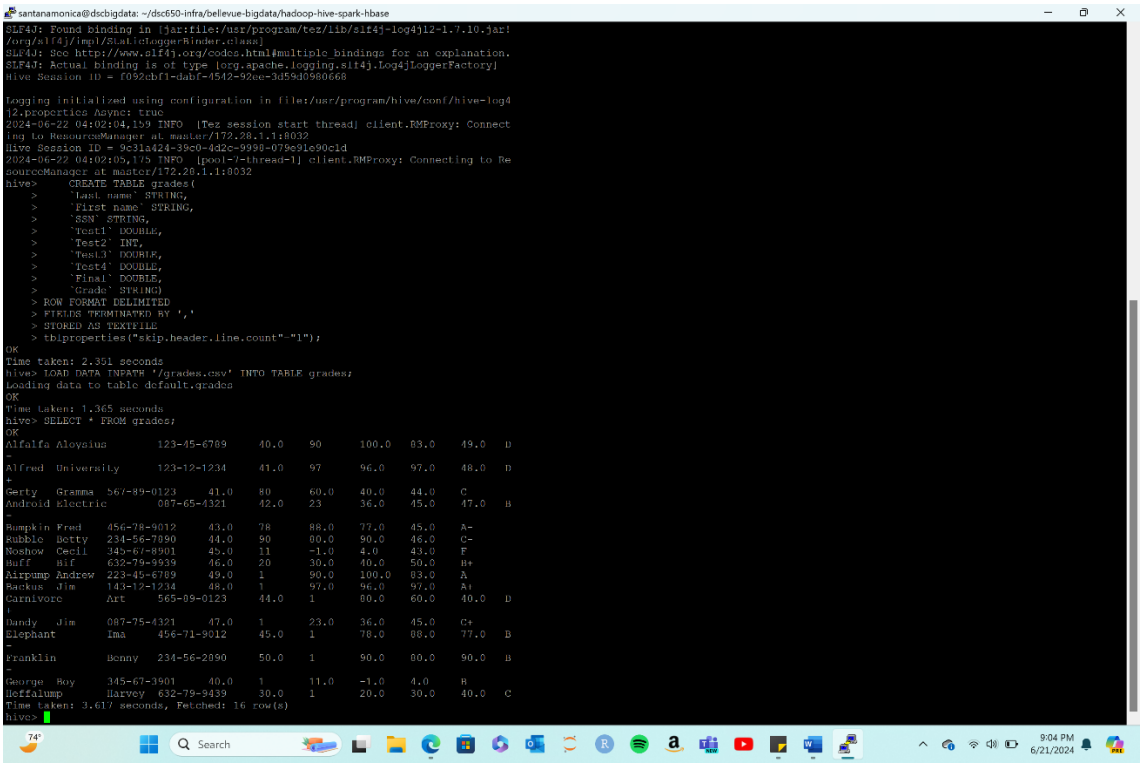
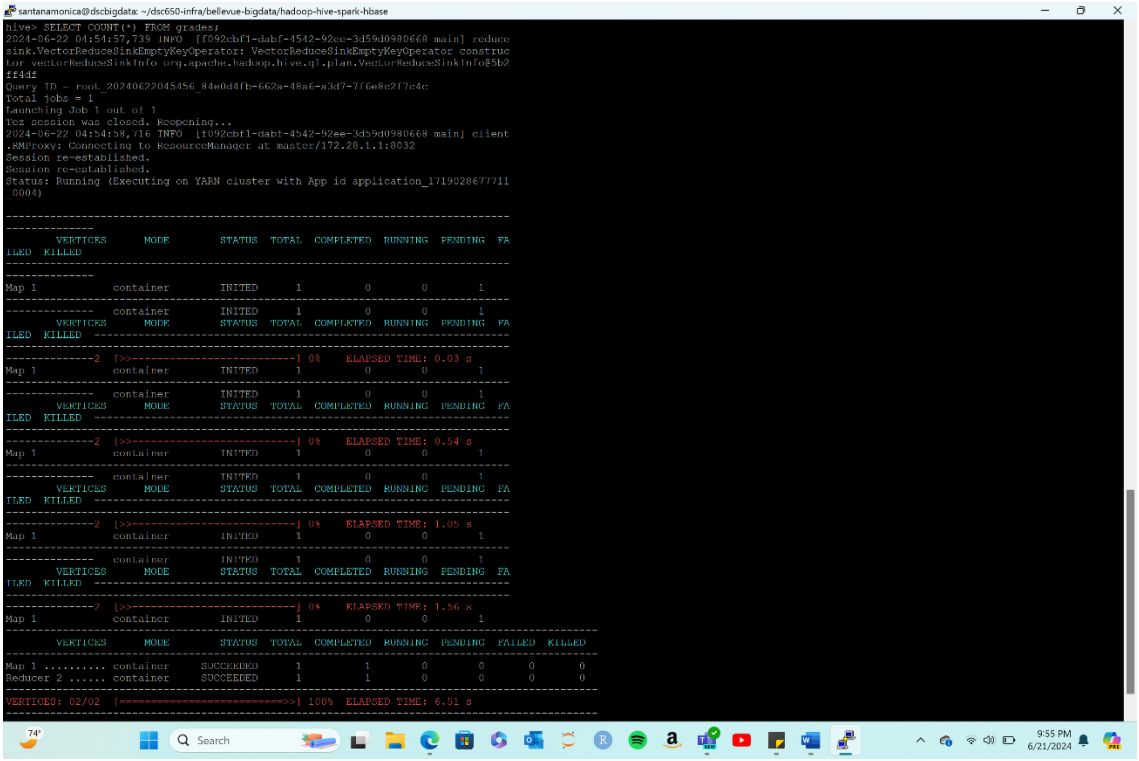
Monica Santana Week 3 -Exercise

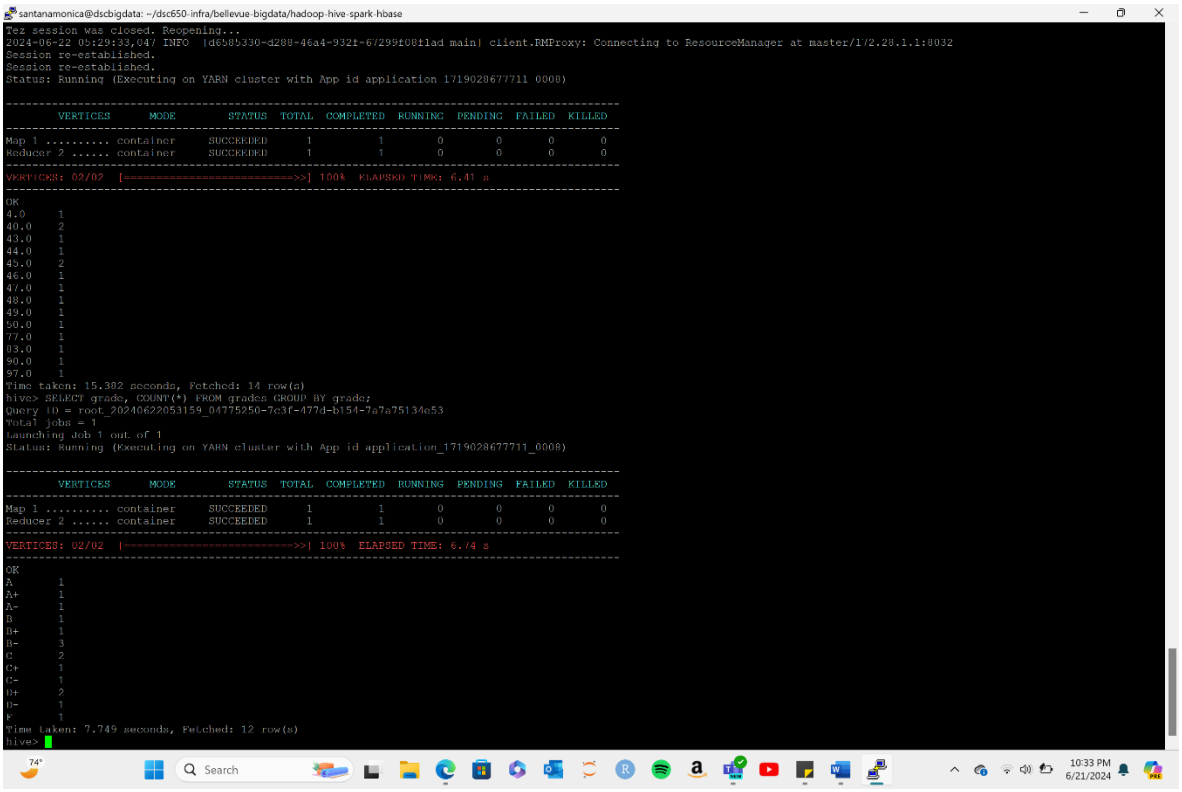
Grades.csv Results



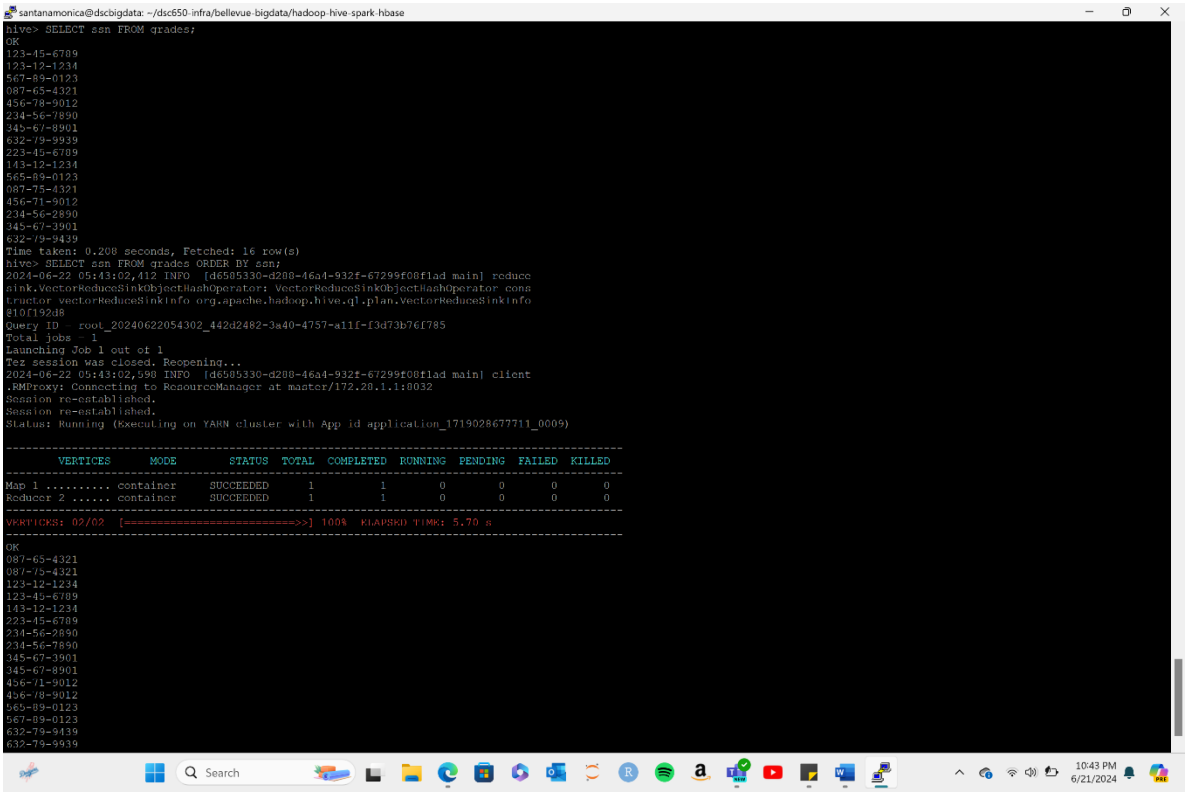
SQL Command #1



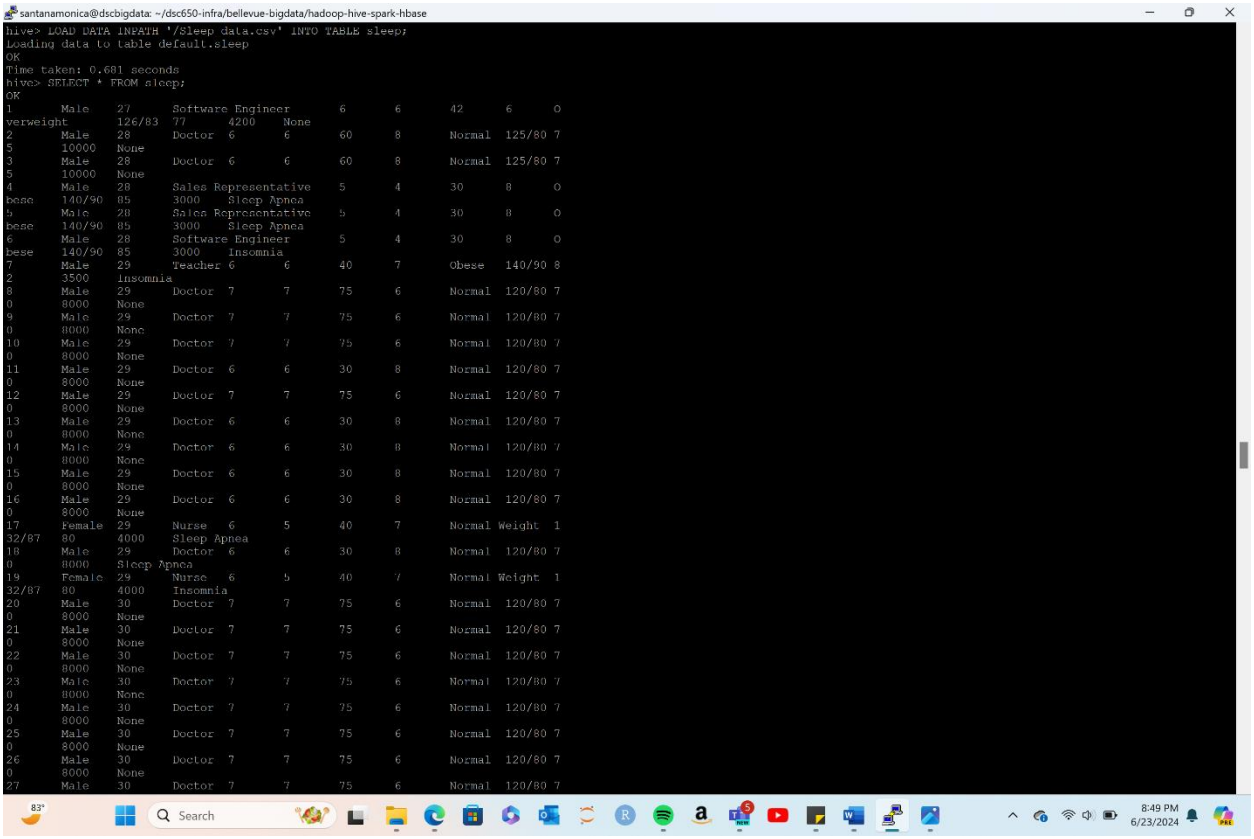
SQL Command #2



SQL Command #3 and #4

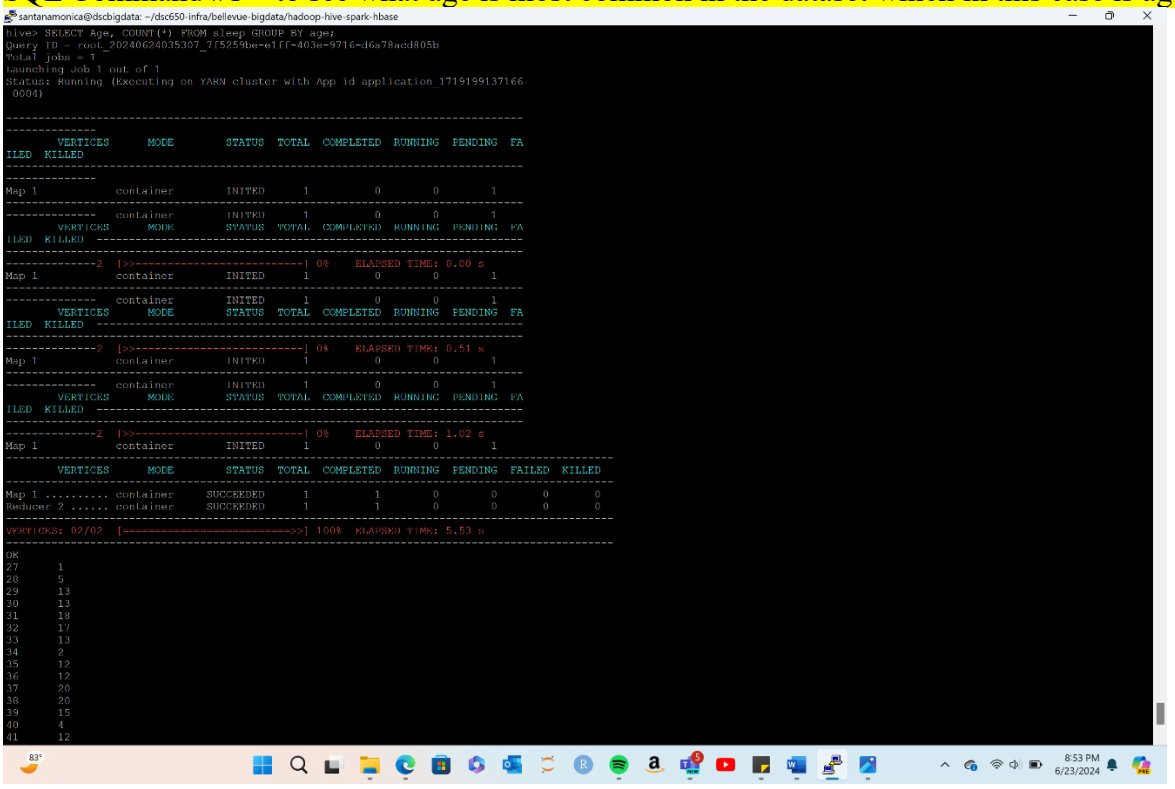


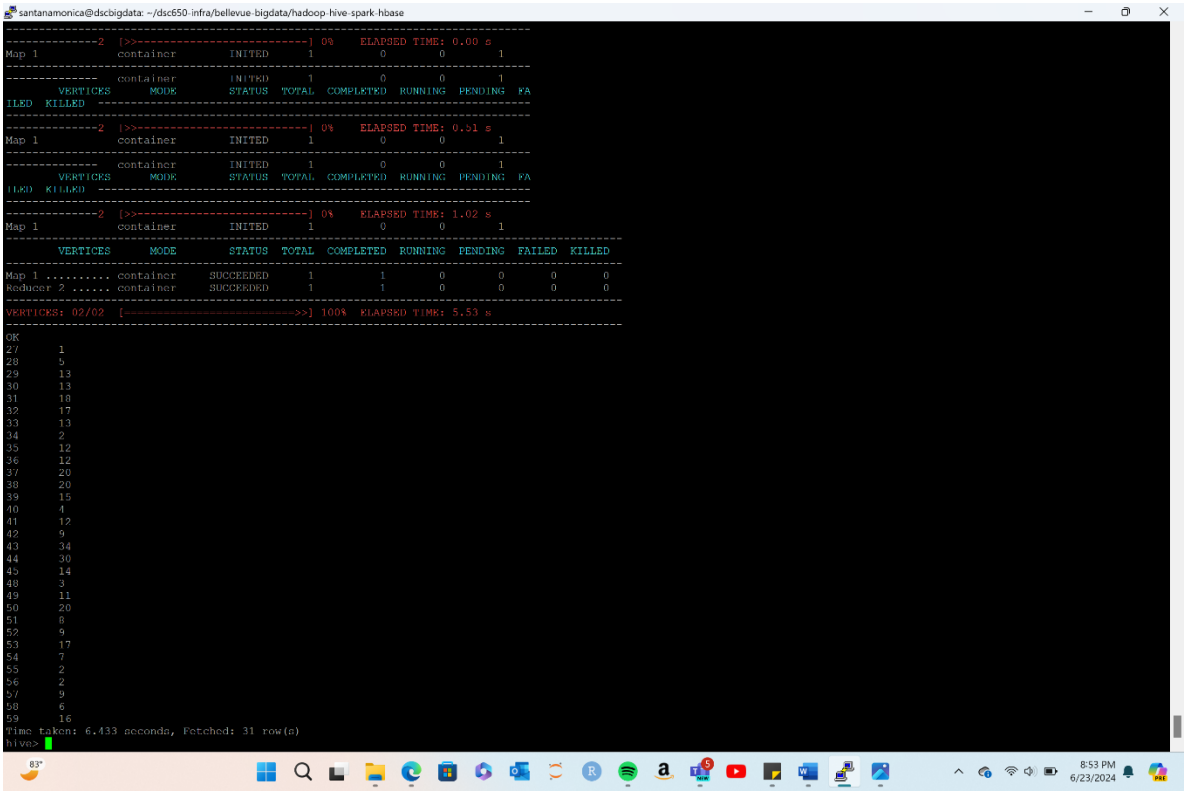
The dataset I chose for my SQL query is on sleep, and what features are used to predict whether someone has a sleep disorder of insomnia, sleep apnea or no sleep disorder at all. I hope to gain insights such as what gender is most apparent in the dataset, maximum age, most common age, and how many people are in the dataset total.



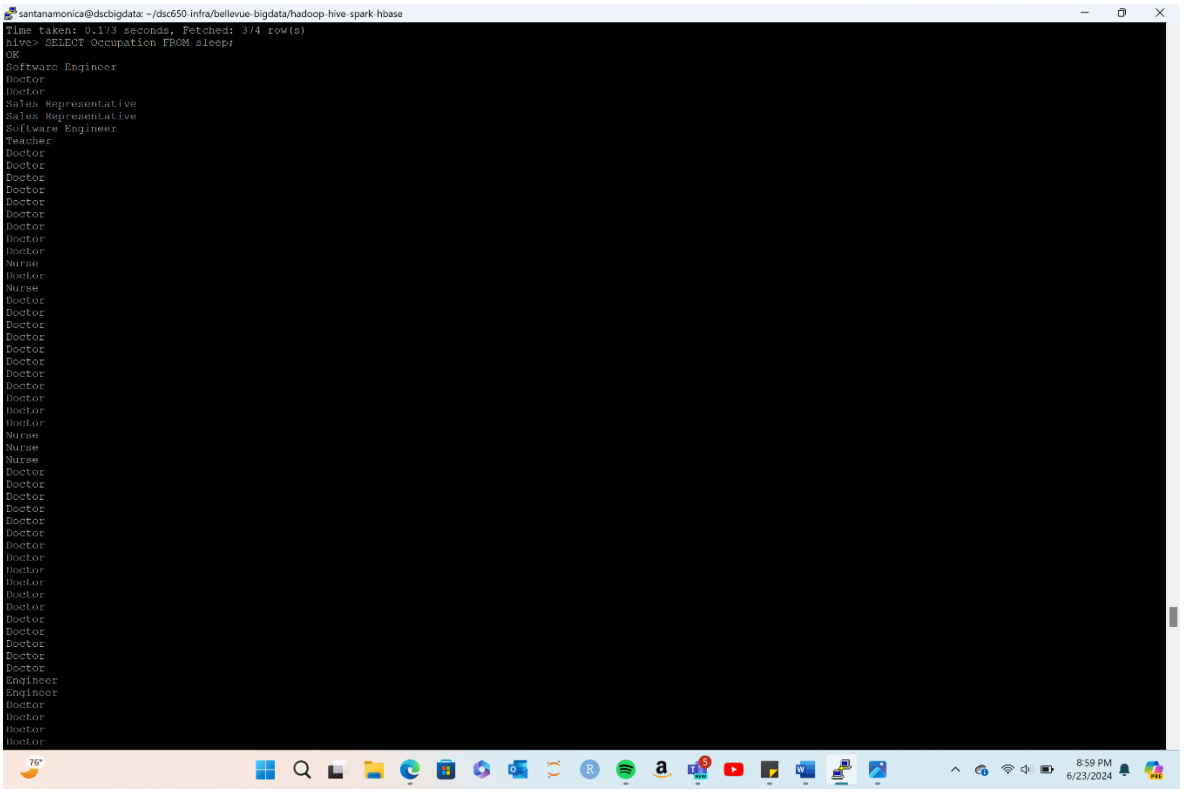

```
CREATE TABLE sleep(
    `Person ID` INT,
    `Gender` STRING,
    `Age` INT,
    `Occupation` STRING,
    `Sleep Duration` INT,
    `Quality of Sleep` INT,
    `Physical Activity Level` INT,
    `Stress Level` INT,
    `BMI Category` STRING,
    `Blood Pressure` STRING,
    `Heart Rate` INT,
    `Daily Steps` INT,
    `Sleep Disorder` STRING)
  ROW FORMAT DELIMITED
  FIELDS TERMINATED BY ','
  STORED AS TEXTFILE
tblproperties("skip.header.line.count"="1");
```

SQL Command #1 – to see what age is most common in the dataset which in this case is age 43

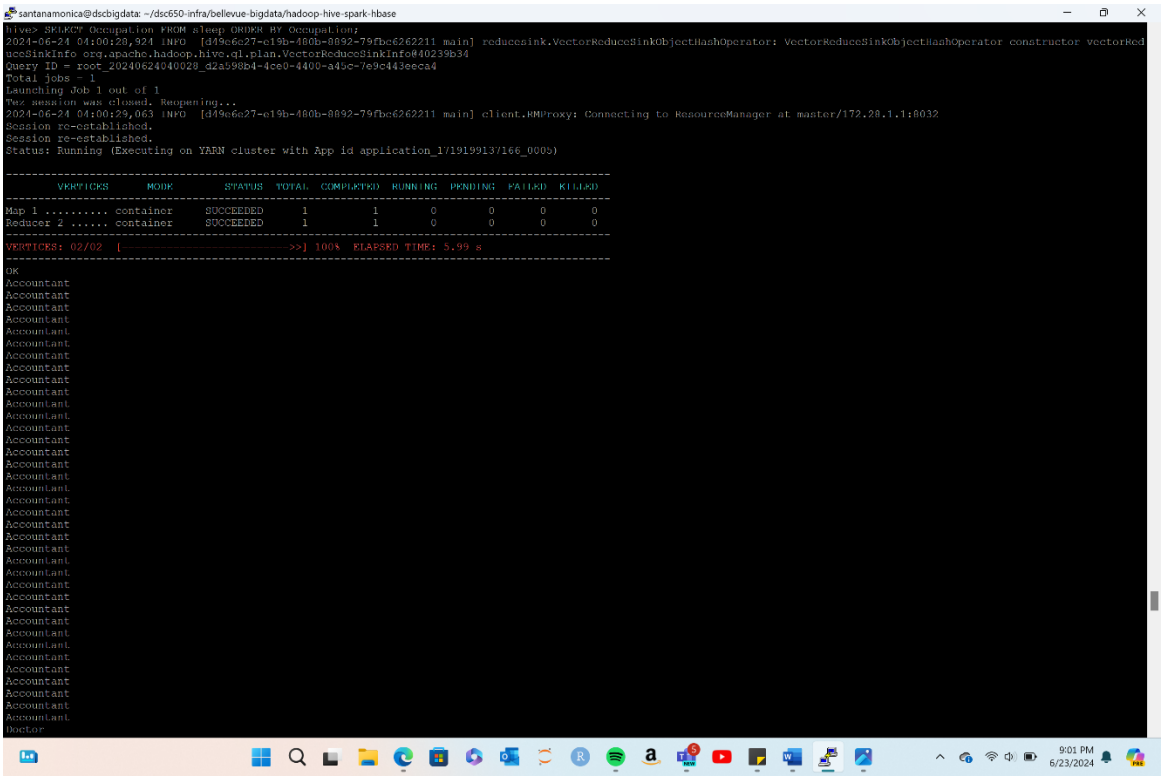




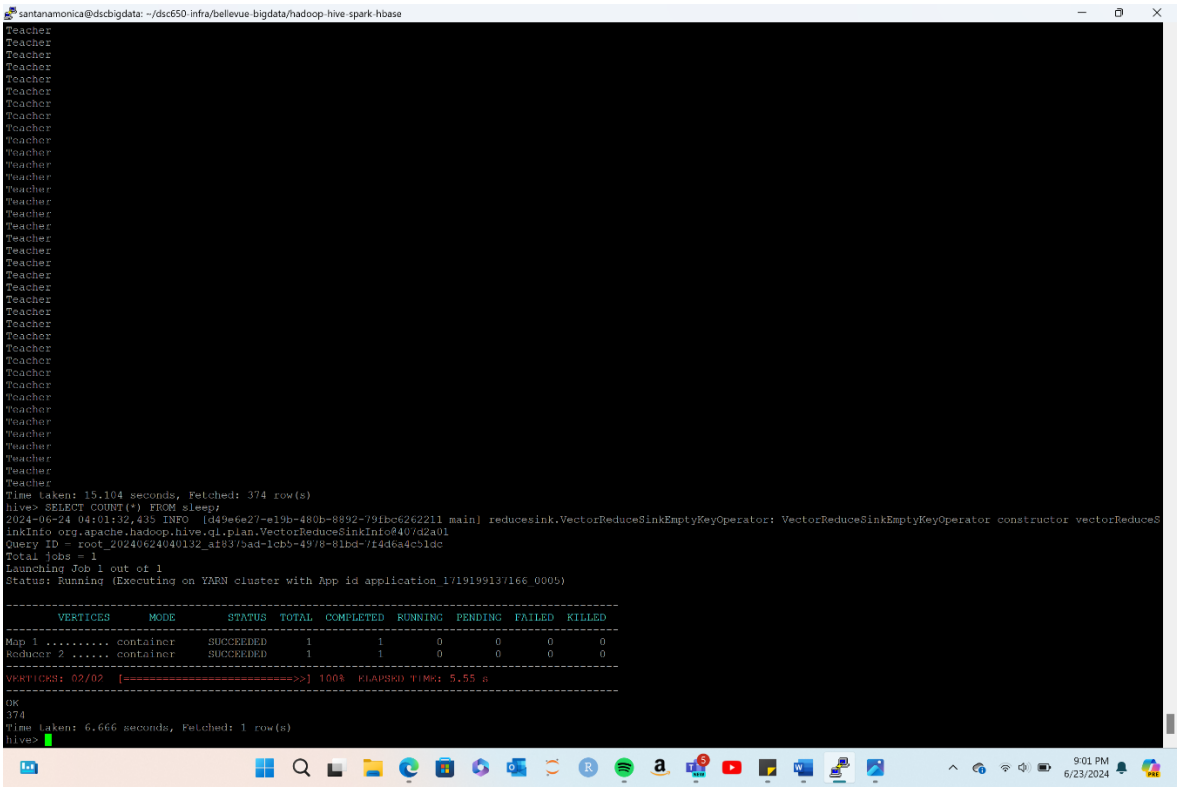
SQL Command #2 – separating the occupations from the sleep dataset



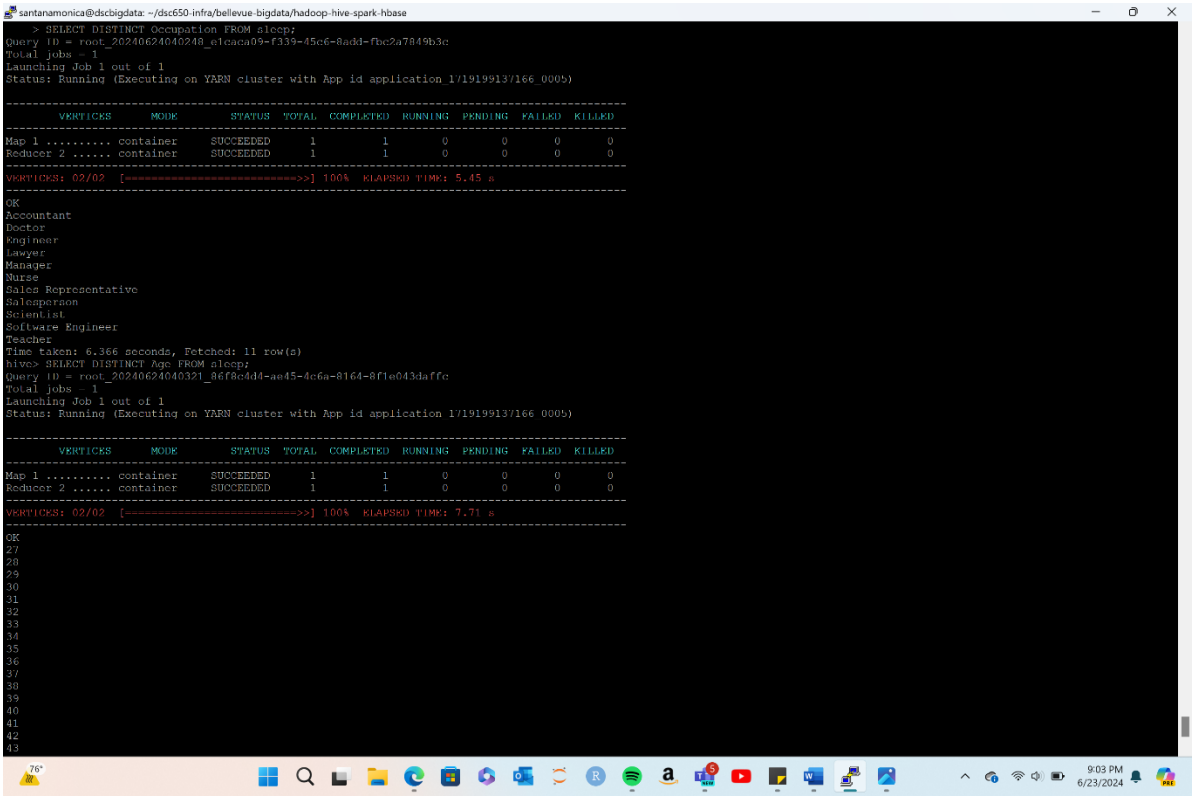
SQL Command #3 – sorting the occupations in alphabetical order


santanamonica@dscbigdata: ~/dsc650-infra/bellevue-bigdata/hadoop-hive-spark-hbase
hive> SELECT Occupation FROM sleep ORDER BY Occupation;
2024-06-24 04:00:28,924 INFO [d49c6a27-e19b-480b-8092-79fbc6262211 main] reducesink.VectorReduceSinkObjectHashOperator: VectorReduceSinkObjectHashOperator constructor vectorReduceSinkInfo org.apache.hadoop.hive.ql.plan.VectorReduceSinkInfo@40239b34
Query ID = root_20240624040028_d2a598b4-4ce0-4400-a4bc-7e9c443eeca4
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
2024-06-24 04:00:29,063 INFO [d49c6a27-e19b-480b-8092-79fbc6262211 main] client.RMProxy: Connecting to ResourceManager at master/172.28.1.1:8032
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1719199137166_0005)
VERTICES MODE STATUS TOTAL COMPLETED RUNNING PENDING FAILED KILLED
Map 1 ..... container SUCCEEDED 1 1 0 0 0 0
Reducer 2 ..... container SUCCEEDED 1 1 0 0 0 0
VERTICES: 02/02 [>>] 100% ELAPSED TIME: 5.99 s
OK
Accountant
Accountant
Accountant

SQL Command #4 – Counting how many people are in the dataset – which is 374



SQL Command #5 – seeing the distinct types of occupations and ages



SQL Command #6 – seeing the maximum value in occupations in the dataset, which is a teacher, seeing the maximum value in age which is age 59 (the oldest person in the data), and the maximum value of gender is male