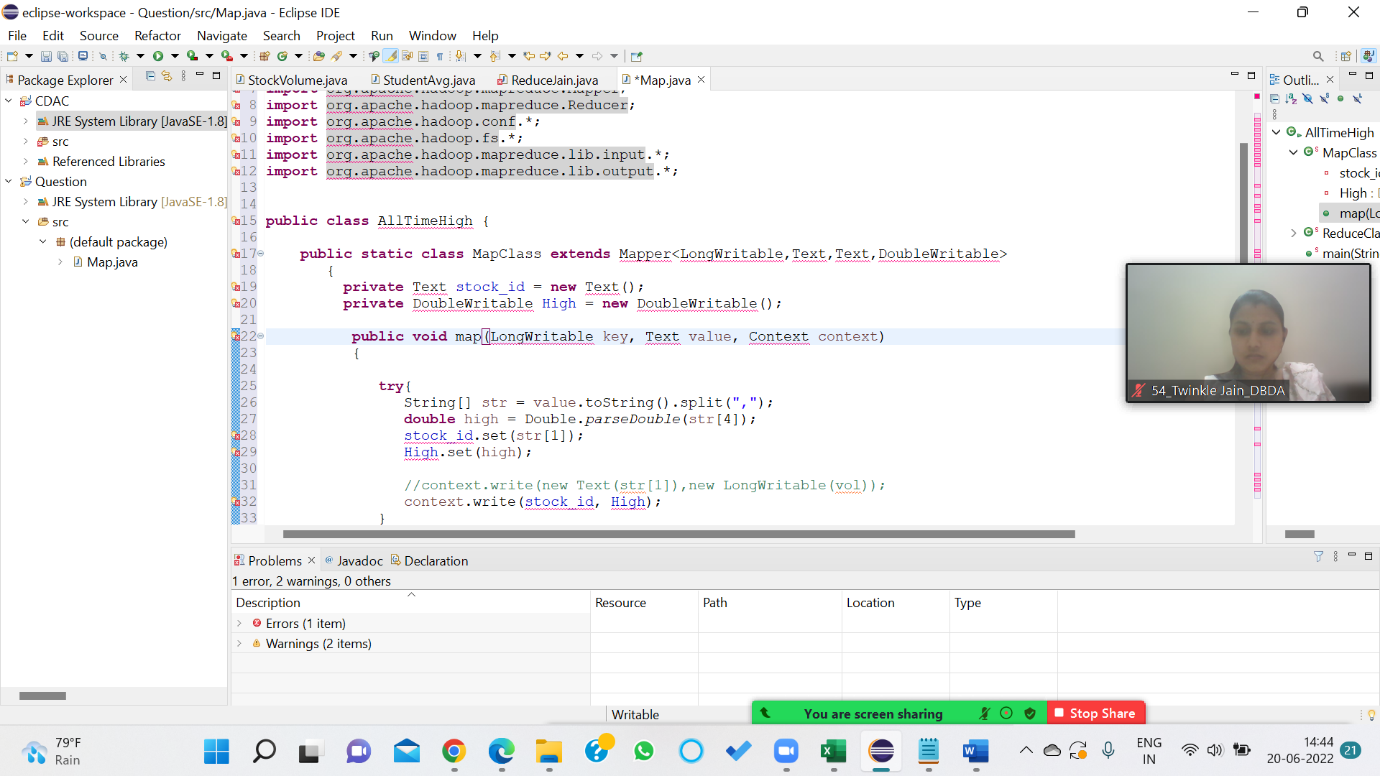
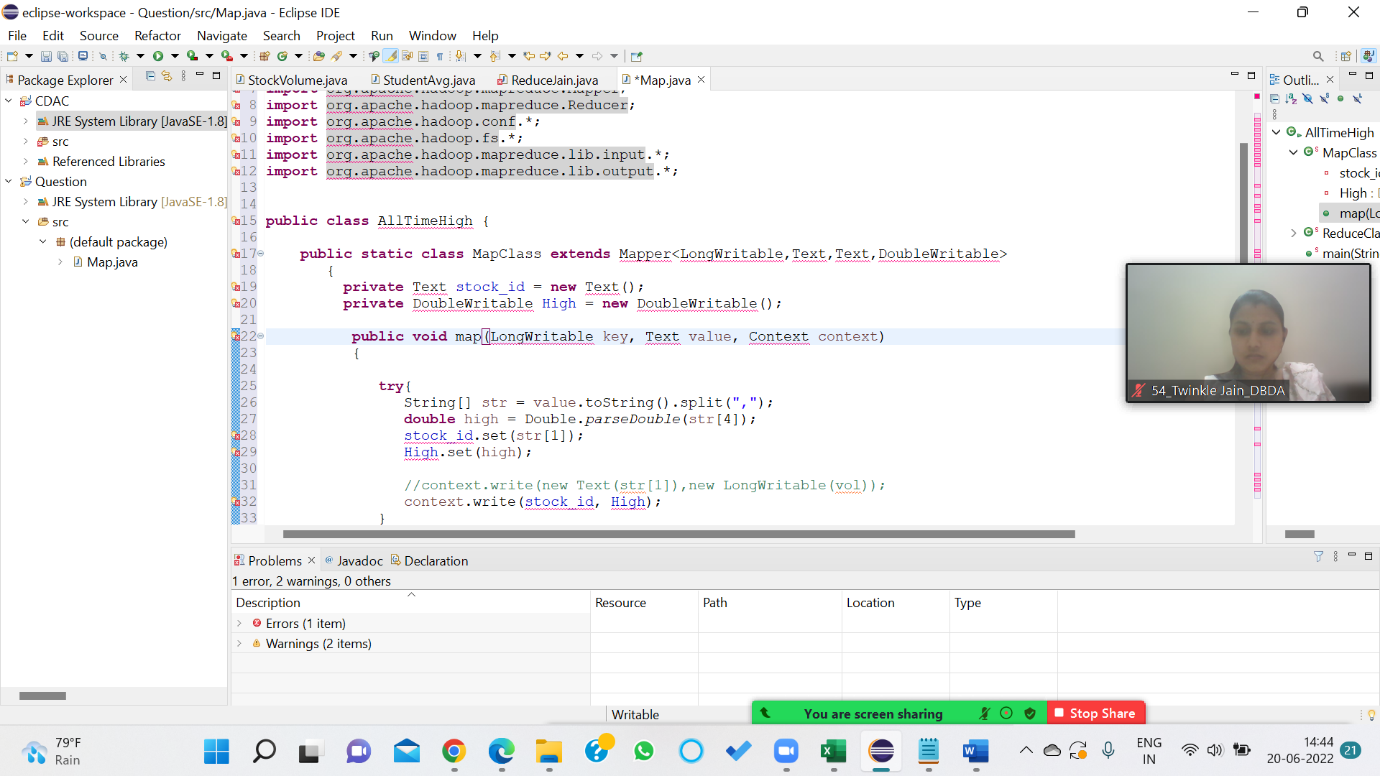
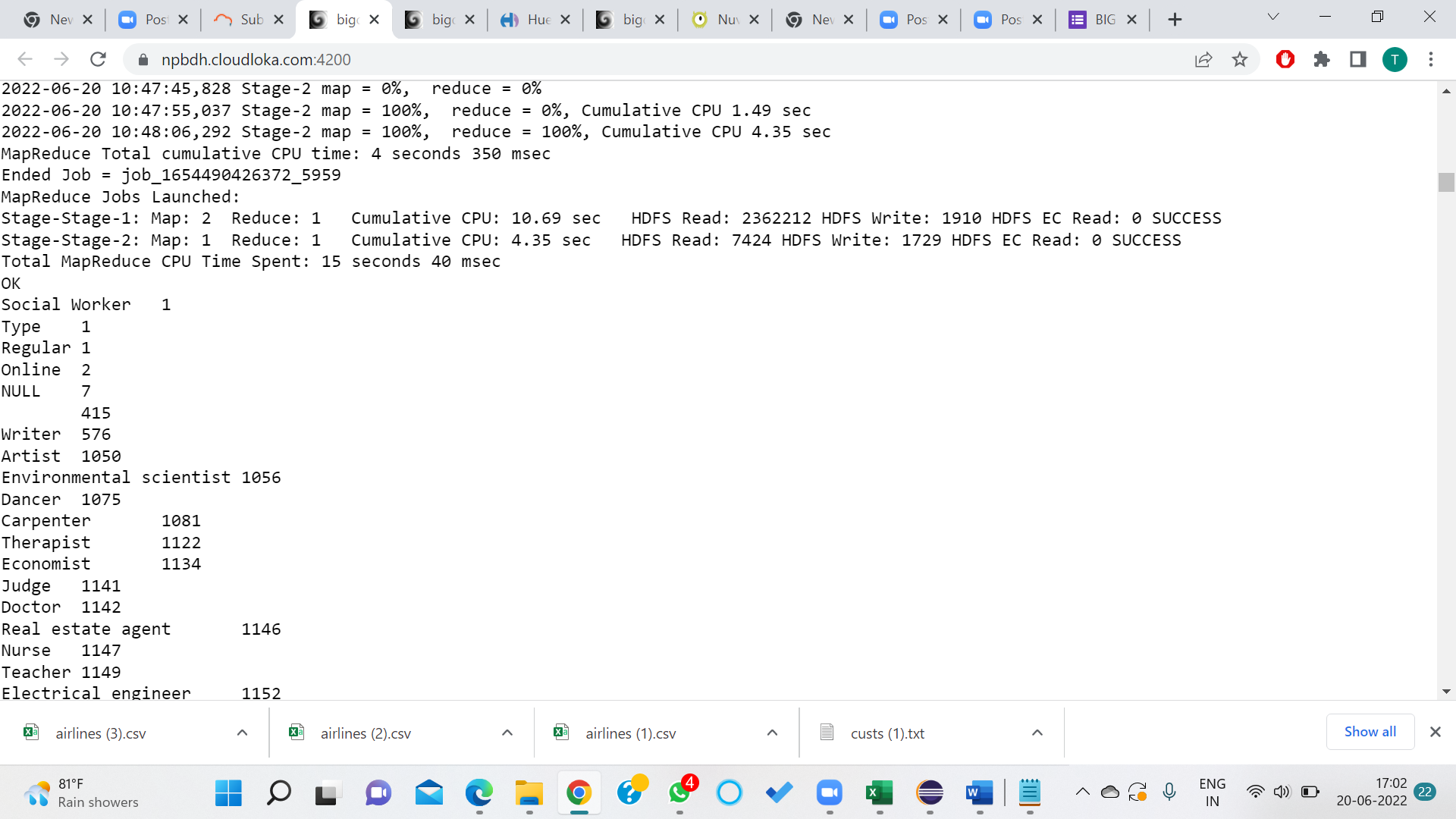
Q1:-> Map reduce

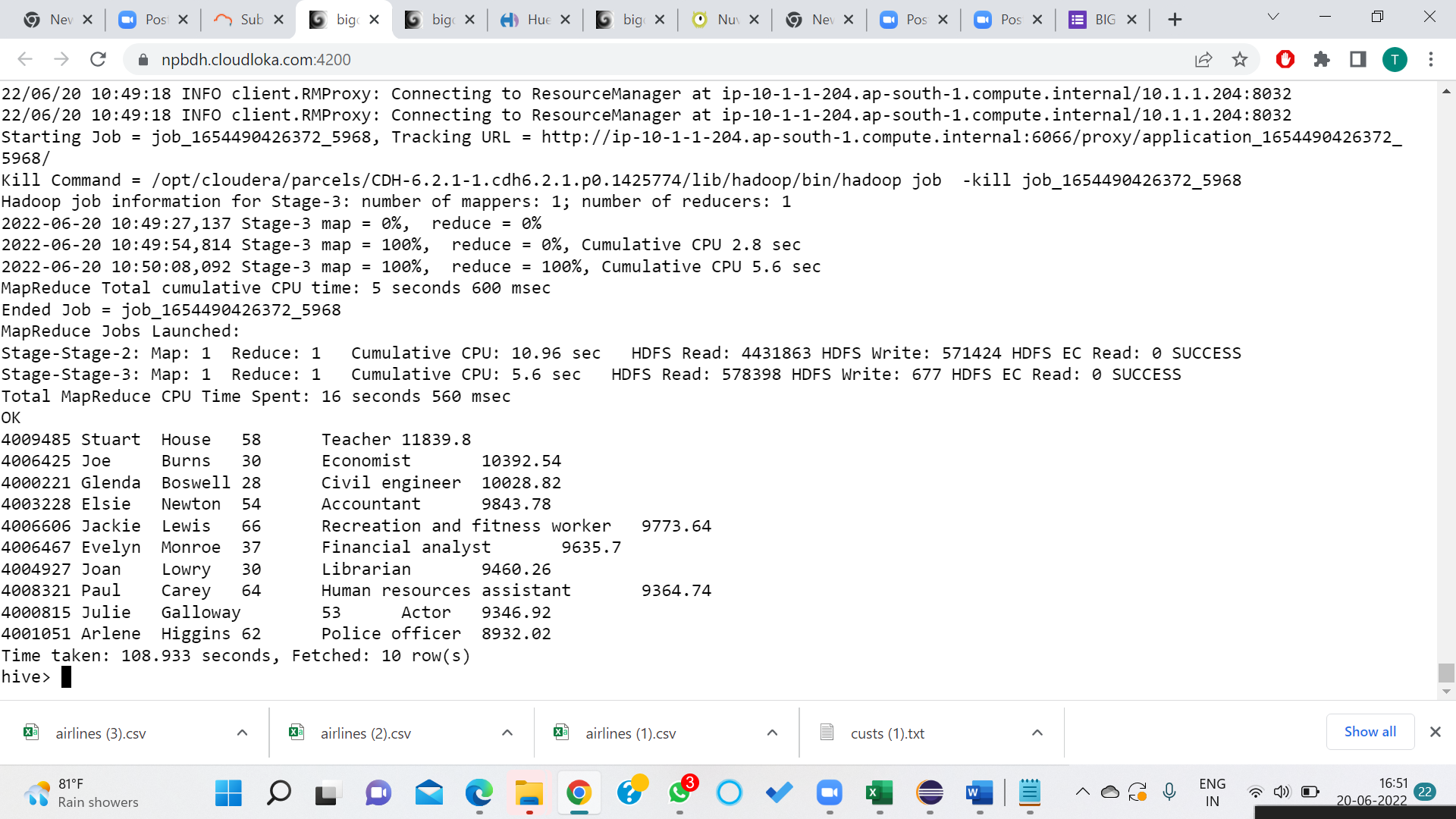




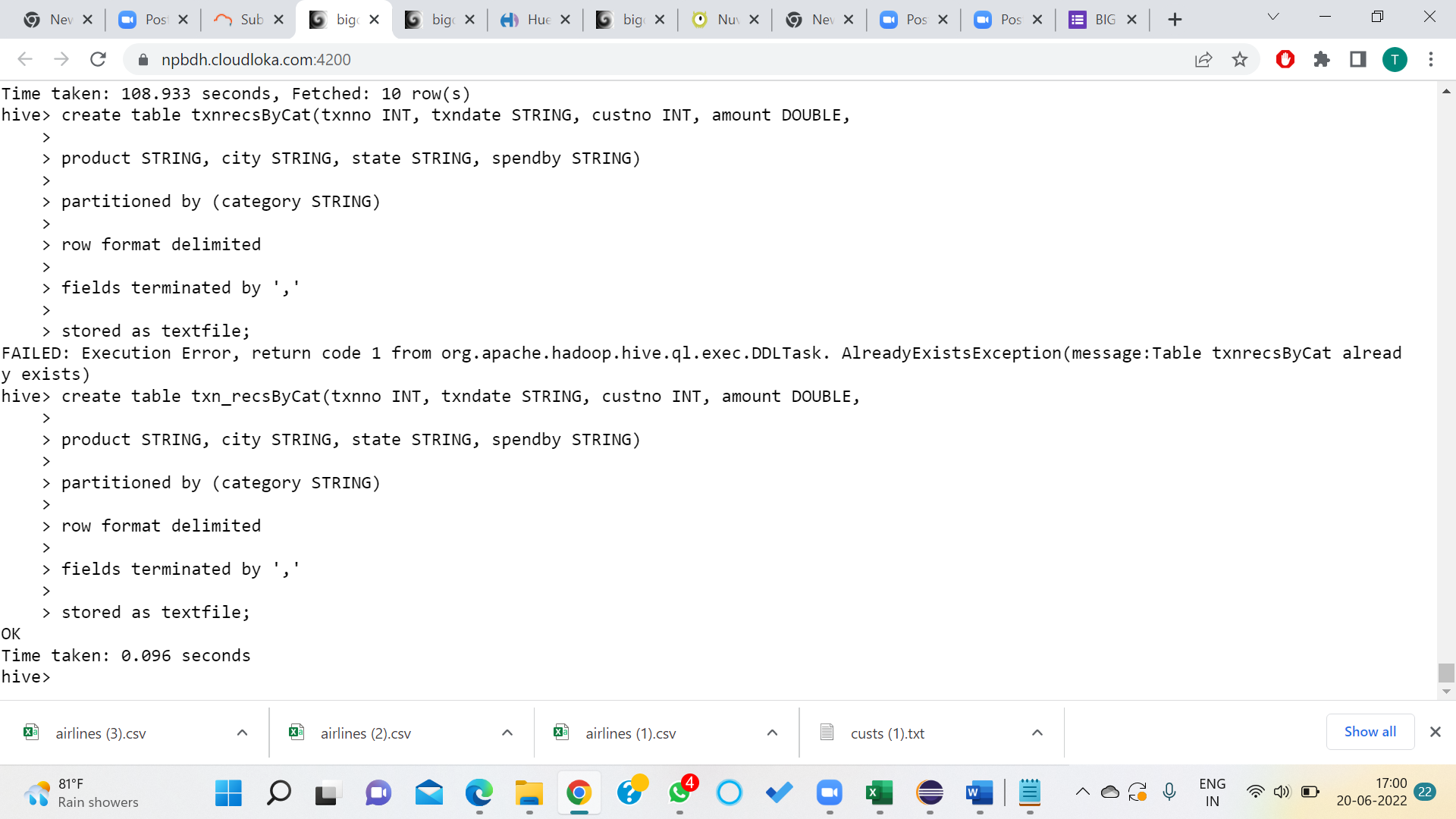
Hive Q1: Write a program to find the count of customers for each profession.



Hive Q2: Write a program to find the top 10 products sales wise



Hive: Q3: Write a program to create partiioned table on category



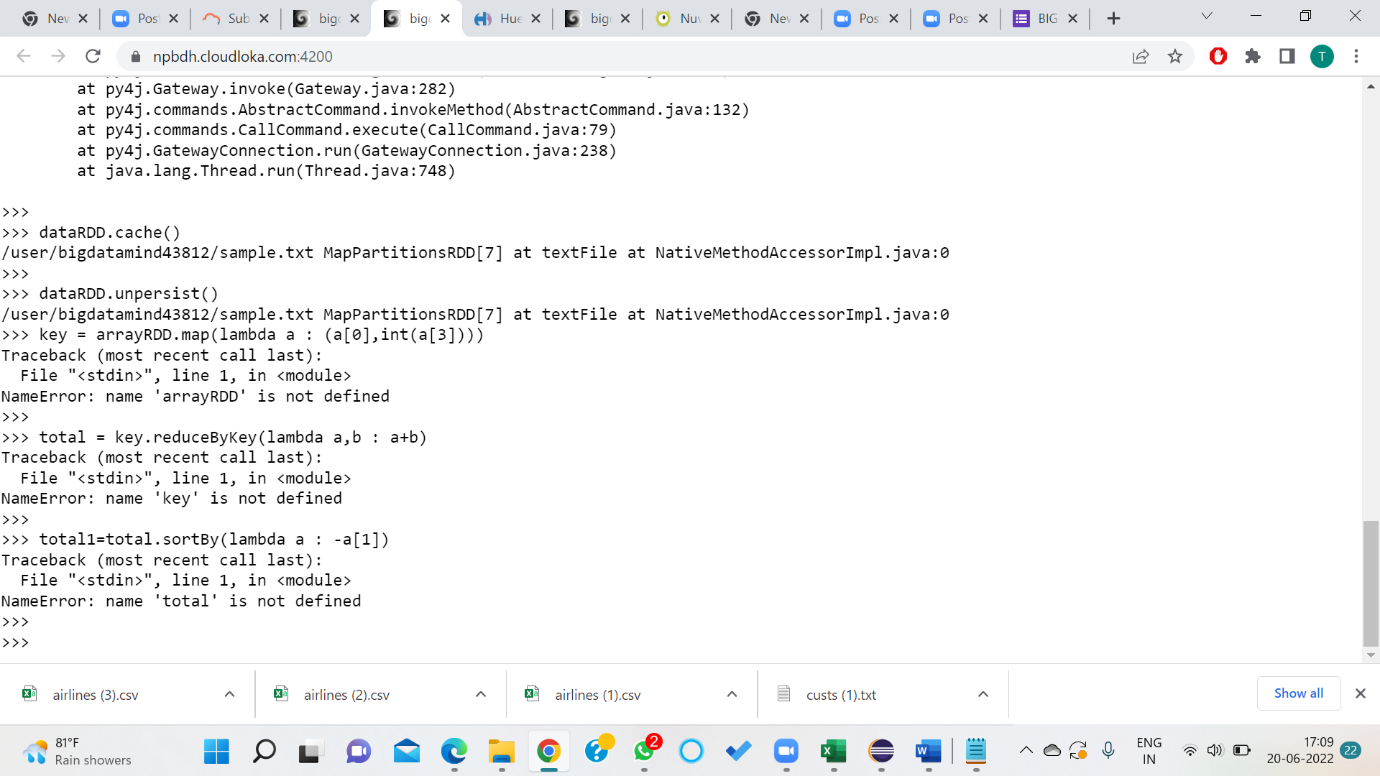
SPARK Q1: What was the highest number of people travelled in which year

key = arrayRDD.map(lambda a : (a[0],int(a[3])))

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

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

)



Q2 : Spark Identifying the highest revenue generation for which year

Ans: key = arrayRDD.map(lambda a : (a[0],float(a[2])\*int(a[3])))

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

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

for i in total1.take(5):

print(i)

Maximum revenue genrated in 2013 ----> 66363208.71

Q3 spark: Identifying the highest revenue generation for which year and quarter (Common group)

Ans:

Data Files ---------> retail--->D01 , D02 , D11 , D12

rdd= sc.textFile("/user/bigdatamind43822/retail")

rdd.count()

rdd1 = rdd.map(lambda a : a.encode("ascii","ignore"))

array = rdd1.map(lambda a : a.split(","))

key = array.map(lambda a : (a[5].strip(), int(a[8]) - int(a[7])))

final = key.reduceByKey(lambda a,b : a+b)

final.count() -------> 23812

loss\_prod = final.filter(lambda a : a[1]<0)

loss\_prod.count() --------> 101

sort = loss\_prod.sortBy(lambda a : a[1])

for i in sort.collect():

print(i)

Answer ---------->Maximum Loss in product---> '4714981010038' , -131002