Question 2 : Find all time High price for each stock

create database end\_module;

use end\_module;

set hive.cli.print.current.db = true;

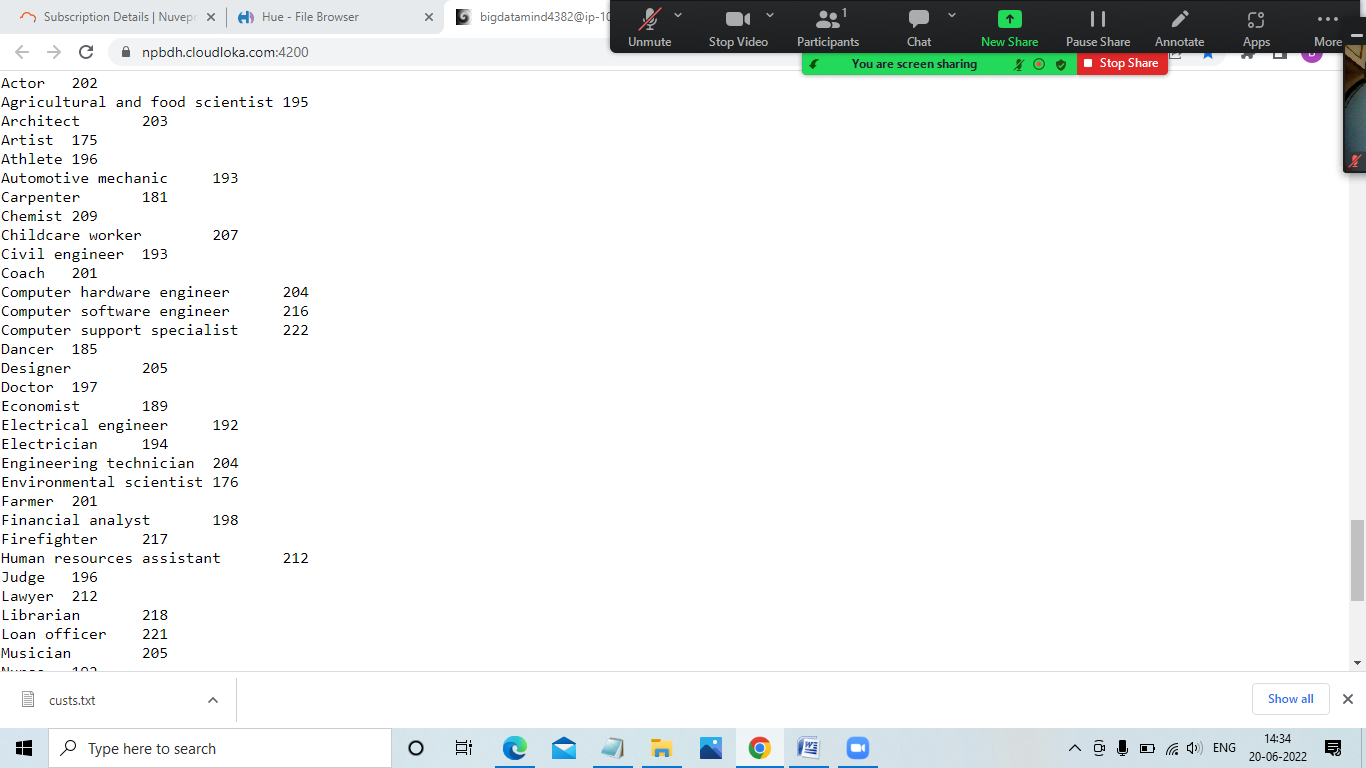
create table customer(cust\_id int,firstname string,las

tname string,age int,profession string)row format delimited fields terminated by ',' stored as textfile;

show tables;

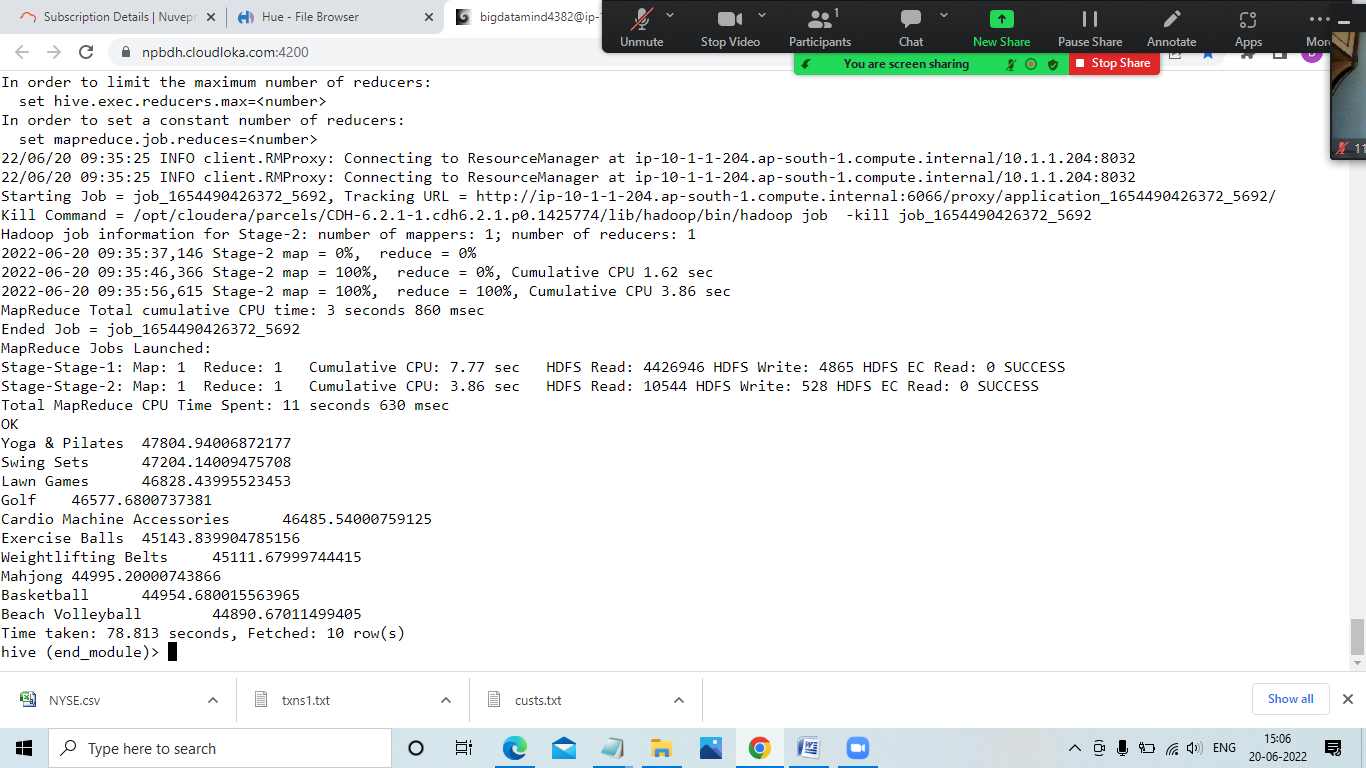
1. Write a program to find the count of customers for each profession.

select profession,count(cust\_id) count from customer group by profession;



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

select product,sum(amount) as total from sales group by product order by total desc limit 10;



1. Write a program to create partiioned table on category

create table sales1(txn\_id int,txn\_date string,cust\_id

int,amount float,product string) partitioned by(category string)row format delimited fields terminated by ',' stored as textfile;

QUESTION 3

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

airrdd = sc.textFile("/user/bigdatamind4382/airlines.csv")

>>> airrdd2 = airrdd.map(lambda a : a.encode("ascii","ignore"))

header = airrdd2.first();

airrdd3 = airrdd2.filter(lambda a : a != header)

for i in airrdd3.take(10):

print(i)

array = airrdd3.map(lambda a : a.split(","))

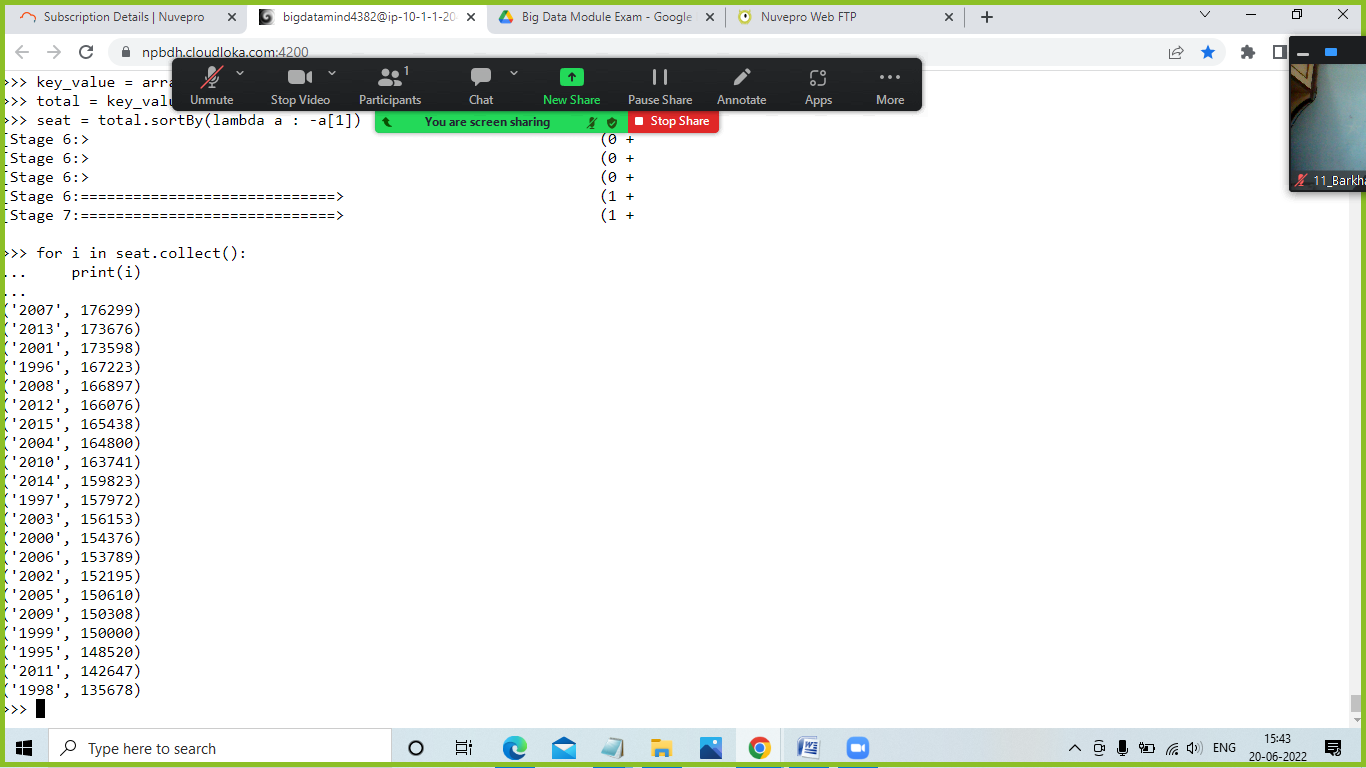
for i in array.take(10):

print(i)

key\_value = array.map(lambda a : (a[0],int(a[3])))

total = key\_value.reduceByKey(lambda a,b : a+b)

seat = total.sortBy(lambda a : -a[1])



1. Identifying the highest revenue generation for which year

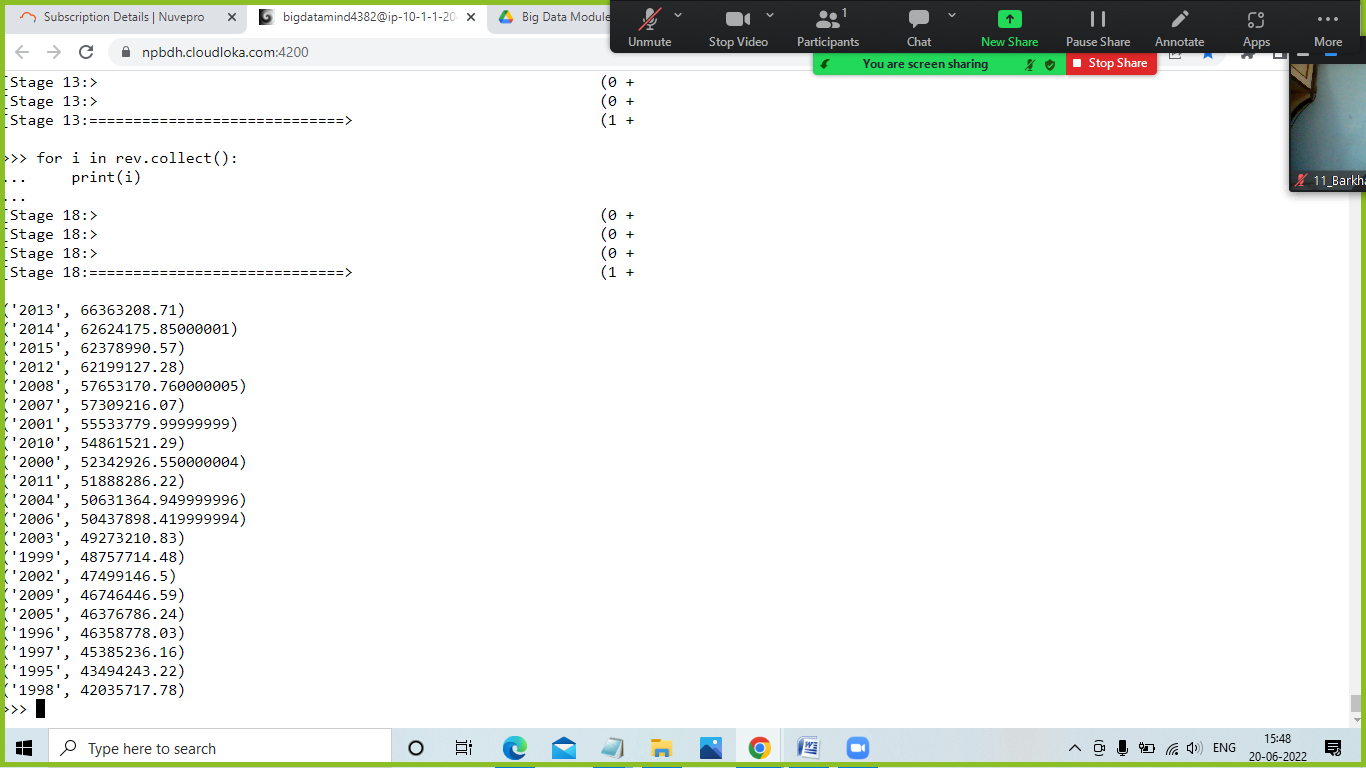
keyvalue = array.map(lambda a : (a[0],float(a[2])\*int(a[3])))

total1 = keyvalue.reduceByKey(lambda a,b : a+b)

rev = total1.sortBy(lambda a : -a[1])

for i in rev.collect():

print(i)



# 

**Q1. MapReduce Problem Statement**

All time high for all stocks -

haddop fs -put NYSE.csv cdac

hadoop jar MyJar.jar AllTimeHigh cdac/NYSE.csv exam/out1

