VCE APPLIED COMPUTING · UNIT 1 OUTCOME 1 · RELATIONAL DATABASES · LESSON 2 EXTENSION / LESSON 3 RETRIEVAL

Untangle the Olympics

You designed the Ridgeview database from a messy spreadsheet. Now try the same thinking on real data that fights back. Work in pairs; write your own justification.

The data

These 12 rows come from a spreadsheet of Olympic swimming medallists (one row per athlete, per event, per Games). The full file has more than 3,000 rows. Weight has been left out of this excerpt to save space. Real people, published results: this is public sporting history, but treat it with the same care you would want for your own records.

ID	Name	Sex	Age	Height	Team	NOC	Games	Year	City	Event	Medal
36840	Dawn Lorraine Fraser	F	19	172	Australia	AUS	1956 Summer	1956	Melbourne	Swimming Women's 100 metres Freestyle	Gold
36840	Dawn Lorraine Fraser	F	19	172	Australia	AUS	1956 Summer	1956	Melbourne	Swimming Women's 400 metres Freestyle	Silver
36840	Dawn Lorraine Fraser	F	19	172	Australia	AUS	1956 Summer	1956	Melbourne	Swimming Women's 4 x 100 metres Freestyle Relay	Gold
36840	Dawn Lorraine Fraser	F	22	172	Australia	AUS	1960 Summer	1960	Roma	Swimming Women's 100 metres Freestyle	Gold
36840	Dawn Lorraine Fraser	F	27	172	Australia	AUS	1964 Summer	1964	Tokyo	Swimming Women's 100 metres Freestyle	Gold
120087	Ian James Thorpe	M	17	196	Australia	AUS	2000 Summer	2000	Sydney	Swimming Men's 400 metres Freestyle	Gold
120087	Ian James Thorpe	M	17	196	Australia	AUS	2000 Summer	2000	Sydney	Swimming Men's 4 x 100 metres Freestyle Relay	Gold
120087	Ian James Thorpe	M	21	196	Australia	AUS	2004 Summer	2004	Athina	Swimming Men's 400 metres Freestyle	Gold
30777	Sarah Frances "Fanny" Durack (-Gately)	F	22	NA	Australasia	ANZ	1912 Summer	1912	Stockholm	Swimming Women's 100 metres Freestyle	Gold
131941	Wilhelmina "Mina" Wylie	F	20	NA	Australasia	ANZ	1912 Summer	1912	Stockholm	Swimming Women's 100 metres Freestyle	Silver
46968	Cecil Patrick Healy	M	30	NA	Australasia	ANZ	1912 Summer	1912	Stockholm	Swimming Men's 4 x 200 metres Freestyle Relay	Gold
46006	Harold Hampton Hardwick	M	23	NA	Australasia	ANZ	1912 Summer	1912	Stockholm	Swimming Men's 4 x 200 metres Freestyle Relay	Gold

Excerpt from Randi Griffin's public dataset of Olympic results, scraped from sports-reference.com (Griffin, 2018).

Warm-up (2 minutes)

Circle three values that are repeated. For each one, say what it is *about* (an athlete? a Games? an event?).

Core

---

1. This table describes five kinds of thing. Sort every column heading into one of these entities: **Athlete, Team, Games, Event, Result**. A heading can only go in one place.
2. Choose a primary key for each entity. Why is **Name** a poor primary key for Athlete? Look at Fanny Durack's name before you answer.

Entity	Attributes (column headings)	Primary key, and why
Athlete		
Team		
Games		
Event		
Result		

Application

1. Draw your ERD. Label every relationship in words as well as symbols, e.g. *One Games has many Results*.
2. **Where does Age belong?** Use Dawn Fraser's three Games to explain why Age cannot sit in Athlete.
3. **Where does NOC belong?** In 1912 Australian swimmers competed for *Australasia* (ANZ). Should NOC sit in Athlete or in Result? What would go wrong with the other choice?

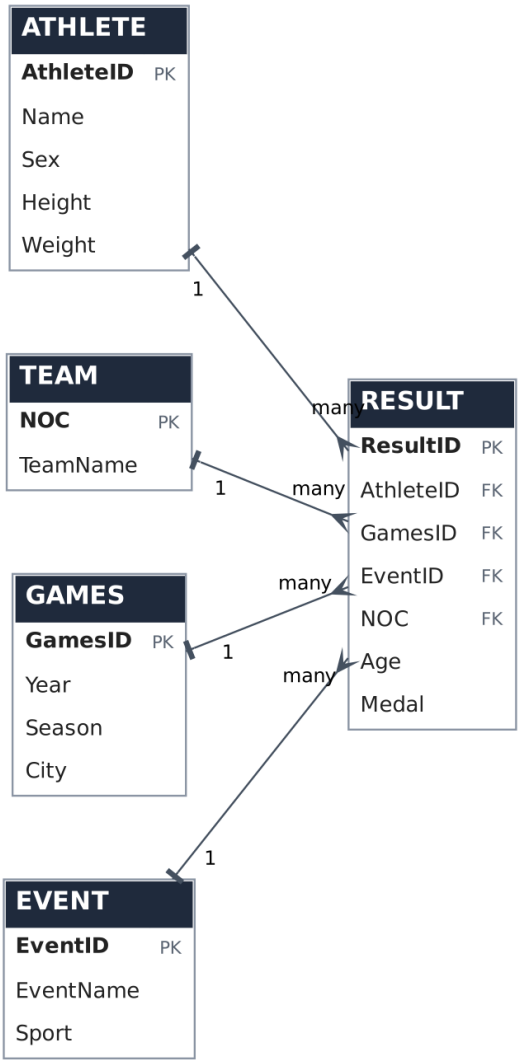
Challenge

1. An athlete can swim many events and an event has many athletes. Explain how your Result entity resolves this many-to-many relationship. Which Ridgeview entity does the same job?
2. Four swimmers share one relay gold. If you count Medal values in Result to make a country's medal tally, what goes wrong? Suggest a fix.
3. Height is 'NA' for every 1912 swimmer. In the full file, some names have lost their accents (e.g. 'Jos Manuel Abascal Gmez'). Which of these problems could a validation rule catch, and which could it never fix?

Sentence starters (same as Lesson 2)

___ should be an entity because…
I selected ___ as the primary key because…
___ needs a foreign key to ___ because…
The relationship between ___ and ___ is one-to-many because…

Teacher notes and answer



One defensible answer. Accept others that are justified.

Decision	What to listen for
Age in Result	Age describes the athlete *at a Games*, so it changes (Fraser: 19, 22, 27). Strong students note it is really derived data: storing a date of birth in Athlete would be better, but this dataset has none.
NOC in Result	Team is chosen per Games, not per person (Australasia 1912; 'Unified Team' 1992 in the full file). Putting NOC in Athlete forces one team for life.

Decision	What to listen for
Year and Games	'1956 Summer' already contains the year, so Year is redundant. A neat example of the redundancy students named in Lesson 1.
Relay tally	Counting Result rows gives four golds for one relay. Fix: count DISTINCT Games + Event + NOC for team events. A good bridge to SQL in Lesson 4.
Validation limits	A rule can reject 'NA' or force a number range; it cannot restore a lost accent or check a value is true. Links to accuracy versus validity (data quality, U1O1).

Where it fits. Extension for pairs who finish the Ridgeview ERD early in Lesson 2, or a 15-minute retrieval task at the start of Lesson 3. The Core rows give entity names, so every student starts from the same scaffold; Challenge questions deepen the same concept rather than adding new content.

Resource by Arvin Hay, part of the Ridgeview database unit. The task applies the ERD process (lists of entities, many-to-many relationships, keys, attribute types) to real data. Griffin, R. (2018). Olympic history: Longitudinal data scraped from www.sports-reference.com [Data set]. figshare. https://figshare.com/articles/dataset/Olympic_history_longitudinal_data_scraped_from_www_sports-reference_com/6121274