Introduction to Database Design: Exercise 2

Björn Þór Jónsson

Preparation

Run the sports.sql script against your PostgreSQL server, to create the exercise database. Note that the same database is used in Homework 1.

Note: As mentioned in the first lecture, we advise you to practice running SQL scripts from the command prompt.

The database holds results for track and field athletes in a sports club. For simplicity there are only seven sports; a number of other simplifying assumptions are also made. The database has 254 athletes of both genders and more than 10K results from 200 competitions. The data is purely fictional and randomly generated, except that the places where competitions have been held actually exist (although spelling is "anglicized" to avoid non-ASCII characters). Note that the records represent "national records" so the best result in the database may not match the record.

The Exercise

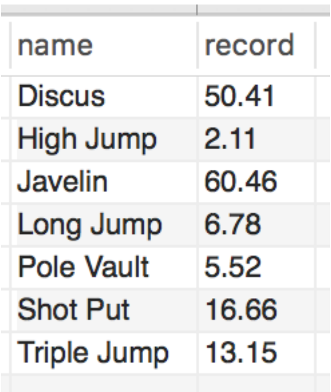
Write SQL commands to retrieve the information requested below from your PostgreSQL database. Each piece of information should be retrieved by a single query. You should make the queries as simple as you possibly can.

The output should, for example, only have the columns and rows asked for, only select from the tables required, ordering should only be used if required, and so on. Generally, the output should contain unique rows; yet, the keyword DISTINCT should not be used unless needed. Note, in particular, that the construct ORDER BY … LIMIT 1 should not be used to find highest or lowest values. Note that when "gender" is requested in the queries, this refers to the description of the gender ("Male" or "Female") from the Gender table (this is to train you in joining tables).

The queries should also be well and consistently formatted, and as readable as possible, as SQL queries are generally part of your code base. We recommend to create a script with all the queries.

Note: We have given hints about the results. In many cases, however, some (incorrect) variants of some queries could return identical results. You should take care to insert additional data to test the queries adequately. Query 2 is a prime example: if there is no sport without a result, then listing all sports will produce the correct output; adding a new sport with no result will test this aspect of the query.

Queries



name	record
Discus	50.41
High Jump	2.11
Javelin	60.46
Long Jump	6.78
Pole Vault	5.52
Shot Put	16.66
Triple Jump	13.15

1. The name and record of all sports, sorted by name. *This query should give the results on the right.*
2. The name of all sports with at least one result.
3. The number of athletes who have competed in at least one sport. *This query should give the result 251.*
4. The ID and name of athletes who have at least twenty results. *This query should return 194 rows.*
5. The **ID,** name and gender description of all athletes that currently hold a record. *This query should return **33** rows.*
6. For each sport, where some athlete holds the record, the name of the sport and the number of athletes that hold a record in that sport; the last column should be named “numathletes”. *The Long Jump has a total of 22 people that have equaled the record, while one sport does not yield a result row. The total number of people that have equaled a record is 38; how can this be larger than the 33 from the previous query?*
7. The ID and name of each athlete that has at least twenty results in the triple jump, their best result, along with the difference between the record and their best result. The second-to-last column should be named “best” while the last column should be named “difference”. The last column should always contain non-negative values and should preferably be formatted to show at least one digit before the decimal point and exactly two digits after the decimal point. *This query should return 7 rows. Why is it that, if you do no formatting, the difference has so many digits?*
8. The ID, name and gender description of all athletes who participated in the competition held in **Hvide Sande** in 2009. *This query should return 84 rows.*
9. The name and gender description of all people with a last name that starts with a “J” and ends with a “sen” (e.g., Jensen, Jansen, Johansen). *This query should return 82 rows.*
10. For each result, the name of the athlete, the name of the sport, and the percentage of the record achieved by the result (a result that is a record should therefore appear as 100; this column should be named “percentage”). Preferably, format the last column to show only whole numbers, as well as the % sign. *This query should return one row for each row in results; if not, then why?*

Deliverables

This exercise is purely for training purposes. No deliverables are needed.