Assignment 1

Due: Friday 7 February at 10:00 pm

Introduction

For this assignment, you will be writing queries in relational algebra on a database. We are providing the schema for you. Later in the course, you will learn about how to develop your own schema based on knowledge of the domain. Even though developing a schema is really the first step in building a database, it is a more advanced task than querying an existing database; this is why we will be learning about it and practising it later.

Schema

The schema below represents data for World Cup 2014, where people can buy tickets for matches they would like to attend. It attempts to represent the domain well, but in order to keep the assignment manageable, some things are simplified and other details are not represented. For example, the different stages of the competition are ignored and only matches between any two countries are considered.

Remember that all your queries will be written with respect to the schema below. Your goal is to produce answers that are correct with respect to this schema, even if they aren't quite correct in the real world.

Relations

- Team(<u>country</u>, coach)
  A tuple in this relation represents the team of a country that participates in the competition. *country* is the team's country name, and *coach* is the name of the team's coach.

- Player(<u>PID</u>, fname, lname, position, goals, country)
  A tuple in this relation represents a football player who is participating in the contest. *PID* is the player's ID, *fname* is their first name, *lname* is their last name, *position* is the position they are playing on the pitch, *goals* is the number of goals the player has scored so far in the competition, and *country* is the country of the team they are playing for.

- Stadium(<u>SID</u>, capacity, city)
  A tuple in this relation represents a stadium where a match takes place. *SID* is the stadium's ID, *capacity* is the capacity of the stadium, and *city* is the city where the stadium is located.

- Match(<u>MID</u>, date, time, SID)
  A tuple in this relation represents a match. *MID* is the id of the match, *date* is the date the match is scheduled on, *time* is the time the match is scheduled on, and *SID* is the SID of the stadium where the match takes place.

- Ticket(<u>TID</u>, dateIssued, timeIssued, MID)
  A tuple in this relation represents a ticket that was purchased. *TID* is the ticket's id, *dateIssued* is the date that it was purchased, *timeIssued* is the time that it was purchased, and *MID* is the MID of the match it was purchased for.

- Competes(<u>MID</u>, country1, country2, goals1, goals2)
  A tuple in this relation represents "who plays against whom" in a match. *MID* is the id of the match, *country1* is the country of the first team, *country2* is the country of the second team, *goals1* is the number of goals scored by the first team, and *goals2* is the number of goals scored by the second team.

Integrity constraints

- Player[country] ⊆ Team[country]

- Match[SID] ⊆ Stadium[SID]

- Ticket[MID] ⊆ Match[MID]

- Competes[MID] ⊆ Match[MID]

- Competes[country1] ⊆ Team[country]

- Competes[country2] ⊆ Team[country]

Part 1 [20% - 4 marks each]: Additional Integrity Constraints

Below are some additional integrity constraints on our schema. Express each of them using the notation from Section 2.5.1 of your textbook. If a constraint cannot be expressed using this notation, simply write "cannot be expressed".

1. No team can play against itself.

2. All tickets for a match have to be purchased before the time of the match.

3. The number of tickets purchased for a match should not exceed the capacity of the stadium where the match takes place.

4. A coach can only coach one team.

5. A player's position should be one of 'G', 'D', 'M' or 'S' representing a goalkeeper, defender, midfielder or striker, respectively.

Part 2 [80% - 10 marks each]: Queries

Write the queries below in relational algebra. There are a number of variations on relational algebra, and different notations for the operations. You may use only the operations defined in Section 2.4 of the text and you must use the same notation as in the textbook. In particular, assume all relations are sets (not bags), and do not use any of the extended relational algebra operations from Chapter 5 (for example, do not use the extended projection).

Some of the queries cannot be expressed in the language that you are using. In those cases, simply write "cannot be expressed". Later in the course, when we learn SQL, it will be interesting to consider whether they can be expressed in SQL.

You are encouraged to use the assignment operator (:=) to define intermediate results, and it is a good idea to add commentary explaining what you're doing before each expression. This way, even if your final answer is not completely correct, you may receive partial marks.

You should assume that both the original constraints enforced by the schema and **the additional ones** given in Part 1 hold for the following queries. However, do not make any assumptions other than those. Your queries should work for any database that satisfies those constraints.

Note: These queries are not in order according to difficulty.

1. Report the country of the team that has played in every stadium. If there are ties report all of them.

2. Report the MID of the match for which the highest number of tickets was purchased. If there are ties report all of them.

3. Report the PID(s) of the player(s) of the team(s) that didn't play in any match.

4. Report the SID(s) of the stadium(s) where exactly one match took place.

5. Report the coaches of the teams with the highest difference in the number of goals when competed with each other at a match. If there are ties, report all of them.

6. Report the fname and lname of the players whose position is 'D' and have scored the highest number of goals among all players (in any team) who play at the same position.

7. Find the winner country of the match for which the very first ticket out of all the tickets in the database was purchased. If there was a tie in the match, report nothing.

8. Report the fname and lname of the player of the country 'Spain' with the second highest number of goals among players of the same country.

9. Report the MID(s) of the matches for which at least two tickets were bought on the date of the match.

10. Consider all teams that have won at least one match. For each of those teams, report its country, the position of its player with the highest number of goals and the nubmer of goals he has scored.

Submission Instructions

Your assignment must be typed; handwritten assignments will **not** be marked. You may use any word-processing software you like. Many academics use LaTeX. It produces beautifully typeset text and handles mathematical notation well. Whatever you choose to use, you need to produce a final document in *pdf*. On the first page of your *pdf* you should declare:

- your team (whether it is a team of one or two students) including *cdflogin*, *first name*, *last name* and *student number*.

- the number of any *grace days* you used for this assignment.

Important Notes:

- If you are working in a pair, only one of you should submit.

- You should submit your document **both** *electronically* **and** as a *hard copy* in class.

Step 1: Electronic Submission

First, you should submit your work electronically using the *submit* command in CDF. For this assignment, you will hand in just one file named **A1.pdf**. When you have completed the assignment, move your **A1.pdf** in CDF, and use the following command to electronically submit your file:

% submit -c csc343h -a A1 A1.pdf

You may submit your solution as many times as you wish prior to the submission deadline. Make sure you name your file exactly as stated (including lower/upper case letters). You may check the status of your submission using the command:

% submit -l -c csc343h -a A1

Once you have submitted, be sure to check that you have submitted the correct version; new or missing files will not be accepted after the due date.

Step 2: Hard-copy Submission

Then, you should print and hand in a hard copy of your **A1.pdf** report during your lecture class on Tuesday or Wednesday after the due date. This should be identical to your electronic submission. **Reports that are not handed in will not be marked**.