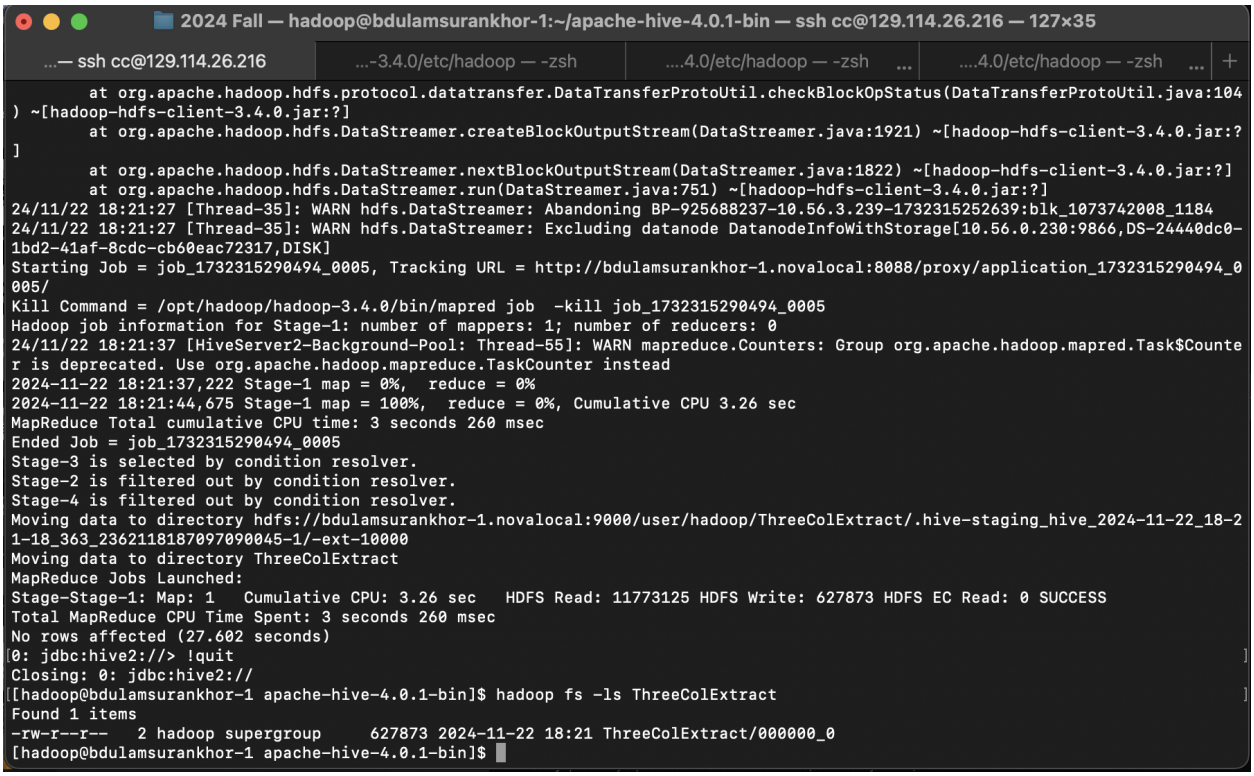

SELECT (barrels08/city08) FROM VehicleData;




INSERT OVERWRITE DIRECTORY 'ThreeColExtract' SELECT barrels08, city08, charge120 FROM VehicleData;



hadoop fs -ls ThreeColExtract

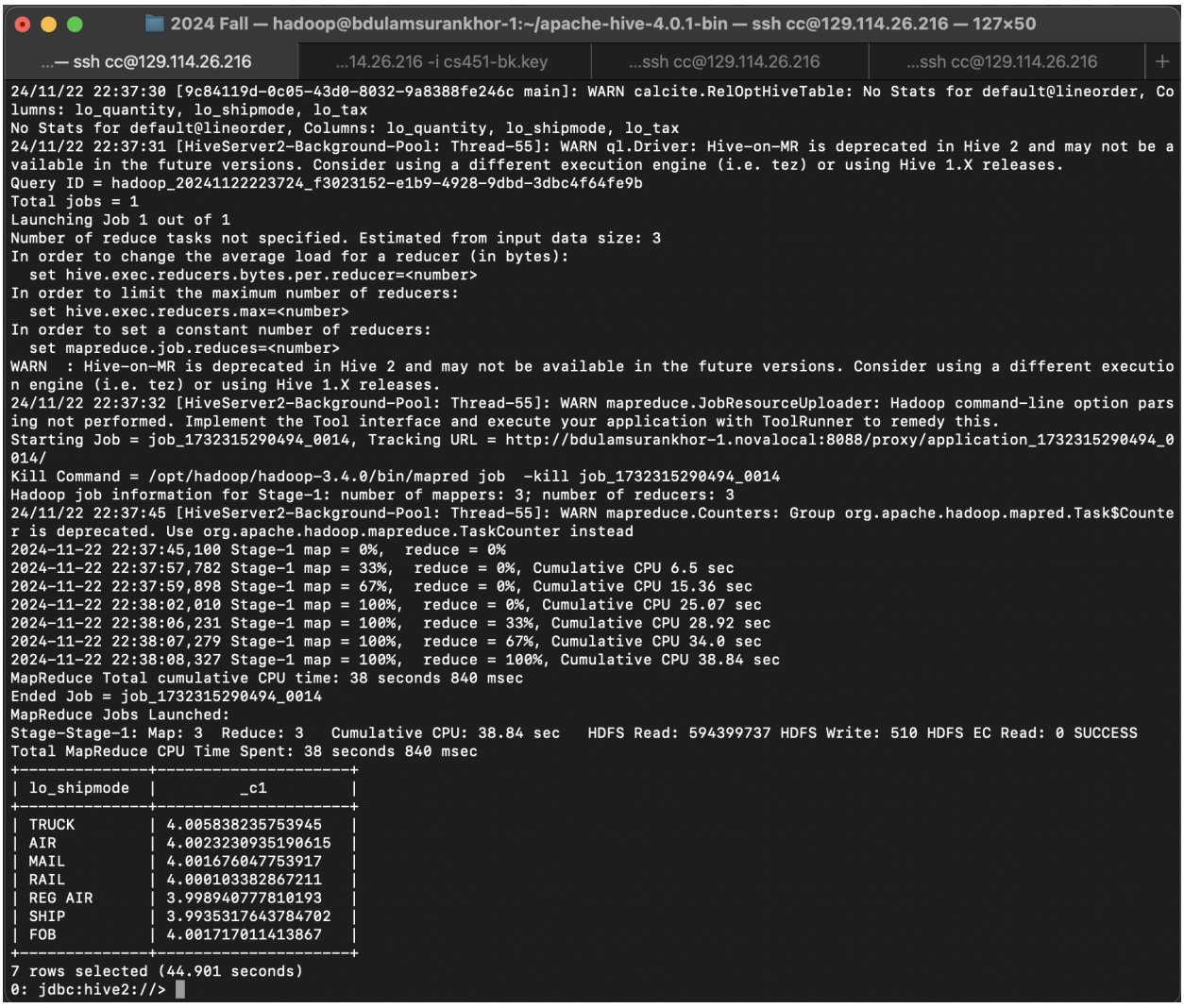



Download data:

wget http://cdmgcsarprd01.dpu.depaul.edu/CSC555/SSBM1/lineorder.tbl

Load data:

LOAD DATA LOCAL INPATH '/opt/hadoop/lineorder.tbl' OVERWRITE INTO TABLE lineorder;




7) Use the same vehicles file. Copy the vehicles.csv file to the HDFS if it is not already there.

VehicleData = LOAD '/data/vehicles.csv' USING PigStorage(',') AS (barrels08:FLOAT, barrelsA08:FLOAT, charge120:FLOAT, charge240:FLOAT, city08:FLOAT);

You can see the table description by

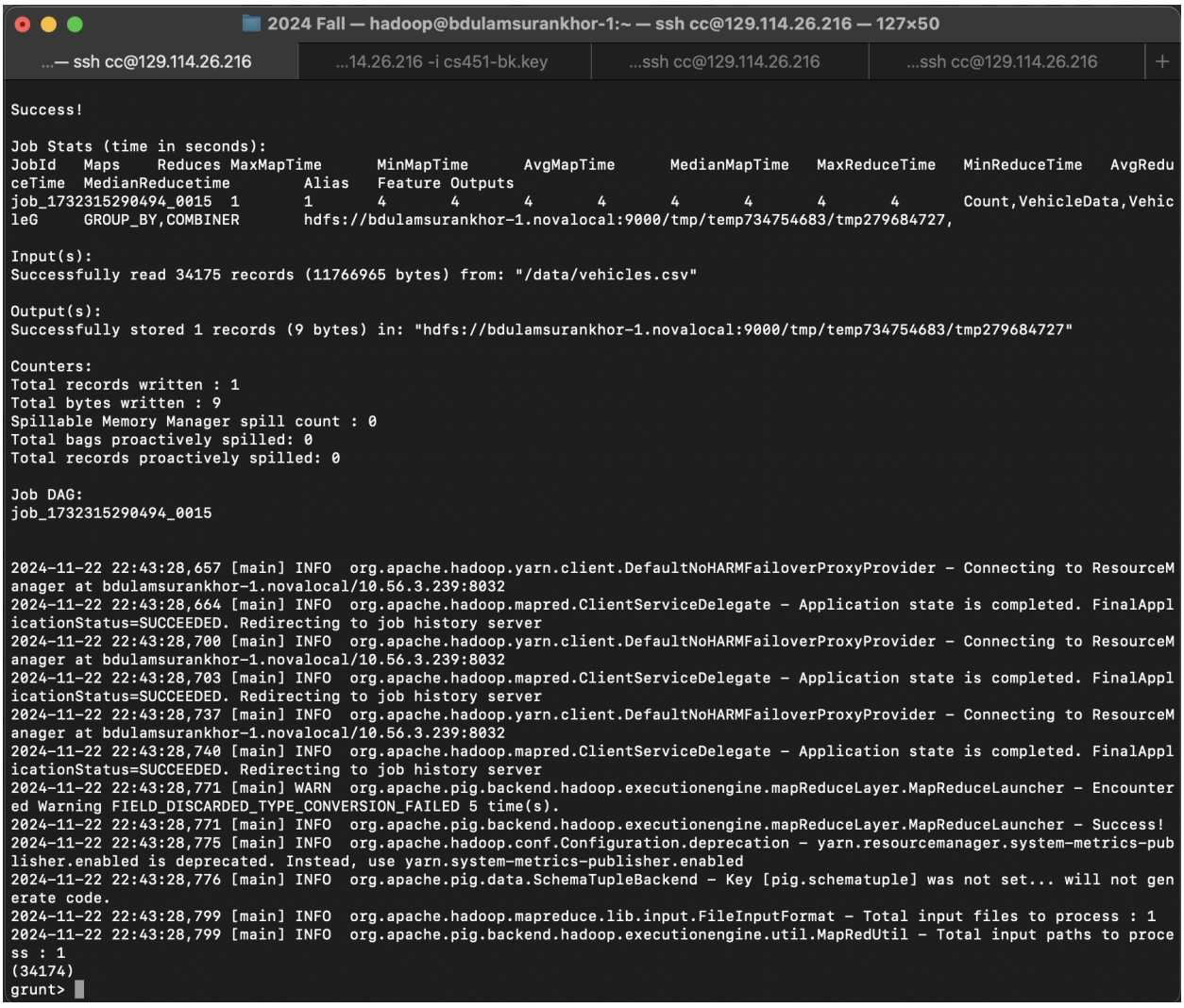
DESCRIBE VehicleData;

Verify that your data has loaded by running:

VehicleG = GROUP VehicleData ALL;

Count = FOREACH VehicleG GENERATE COUNT(VehicleData);

DUMP Count;


2024 Fall — hadoop@bdulamsurankhor-1:~ — ssh cc@129.114.26.216 — 127×50
...— ssh cc@129.114.26.216
...14.26.216 -i cs451-bk.key
...ssh cc@129.114.26.216
...ssh cc@129.114.26.216
Success!
Job Stats (time in seconds):
JobId Maps Reduces MaxMapTime MinMapTime AvgMapTime MedianMapTime MaxReduceTime MinReduceTime AvgReduceTime MedianReducetime Alias Feature Outputs
job_1732315290494_0015 1 1 4 4 4 4 4 4 4 Count,VehicleData,VehicleG GROUP_BY,COMBINER hdfs://bdulamsurankhor-1.novalocal:9000/tmp/temp734754683/tmp279684727,
Input(s):
Successfully read 34175 records (11766965 bytes) from: "/data/vehicles.csv"
Output(s):
Successfully stored 1 records (9 bytes) in: "hdfs://bdulamsurankhor-1.novalocal:9000/tmp/temp734754683/tmp279684727"
Counters:
Total records written : 1
Total bytes written : 9
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
Job DAG:
job_1732315290494_0015
2024-11-22 22:43:28,657 [main] INFO org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at bdulamsurankhor-1.novalocal/10.56.3.239:8032
2024-11-22 22:43:28,664 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2024-11-22 22:43:28,700 [main] INFO org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at bdulamsurankhor-1.novalocal/10.56.3.239:8032
2024-11-22 22:43:28,703 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2024-11-22 22:43:28,737 [main] INFO org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at bdulamsurankhor-1.novalocal/10.56.3.239:8032
2024-11-22 22:43:28,740 [main] INFO org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server
2024-11-22 22:43:28,771 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 5 time(s).
2024-11-22 22:43:28,771 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2024-11-22 22:43:28,775 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2024-11-22 22:43:28,776 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2024-11-22 22:43:28,799 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2024-11-22 22:43:28,799 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(34174)
grunt>