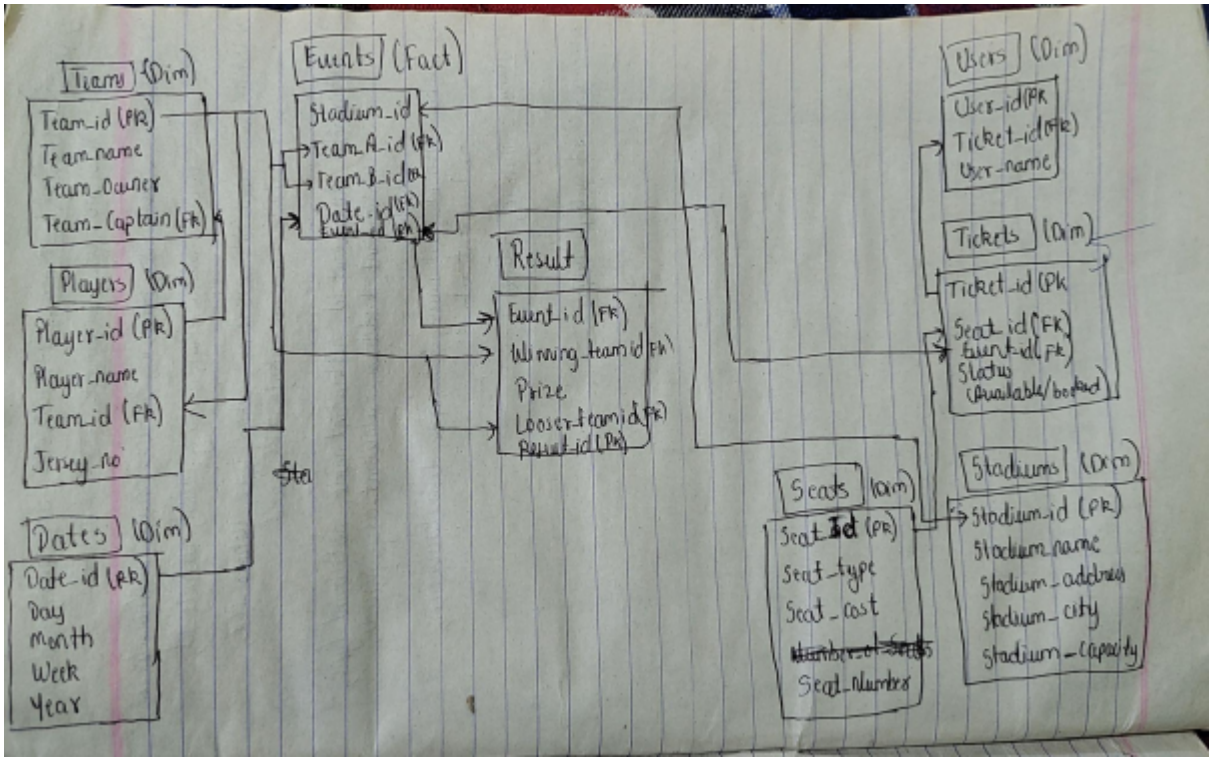
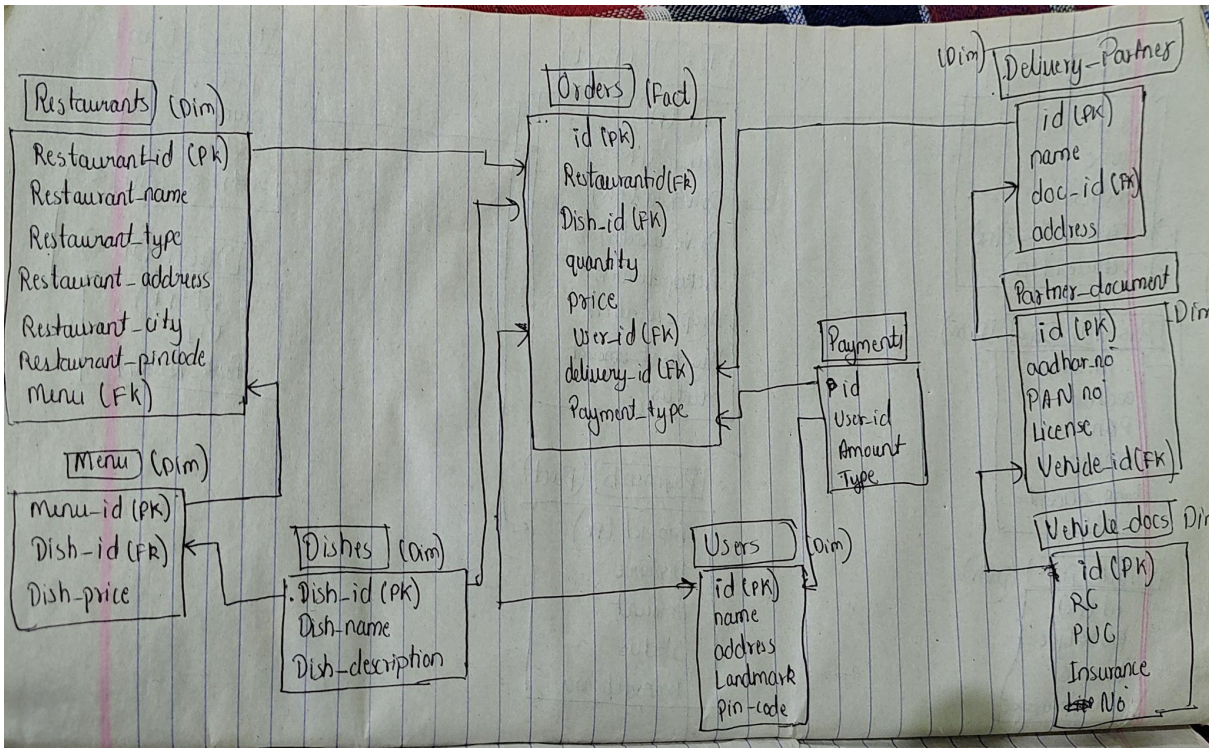
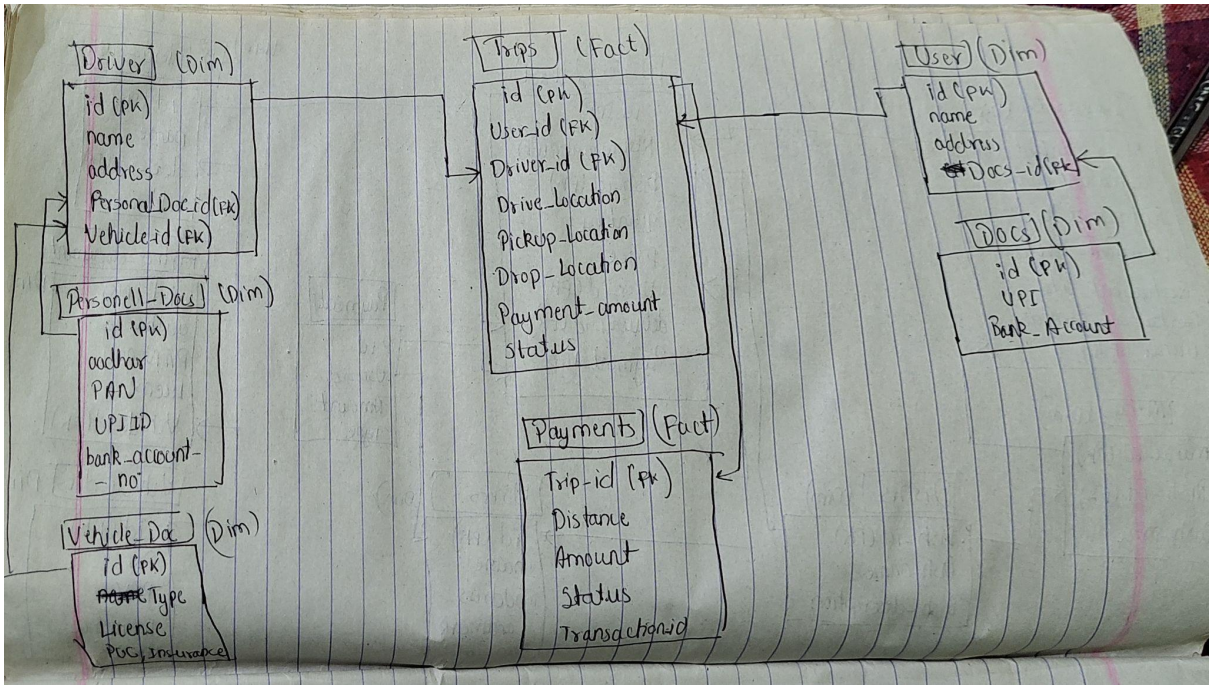
1. Design a Data Warehouse for IPL Cricket Tournament (Asked in Flipkart Interview for Senior Data Engineer role)



2. Design a Data Warehouse for Food delivery app like Swiggy, Zomato (Asked in Grab for Data Engineer role)



3. Design a Data Warehouse for cab ride service like Uber, Lyft (Asked in Google for Data Engineer role)



Data Warehouse Assignment

Scenario I
Use option C

1. What are the strengths and weaknesses of each option?
Ans. Option A

Disadvantages: it would require us to do a join of both tables to get our required results, So this would cause a efficiency issue. In case we have to update the instructors table and the course taught by two teachers, we have to update everywhere that particular teacher is involved.

Advantages: Since the two tables are in different tables, there would be no confusion.

Option B

Disadvantages: This would cause redundant data to exist in the table. We are keeping unnecessarily repeated data which would inefficient operations

Advantages: Since all data is in a single table there is no need for joins.

Option C

Disadvantages: The same data is present in two tables which would mean more storage required for same data.

Advantages: This option would be a better choice as we can segregate operations based on need. There is no need to join them as the queries can be performed independentley

2. Which option would you choose and why?
Ans. I would choose option C because of the advantage it has of performing queries for each requirement. We can give access to a certain table to a certain department.

1. Would your answer to Question 2 be different if the majority of classes had multiple instructors? How

   about if only one or two classes had multiple instructors? (Explain your answer.)

Ans. I would still choose option C as it can handle both the cases.

SCENARIO 2

1. What are the strengths and weaknesses of each option?

Anse. Option A: This would save storage as only one table is being overwritten but disadvantage would be that the old scores are over written so there would be no way to know the old score and compare it with new score.

Option B: This would be a better option because all the scores are in a single dim table and we can perform required operations in a single table but there would be no way to differentiate them without the dates.

Option C: This is the best option as the required queries can be performed using a single join and data is differentiated into different table.

Option D: This is not a better option as we have to join three tables in order to get the proper result. The advantage is that all data is segregated properly

1. Which option would you choose and why?

Ans. I would choose option C as it would mean less joins, less confusion and the queries can be performed using a single join

1. Would your answer to Question 6 be different if the number of customers and/or the time interval between score recalculations was much larger or much smaller? (Explain your answer.)

Ans. No, my answer would remain the same.