**Roll no:- 220320345051**

**Q1.**

**MapReduce**

**Problem Statement**

Here, we have chosen the stock market dataset on which we have performed map-reduce operations. Following is the structure of the data. Kindlyfind the solutions to the questions below.

Data Structure

1. Exchange Name

2 Stock symbol

3. Transaction date

4. Opening price of the stock

5. Intra day high price of the stock

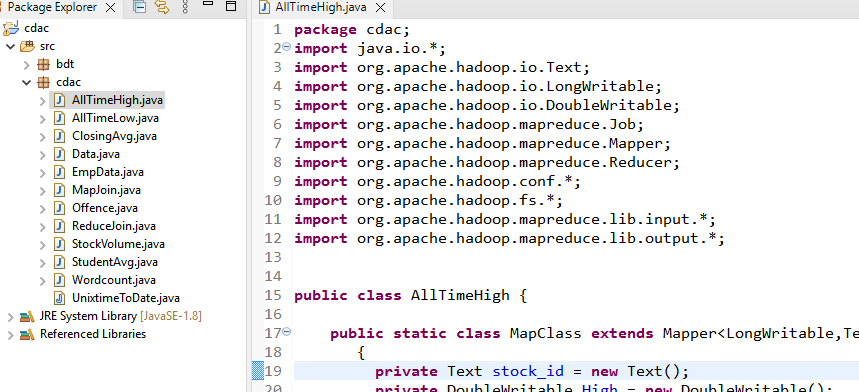
6. Intra day low price of the stock

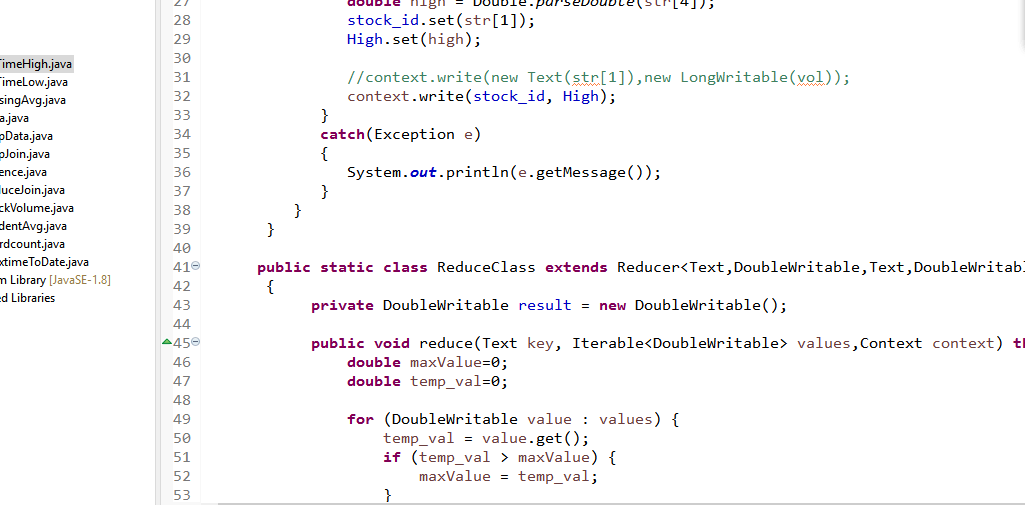
7. Closing price of the stock

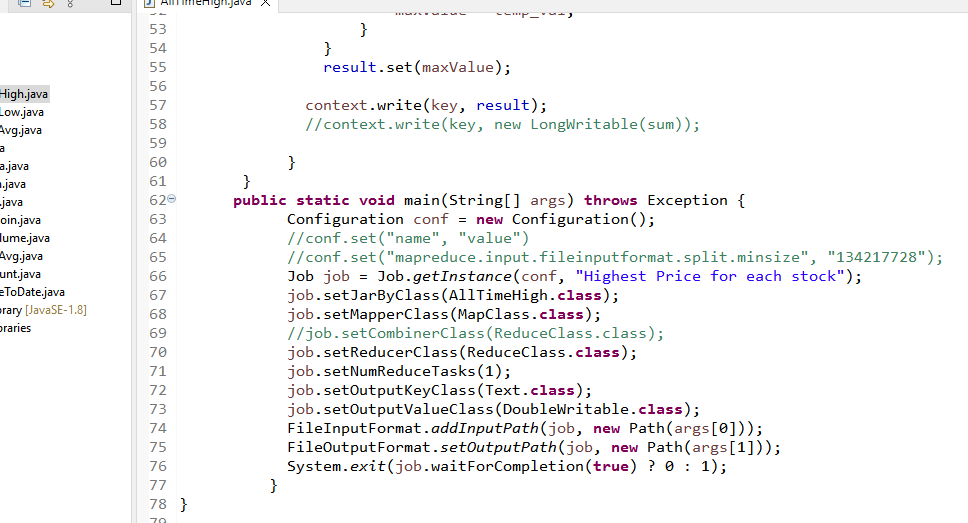
8. Total Volume of the stock on the particular day

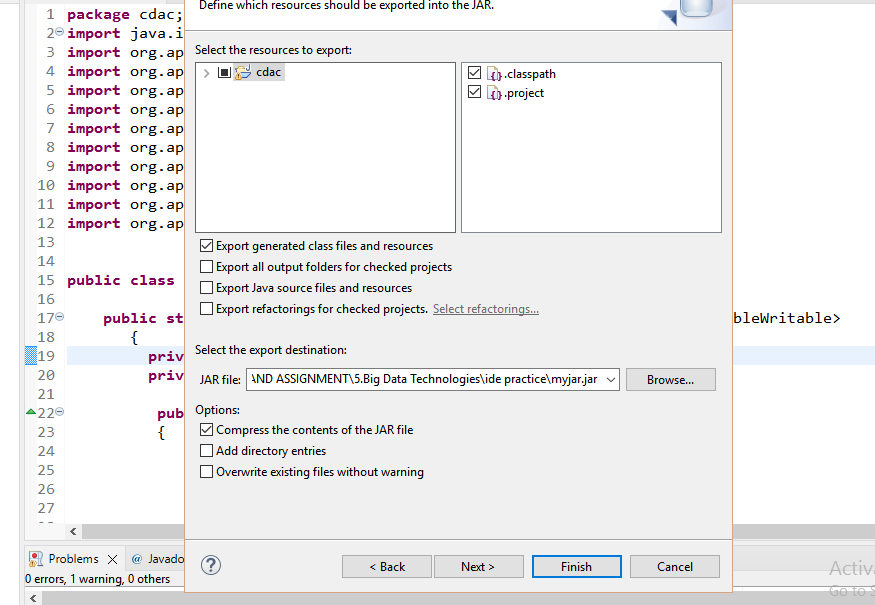
9. Adjustment Closing price of the stock

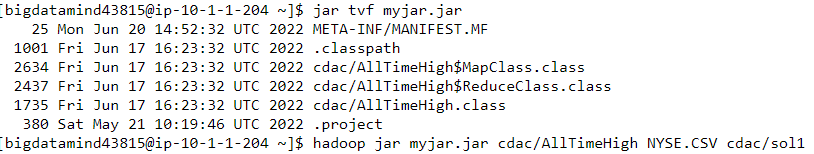
Field Separator – comma











**Question 2 : Find all time High price for each stock**

[15 marks]

**Hive**

Please find the customer data set.

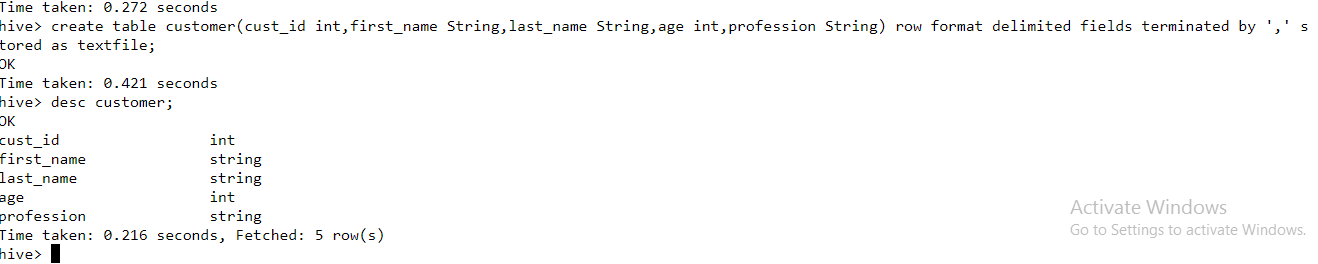
cust id

firstname

lastname

age

profession



**1) Write a program to find the count of customers for each profession.**

Please find the sales data set.

txn id

txn date

cust id

amount

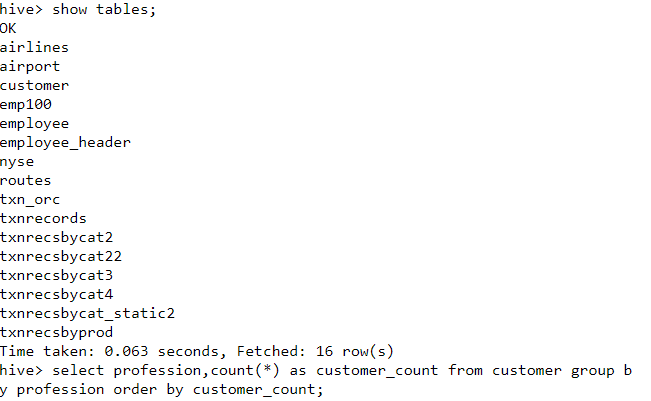
category

product

city

state

spendby



-🡪Output:

Social Worker   1

Writer  101

Artist  175

Environmental scientist 176

Carpenter       181

Dancer  185

Therapist       187

Economist       189

Real estate agent       191

Electrical engineer     192

Nurse   192

Civil engineer  193

Automotive mechanic     193

Psychologist    194

Electrician     194

Agricultural and food scientist 195

Athlete 196

Statistician    196

Judge   196

Doctor  197

Financial analyst       198

Accountant      199

Reporter        200

Secretary       200

Coach   201

Physicist       201

Farmer  201

Actor   202

Architect       203

Computer hardware engineer      204

Teacher 204

Engineering technician  204

Designer        205

Musician        205

Childcare worker        207

Veterinarian    208

Chemist 209

Police officer  210

Recreation and fitness worker   210

Lawyer  212

Social worker   212

Pilot   212

Human resources assistant       212

Pharmacist      213

Computer software engineer      216

Firefighter     217

Librarian       218

Loan officer    221

Photographer    222

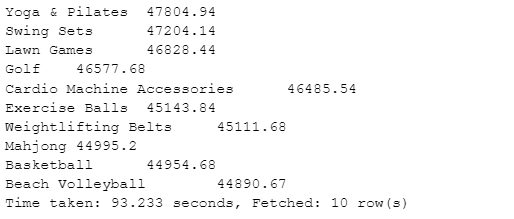
Computer support specialist     222

Politician      228

Time taken: 80.14 seconds, Fetched: 51 row(s)

**2) Write a program to find the top 10 products sales wise**





**3) Write a program to create partiioned table on category**

**QUESTION 3** [15 marks]

**PySpark**

Please find the AIRLINES data set

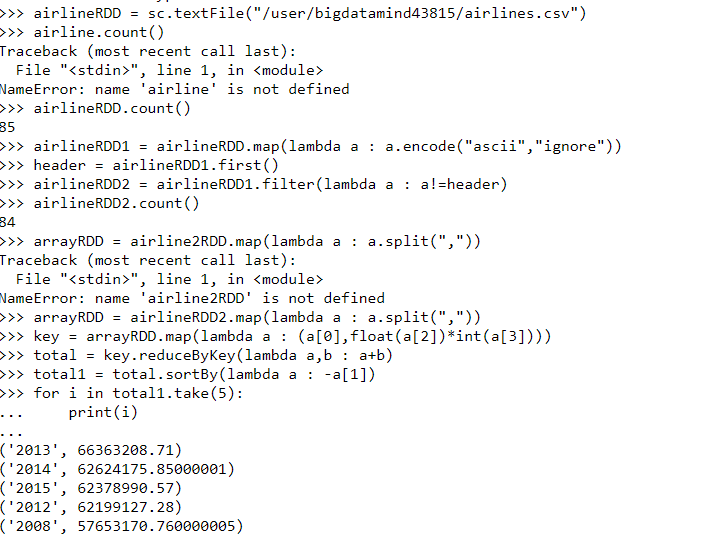
Year

Quarter

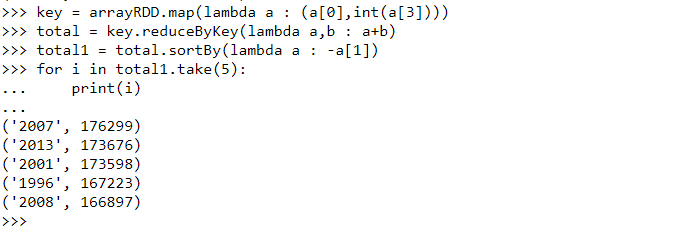
Average revenue per seat

Total number of booked seats

1. **What was the highest number of people travelled in which year?**



1. **Identifying the highest revenue generation for which year**



**3) Identifying the highest revenue generation for which year and quarter (Common group)**

