RMIT Classification: Trusted

RMIT University

SCHOOL OF AISSC

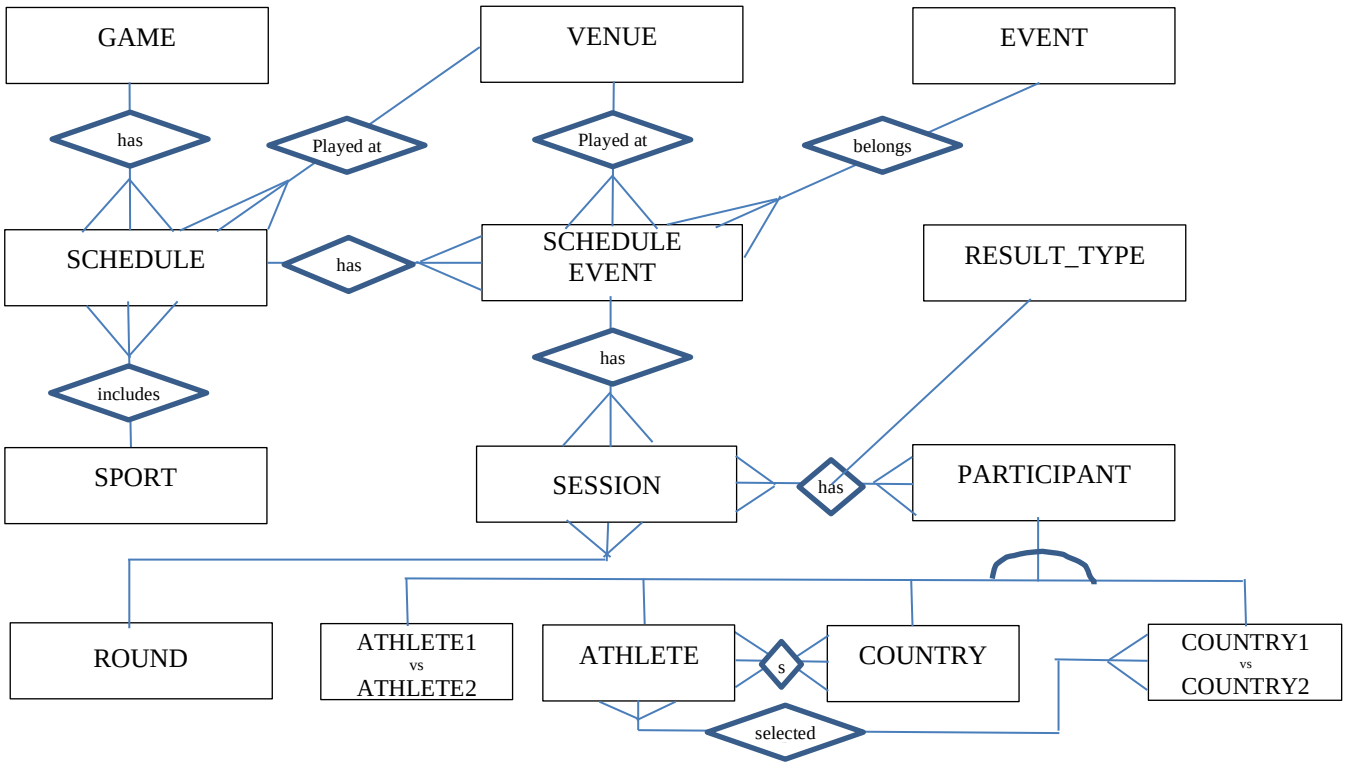
ISYS2421 - Business Data Management and Analytics

Assignment 1 – SQL Queries – Part B

Due Date: End of Week 7

CASE STUDY

ER Diagram



Relational Model

COMPETE	RESULT_TYPE	SCHEDULE	SCHEDULE_EVENT	COUNTRY
compete_id	result_code	schedule_id	schedule_event_id	code
session_id	description	game_id	schedule_id	name
athlete1_id		sport_id	event_id	capital
athlete2_id	SPORT	venue_id	venue_id	continent
team1_code	sport_id		gender	countrycode
team2_code	name			capital_latitude
score1		ATHLETE	GAME	capital_longitude
score2	EVENT	athlete_id	game_id	
place1	event_id	first_name	olympiad	ROUND
place2	name	last_name	year_of_game	round_id
		country	place	name
SELECTED	SESSION	gender	nations_participating	
selected_id	session_id	height	athletes_participating	VENUE
compete_id	round_id	weight	start	venue_id
athelete_id	schedule_event_id		end	name
role			opened_by	
			men	
			women	

Data Modelling Assumption: COUNTRY is the country team playing, unsure whether to call it COUNTRY or TEAM; An ATHLETE belongs to a country (country.code) and when competing the compete.team1_code links to country.code; Athlete's "height" is in centimetres and "weight" is in kilograms. PARTICIPANT (supertype) and ATHLETE & COUNTRY (subtypes) were rolled down with COMPETE (many2many associative table) related to ATHELETE & COUNTRY to cover individual ATHLETE sports (eg. Swimming – athlete1_id) + ATHLETE vs ATHLETE sport (eg. Boxing – athlete1_id + athlete2_id) + COUNTRY team sport (eg. Team relay – team1_code) + COUNTRY team vs COUNTRY team sports (eg. Basketball – team1_code + team2_code). Score1 and score2 can be a rank, goals scored, points, time taken, or one of the RESULT_TYPEs; these fields are stored as CHAR.
Olympic schedule data from: https://www.olympiandatabase.com/; Athlete data from: https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results

SPECIFICATIONS

Read the following questions carefully. You will be asked to specify SQL queries to answer them.

QUESTIONS

You will be working with a set of tables for an Olympic database. You can access these tables by using the Olympic database on the MySQL server (mo.its.rmit.edu.au). You are to prepare 11 SQL query statements and 2 visualisations that will provide answers to the following 12 requests.

(1 marks per question)

1. Create a view in your database (you cannot create view in the Olympics database!) called gameseventschedule; This view will contain a list of all details for sporting events at an Olympic games. You will need to include the following details: game_id, Olympiad, place where Olympics are being held, the name of the sports, the event name for the sport, the venue where the "event" is taking place, schedule_event_id, and the gender competing in that event. Show the gender as information, not M or W. Provide appropriate names for the columns in the view created.

```
SELECT * FROM gameseventschedule WHERE game_id = 1;
```

game_id	olympiad	place	sport	venue	event	gender
1	XXXI	Rio	Athletics	Joao Havelange Olympic Stadium	100 m	female
1	XXXI	Rio	Athletics	Joao Havelange Olympic Stadium	100 m	male
1	XXXI	Rio	Athletics	Joao Havelange Olympic Stadium	200 m	female
1	XXXI	Rio	Athletics	Joao Havelange Olympic Stadium	200 m	male
1	XXXI	Rio	Athletics	Joao Havelange Olympic Stadium	400 m	female
1	XXXI	Rio	Athletics	Joao Havelange Olympic Stadium	400 m	male
1	XXXI	Rio	Athletics	Joao Havelange Olympic Stadium	800 m	female

Note: schedule_event_id not shown in the above, it was cropped for readability!

2. Improve question 4 from assignment 1a, by providing more details. Include the sport name, venue name where the sport is generally being held. Display the data in venue name order.
3. Improve question 6 from assignment 1a, by add the sports name.
4. Show all events scheduled across all Olympic games, displaying the event name and the number of times that event has been scheduled. Show the data from the most scheduled event to the least scheduled event.
5. List all the sporting events where only one gender is participating in the London 30th Olympic games, ignore sporting events where gender is not provided (i.e. not M or W). for example, "Greco-Roman bantamweight" only has a men's event, where "100 m" athletics event has both men and women. Show the place of the games, i.e. London, the name of the sport, and the event together (eg. "Athletics:100 m hurdles") in the one column named "sporting_event".

6. Show a list of all the sports that have not been scheduled for the Rio 31st Olympics games.
7. Provide a list of athletes who have competed at Olympics games 20 or more times (this count should include all sessions of a scheduled event for any sport competed in). Show the athlete's full name, country, and the number of sessions they have competed in an Olympics.
   Note: use athelete1_id only, ignore athelete2_id which is currently null.
8. Show the country that has had the greatest number of athlete's represent them and the country that has had the least amount of athlete's represent them. Display the country code and name along with the count of athletes that have represented them across all Olympics.
   Note 1: there is a relationship between country and athlete, but fieldnames are not the same (which is normally done, but someone did not do it, show ambiguity in data model!), you will have to check the data stored in the fields to discover the appropriate relationship. **Note 2**: create a view first (like shown in lecture) will make this easier!
9. A number of officials have extract the data manually, to check data stored for the Rio Olympics games (game_id = 1) showing the football matches and scores, not all the Football (sport_id 14) scores have been stored, only the finals games…

round	gender	team1	score1	team2	score2
Finals	W	Germany (GER)	2	Canada (CAN)	0
Finals	W	Sweden (SWE)	4	Brazil (BRA)	3

   Can you please generate the above data for the 2012 London and 2008 Beijing Olympics games. The stakeholders want to see the round name, gender, team 1 (country name + country code in brackets), score of team 1, team 2 (country name + country code in brackets), and score of team 2.
10. Your business stakeholders were not happy with the result of question 3. The list provided was missing several sports that are currently under consideration for upcoming Olympic games but have not been scheduled yet at an Olympics. Please improve the report provided in question 3 to include these missing sports.

(2 marks)

11. Choose one questions (from questions 1-10 of this assignment) and create a visualisation, using Excel, Tableau or Orange. Attach the created image ONLY to your submission. The visualisation will be judged on how well it represents the question chosen, how clear the various axis and legends are label, and the appropriateness of the title of the visualisation. Feel free to tweak the output of the query to improve visualisation you want to generate.

(3 marks)

12. Produce a report of your own design and write a query to solve it. Marks will be awarded for report design (i.e., How useful is the report), complexity of the query and originality. Please provide:
    a) Business question
    b) SQL query
    c) Visualisation using Excel, Tableau or Orange (attached image only to submission)

Note: you cannot use the MySQL LIMIT operator because it is NOT standard SQL.

REQUIREMENTS

11 SQL queries that answer the questions asked, based on the data model, and implemented database (on mo.its.rmit.edu.au) provided.
1 visualisation (using Excel, PowerBI, Tableau or Orange) will be created and an image will be submitted

ASSESSMENT

Assessment of the data model will be based on the following areas (by the tutor):

- How well the query answers the questions, in relation to the case study provided.
- Understanding of data structure
- Efficiency and simplicity of resulting query

DEMONSTRATION

Selected students will be required to attend a demonstration session where they will be asked to demonstrate and explain the queries they have written and to write several new queries for the same database. Failure to explain the queries written or the inability to write new queries will result in a FAIL mark being recorded for assignment 1b.

SUBMISSION

- SQL queries (output not required) only.
- Screen dump of Visualisation (can be in a word document or separate image files)
- Assignment will be submitted online using the learning hub (i.e., CANVAS).