SQL CoP Problem Sheet 1 - Basic Query Statements

George Melrose: gam55@cam.ac.uk (Research Information Office)

Introduction

These problems aim to test your basic SQL knowledge, steadily building up in complexity. The questions and solutions are common ones you will come across when querying different datasets. For the purposes of this series of problem sheets, a database of dummy Marathon results data has been generated. More information on the **Marathon** database is presented below.

The concepts tested in this sheet are covered by the LinkedIn learning course **SQL Server Fundamentals: Master Basic Query Techniques** - (https://www.linkedin.com/learning/sql-server-fundamentals-master-basic-query-techniques) .

Useful Prepatory Resources

In addition to this problem sheet, there are two useful resources you can draw upon to better understand these SQL concepts:

- **Two RMarkdown documents** - one to generate some dummy 'Universities' data (https://github.com/georgemelrose/SQL_Practice/blob/main/0_generating_databasestar_dummy_data.Rmd). This was copied from the excellent SQL learning resource databasestar (https://github.com/bbrumm/databasestar/tree/main/sample_databases/sample_db_university/sqlite).

  The other document is an RMD HTML I generated walking you through basic SQL concepts and how they can be applied to this databasestar dummy data (https://github.com/georgemelrose/SQL_Practice/blob/main/01_Basic_Query_Statements.html).

- **A video presentation** - a recording of a meeting in which I presented the **Basic Query Statements** HTML , explaining key SQL concepts - (https://universityofcambridgecloud.sharepoint.com/sites/AD_Progress/SitePages/Learning-SQL-in-a-New-Format.aspx). This video can be found on the aforementioned page under the **SQL and R** title. *Note that this can only be viewed by SQL" Community of Practice Members, email gam55@cam.ac.uk for access.*

Marathon Database

Firstly, the data to be put into the Marathon database was formulated from the following Python script - (https://github.com/georgemelrose/SQL_Practice/blob/main/Dummy_Marathon_Data/marathon_data_generation.ipynb).

The **marathon data generation** python script generates the following tables:

1. Runners - Randomly generate 1000 runners with names common in their locale/country, together with their birth date and sex.

2. Events - The 6 Major World marathons (Berlin, Boston, Chicago,London,New York City, Tokyo), with an event per year from 2012 to 2023.

3. Results - Gives results for runners in hh:mm:ss format, ensuring there aren't duplicate results for each runner per event. Prevents any results breaking either the male marathon world-record (2:00:35 Eliud Kipchoge 2023) or the female marathon world-record (2:11:53 Brigid Kosgei 2019). Also determines, with a True/False column, if a result is elite by the male standard (below 02:15:00) or the female standard (below 02:30:00).

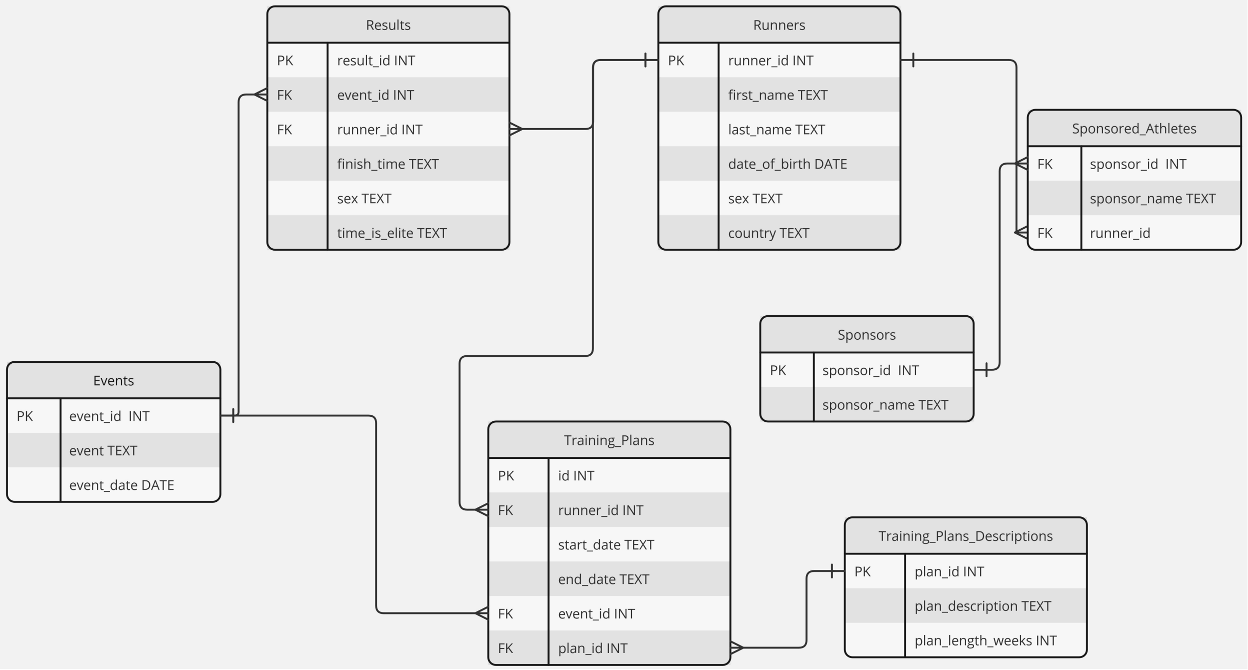
4. Sponsors - Lists the following 10 major companies that typically act as sponsors to runners - "Nike", "Adidas", "Asics", "Saucony", "Hoka","Brooks", "New Balance", "Puma", "Under Armour", "Tracksmith".

5. Sponsored Athletes - A table listing the fraction of the elite athletes that have a sponsor.

6. Training Plans Descriptions - The descriptions of 10 different training plans and their respective lengths in weeks.

7. Training Plans - The training plans of athletes.

Marathon Database Entity Relationship Diagram



Basic Query Statment Problems

Single Table 'SELECT' Statements

Q1. - Get all the tables in the database, inspect their layout, and get the first 5 rows of each table?

Q2.i. - Obtain the full names and countries of the runners?

Q2.ii. - Find the top 10 Marathon results by time, male or female?

Filtering on Single Conditions

Q1 - Find the top 20 male marathon results and the top 20 female marathon results?

Q2.i. - Find all the runners from Lithuania, Latvia, and Estonia?

Q2.ii. - Find all the runners *not* from Poland, Czechia, Slovakia or Hungary?

Q3.i. - Find all the runners whose names begin with ‘J’?

Q3.ii. - Find all the runners whose names don’t begin with ‘J’?

Q3.iii. - Find all the runners whose surnames contain ‘son’?

Q4.i. - Gather all training plans that have a length between 10 and 12 weeks?

Q4.ii. - Gather all training plans that have a length less than 12 weeks?

Q4.iii. - Gather all training plans that have a length more than 12 weeks?

Q4.iv. - Gather all training plans that have a length more than or equal to 12 weeks?

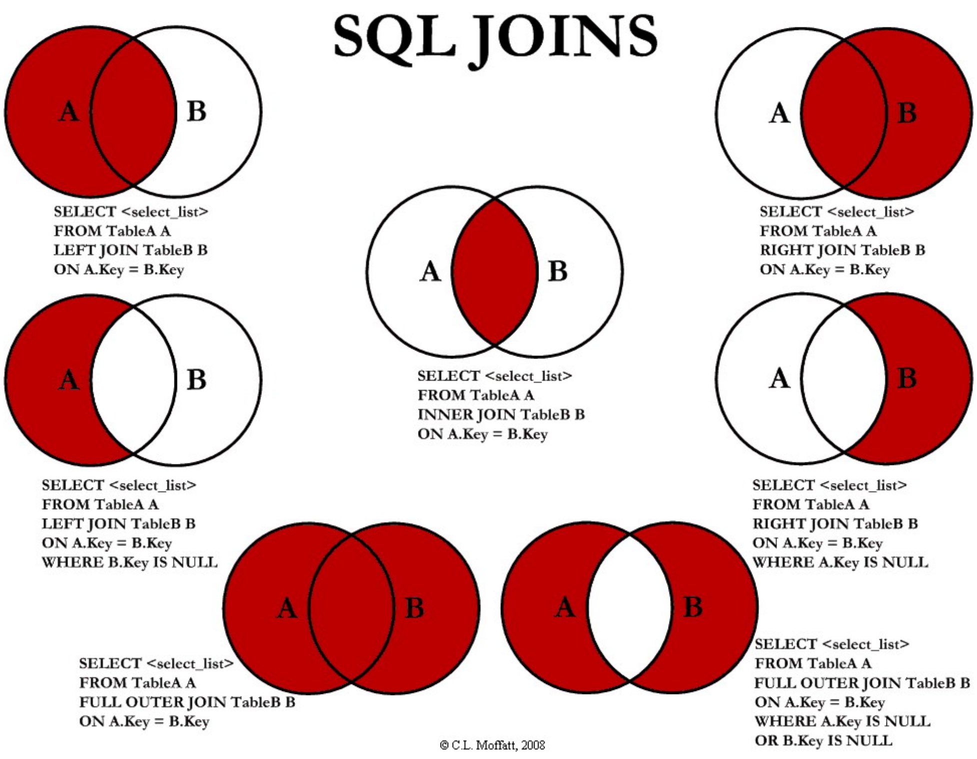
Filtering on Multiple Conditions

Q1.i. - Find all female runners from the United Kingdom born in the 20th century?

Q1.ii. - Find all male runners from Brazil born in 2000?

Q2. - Find the full names of all runners whose names begin with the letters G or J?

Single Inner Joins



```
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
```

```
SELECT <select_list>
FROM TableA A
RIGHT JOIN TableB B
ON A.Key = B.Key
```

```
SELECT <select_list>
FROM TableA A
INNER JOIN TableB B
ON A.Key = B.Key
```

```
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
WHERE B.Key IS NULL
```

```
SELECT <select_list>
FROM TableA A
RIGHT JOIN TableB B
ON A.Key = B.Key
WHERE A.Key IS NULL
```

```
SELECT <select_list>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.Key = B.Key
```

```
SELECT <select_list>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.Key = B.Key
WHERE A.Key IS NULL
OR B.Key IS NULL
```

Q1.i. - Obtain the runner ids of runners with an elite time in the London Marathon(any year)?

Q1.ii. - Obtain the number of runner ids of runners with an elite time in the London Marathon(any year)?.

Q1.iii. - Obtain the number of unique runner ids of runners with an elite time in the London Marathon(any year)?

Q2. - Find the the following information for male runners of any of the Boston Marathon events in the database: runner_id; finish_time; time_is_elite?

Multiple Inner Joins

Q1.i. - Find the full names of runners from the UK who had an elite time in the London Marathon (any year)?

Q1.ii. - Find the number of male and female runners respectively, from the UK, who had an elite time in the London Marathon (any year)?

Q2. - Find the full names, finish time and countries of sponsored runners from the UK who had an elite time in the Berlin Marathon (any year)?

Left Outer Joins

Q1. - Find the full names,sex, and country of runners sponsored by Nike?

Q2.i. - Find dates and times of results that had a finish time under 02:10:00 in Chicago?

Q2.ii. - Count the number of distinct results that had a finish time under 02:10:00 in Chicago?