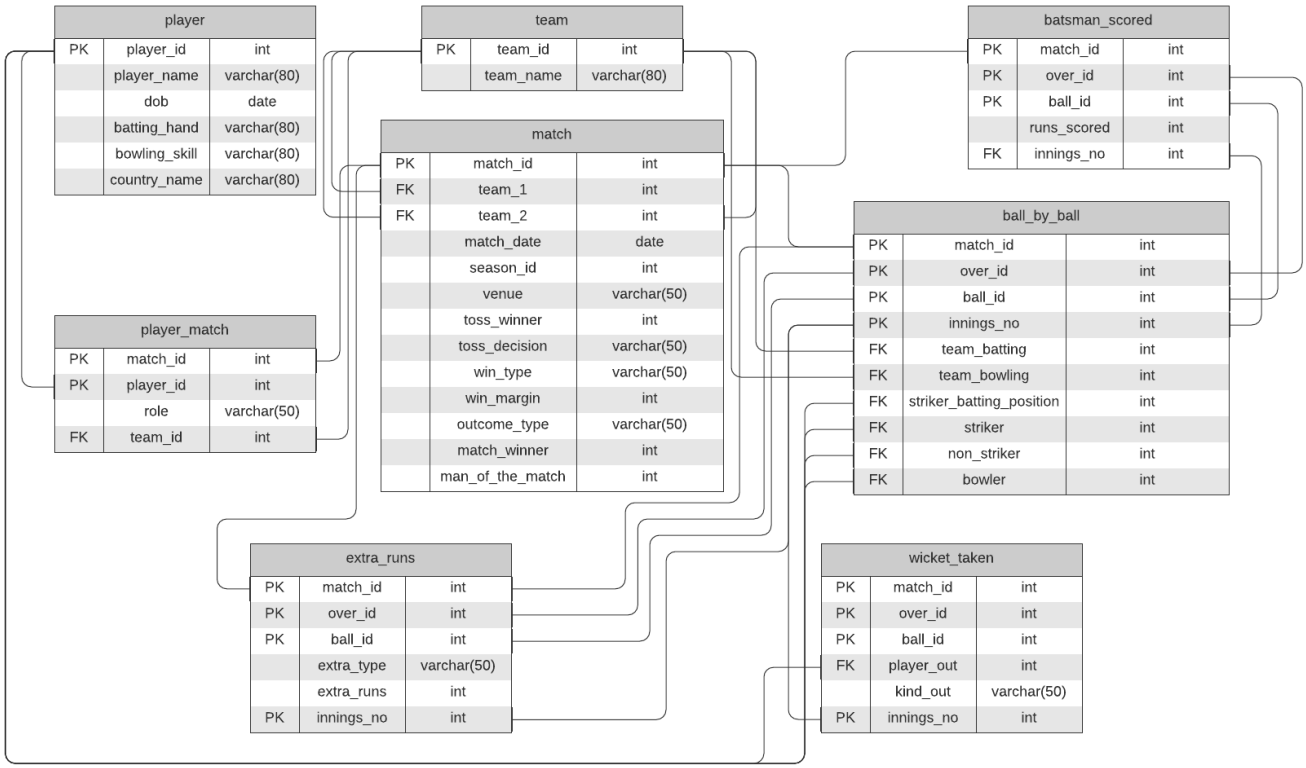
Mohamed Shamir | 17110084

Contents

Database Schema

The Schema for the following problems is as shown below. The integrity constraints are also mentioned in the diagram.



Creating the database

- Python Scripts are used to generate sql insertion codes.
- Python Script can be accessed using this Link
- All the sql insertion codes are avalible in the insert directory.

```
├── create_table.sql
├── data_insert.ipynb
├── dataset
│   ├── ball_by_ball.csv
│   ├── batsman_scored.csv
│   ├── extra_runs.csv
│   ├── match.csv
│   ├── player.csv
│   ├── player_match.csv
│   ├── team.csv
│   └── wicket_taken.csv
├── images
│   └── schema.png
├── insert
│   ├── ball_by_ball.sql
│   ├── batsman_scored.sql
│   ├── extra_runs.sql
│   ├── match.sql
│   ├── player_match.sql
│   ├── player.sql
│   ├── team.sql
│   └── wicket_taken.sql
```

```
├── readme.md
└── run_all.sh
```

- To create the database and insert all the data. Run run_all.sh

```
source run_all.sh
```

Problem 1

Write SQL queries for the following questions. Questions ‘a’ to ‘g’ carry 2 marks each. Questions ‘h’-‘j’ carry 1 mark each. (17 marks+5 marks)

Note: Required output attribute(s) are given next to each query, also export each output in Q1X.csv, where X is a,b...j. Any deviation from the given format would result in zero marks.

a. For all the matches_id(entire IPL), find the minimum runs scored in any over and the bowler who bowled that over. Sort by increasing match_id, followed by increasing innings_no, then finally by increasing over_ids. Output:

Note: Runs scored in an over is the sum of the batsmen_scored+ extra_runs(wides and "no_balls" only. It should not be match specific)

Solution:

```

```

b. Find the names of all the batsmen(players) and the frequency of their “caught” out in increasing order of the number of “caught”. If a tie occurs, sort names alphabetically. Hint: Frequency can be 0 too.

c. List the stadium(s) where the maximum number of “legbyes” (runs) is taken. If ties occur, show alphabetical order. <venue_name><number_of_legbye_runs>

d. Find the bowler(s)(players) who has the best average(no. of runs given/wickets taken) in edition 5. If a tie occurs, sort names alphabetically. <bowler_name>

e. Find out the names of all batsmen(players) who scored more than 100 runs in a match and, their runs scored. Sort names alphabetically. (if multiple entries of the same player, show the one with the highest runs).<batsmen_name>

f. Find out the top 3 batsmen(players) whose [number of runs scored/number of matches played] is the best in edition 2. Sort alphabetically. <batsman_name>

g. Find out the batting average(as calculated in the above question (f)) of all players. Then only show the list of the top 3 countries with the highest country batting average(∑batting average/Total number of

players in that country)

h. Write down a simple query to make a copy of the player table(with data).

i. Using view, create a table say “Indian Players” which contains information about the total runs scored by all the Indian players till now and sort them alphabetically.

j. List all captains who scored more than 50 runs in edition 3.

Problem 2

Suppose a user creates a new relation r1 with a foreign key referencing another relatio>n r2. What authorization privilege does the user need on r2? Why should this not simply be allowed without any such authorization? (max 500 words) (4 marks)

Problem 3

Explain the difference between integrity constraints and authorization constraints. (explain them with examples) (max 500 words) (4 marks)

Problem 4

Consider a set of users A, B, C, D, and E. Suppose the user A creates a table T and thus is the owner of T. Now suppose the following set of statements is executed in order:

1. User A: grant select on T to B, C with grant option
2. User B: grant select on T to C
3. User C: grant select on T to D, E
4. User A: grant select on T to E
5. User A: revoke select on T from B restrict
6. User A: revoke select on T from C cascade