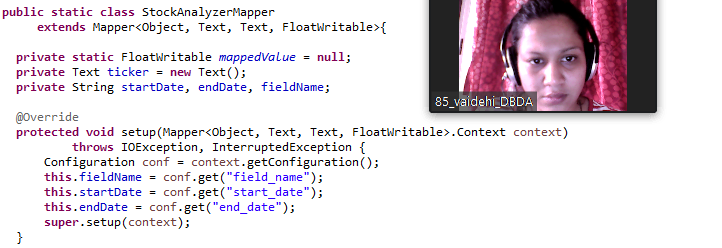
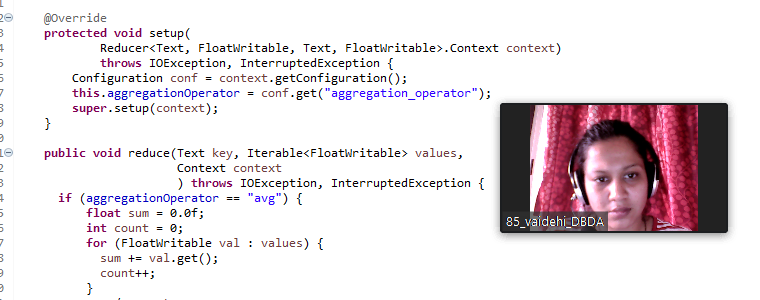
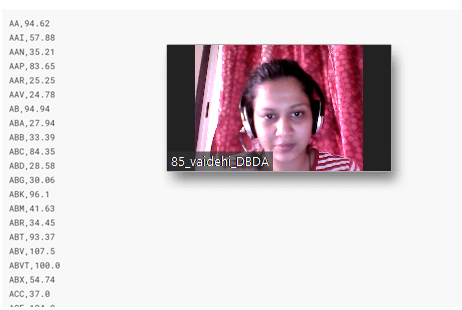
**Q1.**

**MapReduce**

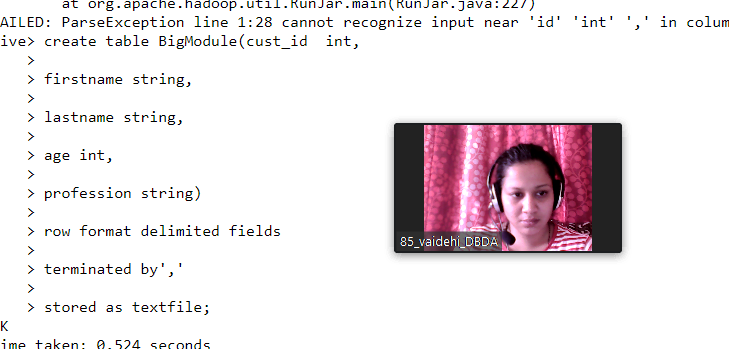
**Problem Statement**

****

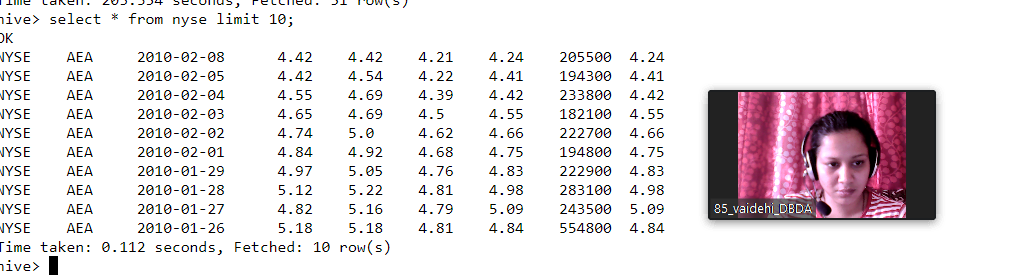
****

****

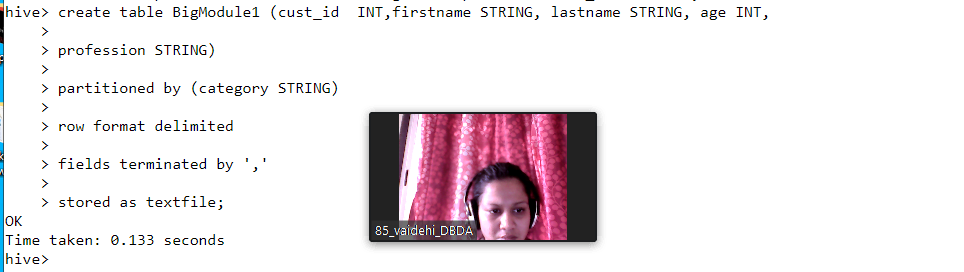
**1) Write a program to find the count of customers for each profession.**

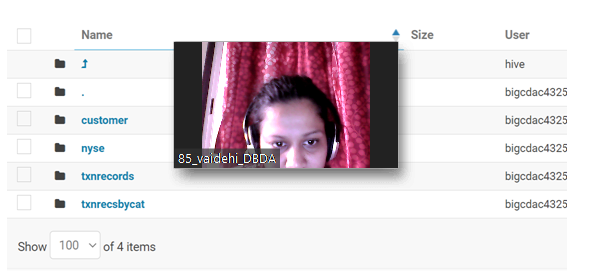


**2) Write a program to find the top 10 products sales wise**



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





3) Write a program to create partiioned table on category

QUESTION 3

PySpark

df.registerTempTable("airlines")

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

select year,sum(Total number of booked seats) as totalSeats from airlines group by year having totalSeats in

(select max(totalSeats) from ((select year,sum(Total number of booked seats) as totalSeats from airlines group by year)abctable));

2) Identifying the highest revenue generation for which year

select year,sum(Average revenue per seat \* Total number of booked seats) as avgRev from airlines group by year order by avgRev desc limit 1;

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

select year,quarter,sum(Average revenue per seat \* Total number of booked seats) as Rev from airlines group by year,quarter order by Rev desc limit 1;