Homework - 1

COL362 - Introduction to DBMS

Due Date: 11:59 PM January 23, 2023

Instructions

Instructions

1. This homework is **not** graded. But it is **mandatory** to submit the homework.
2. All work has to be done individually – this is not group activity.
3. Command to run .sql files: `psql \i path/to/file.sql`.
4. Submit Entry-Number.txt and Entry-Number.sql files. For 'Installation and Compilation' part and Exercise - 1, submit your output in Entry-Number.txt (a single PDF file). For Exercise - 2, submit Entry-Number.sql file. Follow the given format for .sql file. Leave an empty line after each query
   -1-
   SQL Query-1
   -2-
   SQL Query-2
   -3-
   SQL Query-3
5. **Use only piazza for your queries. Don't contact TAs through any other platform. Tag posts related to the homework with "homework1" tag.**

1 Installation and Compilation

In this part of the homework, you will be required to download, compile and install the PostgreSQL 8.3.23 database management system on your machine. Please make sure you do not have any previous installation of PostgreSQL in your machine before you start this – if you have, you must uninstall.

1. Download the source files of PostgreSQL 8.3.23 from `https://www.postgresql.org/ftp/source/v8.3.23/`.
2. Make sure you have the clang 14 compiler on your system. You can install it easily on Linux-like systems (it is default in Mac). For Windows 10 users, we have tested using Windows Subsystem for Linux (WSL). If you have a different set up, your mileage may vary.
3. Compile/build from sources **PostreSQL 8.3.23** using the instructions provided in the documentation **here**. You also have installation instructions in the INSTALL document in the sources you have downloaded. Follow the instructions carefully.
   - You can install PostgreSQL in the default directory. But if you would like, you can also install it in as separate directory (for ease of removal later), by passing appropriate arguments to the `configure` script. Please read the configure help for more details.

- On some machines – namely, the Mac machines with M1/M2 chip – you may get an error saying the platform does not support spinlocks. You can turn it off by passing the appropriate argument to the `configure` script. On all other machines you will not have this issue.

4. After successful installation, run `psql` in the terminal and in the prompt run query: `SELECT version();` and submit the output in the report. Subsequently, all the SQL queries should be run in psql prompt.

2 Dataset Description

We all still fondly remember the exciting finals game of FIFA WC'22 between Argentina and France which is considered one of the greatest football matches ever. In this homework, we take a little bit of history tour and go back in time to look at the past FIFA WCs.

This is a comprehensive database about the FIFA World Cups, that covers **21** World Cup tournaments (1930-2018) with 7 tables. The description is given only for a subset of columns. Other columns are self-explanatory. The schema of the database can be found **here**

1. **players**: This table has details about the players who participated in FIFA WCs.
2. **goals**: This table records all goals.

Column	Description
team_id	The unique ID number for the team that scored the goal. For own goals, this is the team that is awarded the goal, not the team of the player who scored the own goal
player_id	The unique ID number for the player who scored the goal
player_team_id	The unique ID number for the team of the player who scored the goal

3. **matches**: This table records all World Cup matches.

Column	Description
stage_name	The stage of the tournament in which the match occurred
home_team_id	The unique ID number for the home team
away_team_id	The unique ID number for the away team
home_team_score	Number of goals scored by the home team
away_team_score	Number of goals scored by the away team
home_team_score_penalties	The score of the home team in the penalty shootout
away_team_score_penalties	The score of the away team in the penalty shootout

4. **penalty_kicks**: This table records all penalty kicks taken during penalty shootouts. This table does not include attempted penalty kicks during matches.

Column	Description
converted	Whether the penalty kick was converted

5. **stadiums**: This table records all stadiums that have hosted a World Cup match.
6. **teams**: This table records all teams who have participated in a World Cup match.
7. **tournaments**: This table records all World Cup tournaments.

3 Loading the Dataset

The 'data' folder provided has 7 csv files. Run 'create_table.sql' (also provided in data folder) to create the tables. To load the data in the csv files into these tables, use the command: `COPY table-name FROM path/to/file.csv DELIMITER ',' CSV HEADER;`

> Note that you need to load CSV files in some order. It is left as an exercise for you to identify the right order for loading the data.

4 SQL

4.1 Exercise - 1

Run the following queries, submit the output and describe what each query is doing in the report

1. SELECT COUNT(*) FROM Matches JOIN Tournaments ON Matches.tournament_id = Tournaments.tournament_id AND tournament_name = '2014 FIFA World Cup'
2. SELECT COUNT(*) FROM (SELECT DISTINCT Goals.match_id FROM Players JOIN Goals ON Players.player_id = Goals.player_id AND Players.family_name = 'Mbappé' AND given_name = 'Kylian') AS t;
3. SELECT DISTINCT team_name FROM Teams JOIN Matches ON (Teams.team_id = Matches.home_team_id OR Teams.team_id = Matches.away_team_id) AND Matches.stage_name = 'final'
4. SELECT COUNT(*) FROM Teams JOIN (SELECT * FROM Matches JOIN Teams ON ((Teams.team_id = Matches.home_team_id OR Teams.team_id = Matches.away_team_id) AND team_name = 'Germany')) AS t ON ((Teams.team_id = t.home_team_id OR Teams.team_id = t.away_team_id) AND Teams.team_name = 'France' AND t.stage_name != 'group stage')
5. SELECT DISTINCT player_id FROM Goals JOIN (SELECT * FROM Matches JOIN Tournaments ON Matches.tournament_id = Tournaments.tournament_id AND tournament_name = '1930 FIFA World Cup') AS t1 ON Goals.match_id = t1.match_id AND Goals.own_goal = FALSE

4.2 Exercise - 2

Write SQL Queries for the following

1. Tournaments in which the host country is the winner of the tournament
   Output: tournament_id, tournament_name, year, winner
2. Players who played atleast 4 WCs
   Output: player_id, family_name, given_name, count_tournaments
3. Number of draw matches Croatia was a part of
   Output: num_matches
4. Stadium in which the final match of "1990 FIFA World Cup" tournament was held
   Output: stadium_name,city_name,country_name
5. Number of goals scored by Ronaldo (family_name: Ronaldo, given_name: Cristiano) in all WCs (Don't include self goals)
   Output: num_goals
6. Player with highest number of goals in WCs from 2002 - 2018 (both years inclusive, don't include self goals)
   Output: player_id, family_name, given_name, num_goals

7. Team which scored highest number of self goals (lost point to opponent) in last 3 WCs
   Output: team_id, team_name, num_self_goals

Note: In case of a tie for a query, output all the answers.