EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Database Systems 1
Summerterm 2023
Prof. Torsten Grust, Louisa Lambrecht, Tim Fischer
WSI — Database Systems Research Group

Assignment 10

Hand in this assignment until **Thursday, 13 July 2023, 12:00** at the latest.

Ⓔxam-style Exercises

Exercises marked with Ⓔ are similar in style to those you will find in the exam. You can use these to hone your expectations and gauge your skills.

Running out of ideas?

Are you hitting a roadblock? Are some of the exercises unclear? Do you just need that one hint to get the ball rolling? Refer to the `#forum` channel on our Discord server and check the tag for this assignment —maybe you'll find just the help you need.

Task 1: Cell Phone Tracking

Cell tower networks continuously collect and store cell phone location data. In this task, you will explore similar data. We've provided you with a dataset in the form of a schema file, **`cell_phone_tracking.sql`**, and corresponding data files. The dataset consists of two tables, celltowers[1] and cells.

celltowers

mcc	mnc	lac	cell_id	net_type	lat	lon	…
262	1	25510	27997442	LTE	48.504472	9.037483	…
…	…	…	…	…	…	…	…

cells

mcc	mnc	lac	cell_id	net_type	measured_at
262	1	25510	27997442	LTE	2015-12-22 16:24:50.828
…	…	…	…	…	…

The cells table stores the set of connections that a particular cell phone had with cell towers in the vicinity. These records contain attributes that identify a particular cell tower, and the timestamp indicates when the connection was made. It is possible to roughly track the movements of the phone by combining the timestamp and the position of the cell tower that was in range at the time. Note that a cell phone can only record **one** connection at a time!

Attributes

mcc	The **mobile country code** describes which country a particular cell network belongs to.
mnc	The **mobile network code** identifies a specific cell network operator.
lac	The **location area code** is a rough estimate of the current location of a cell phone in the network of a particular cell phone operator.
cell_id	A unique identifier that represents a cell tower in the network of a particular cell phone operator.
lat lon	The geographic coordinates of a cell tower.

Hint

The celltowers table actually contains more columns than depicted, but they are not strictly necessary to solve this task.

[1]Which we sourced from `http://opencellid.org/`.

Please construct the following queries:

(a) Calculate the number of *cell towers* that are located within a **10000 m** radius from point **(48.521460, 9.053271)** in the center of Tübingen. In order to compute the distance in meters between points p_1 and p_2 on the surface of a sphere, we've provided you with with an implementation of the Haversine formula[2] in form of the **SQL** function

```
haversine(lat_p1, lon_p1, lat_p2, lon_p2)
```

in `haversine.sql`.

count
1788

(b) For each **day** in the table ⊞ cells, calculate which **cell tower** was used the most.

day	cell_id	count
2015-12-22	28016897	39
...	...	...

(c) Write a query to calculate the approximate distance (in meters) traveled by the tracked cell phone per day. The distance can be approximated by summing up the distances between the *cell towers* that were in range in ascending time order.

day	distance
2015-12-22	34037.28...
...	...

Note

- Do **not** use SQL window functions!
- To extract the `date` from a `timestamp` use the built-in function `date(timestamp)`. For example:

  `date('2015-12-22 16:24:50.828')` → `'2015-12-22'`

- Because the queries in this exercise can become quite complex, you may want to consider breaking the problem into into smaller, more manageable chunks via `WITH`.

Task 2: Entity Relationship Model Ⓔ

⚠ Caution

Carefully read the entire task before getting started! There may be some nuance required to get this one right. 😉

Consider the following description of a database that stores information about hospitals:

i A hospital's has a unique name and an address.

ii A ward in a hospital has a ward number (e.g., "A2") as well as a specialty designation (e.g., "delivery ward").

iii Every hospital consists of one or multiple wards. Every ward belongs to exactly one hospital. It is uniquely identified by its ward number together with the name of the hospital.

iv Each hospital has employees which are uniquely identified by their Social Security Number (SSN). For each employee, given name, last name, date of birth and salary need to be stored.

v Each employee works at exactly one hospital.

vi A ward has an arbitrary number of employees assigned, while an employee might be assigned to multiple wards, too.

vii Hospitals also employ doctors which are uniquely identified by the number of their medical license. A doctor may be a specialist of a certain medical field (e.g., "gynecologist") and has an academic title.

viii An employee might be a doctor. Every doctor is an employee.

ix Patients treated in a hospital are identified by their health insurance number. A patient's personal information should be on record: given name and last name, date of birth and address.

x Every patient is treated by at least one doctor who is primarily responsible for him. Additionally (non-responsible) doctors may be consulted which treat the patient as well. A doctor may treat multiple patients at a time.

[2] For an in-depth explanation of the formula have a look at the https://en.wikipedia.org/wiki/Haversine_formula.

Model this scenario in terms of an Entity Relationship diagram. Identify the entity and relationship types as well as their attributes. Underline key attributes and add cardinality constraints in (min, max) notation. Document any additional assumptions you make.

Note
Two entity types may be connected by more than one relationship type.

🔧 Diagramming Tools
While it is certainly possible to draw diagrams using Unicode symbols, some may find this option less than ergonomic. There are a variety of diagramming tools available; the following are some recommendations to get you started:

- yEd live — A web-based WYSIWYG editor from yWorks (in Tübingen 😉).
- diagrams.net — A web-based WYSIWYG editor formerly known as draw.io.
- Graphviz Visual Editor — A web-based graph visualization tool, based on Graphviz.

Because these tools produce image or PDF files, we've relaxed the file type restriction for this task. You can submit any easily accessible and portable file type for. Please note, that this explicitly excludes unportable files, such as `.pptx`, `.odp`, `.pages`, etc.!

Task 3: Translating ER to SQL Ⓔ

Building an ER model from the note of a domain expert is only half of the problem. The other half is actually translating that model into an appropriate SQL schema. This can be especially difficult in situations where the model contains features that are difficult or impossible to express using relational constraints.

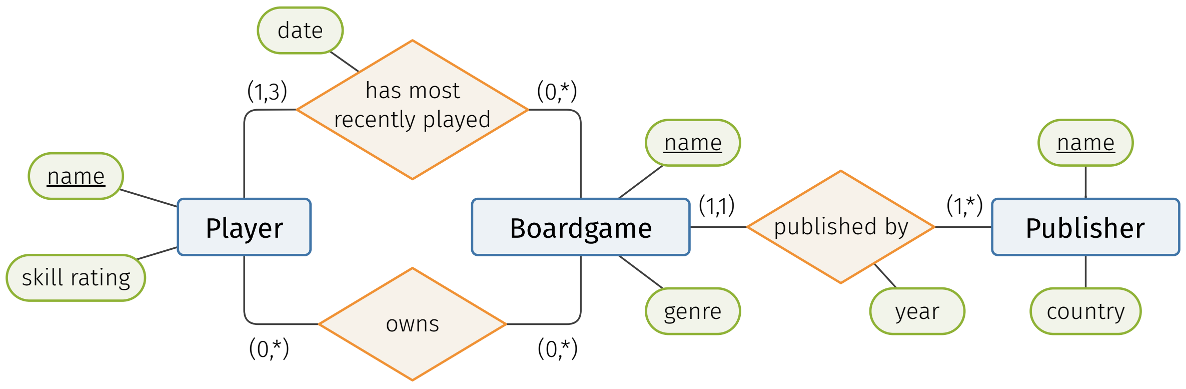

Figure 1: ER model of a board game collection

Consider the ER model in Figure 1, which describes a world of players, their board games, and their publishers.

(a) Which features of the ER model are impossible to express using relational constraints (key, foreign key, `NOT NULL`, `UNIQUE`)?

(b) Translate the ER model into a sequence of SQL DDL statements, excluding the problematic features you identified in subtask (a).