# Introduction to Database Systems
Homework 2

Omar Shahbaz Khan and Eleni Tzirita Zacharatou

Autumn 2024

You should submit the homework as a quiz on LearnIT. In this homework, you will practice ER diagram interpretation, ER diagram design, and SQL queries. Recommended readings are Chapters 3.1-3.3, 6.3-6.4, 7.1-7.4, 9.2, 14.1-14.2, and 14.5 from the PDBM book.

1 ER Diagram Interpretation

Note: This exercise is intentionally designed to be confusing, to prepare you for the real world where you may encounter counter-intuitive or outright nonsensical database designs!
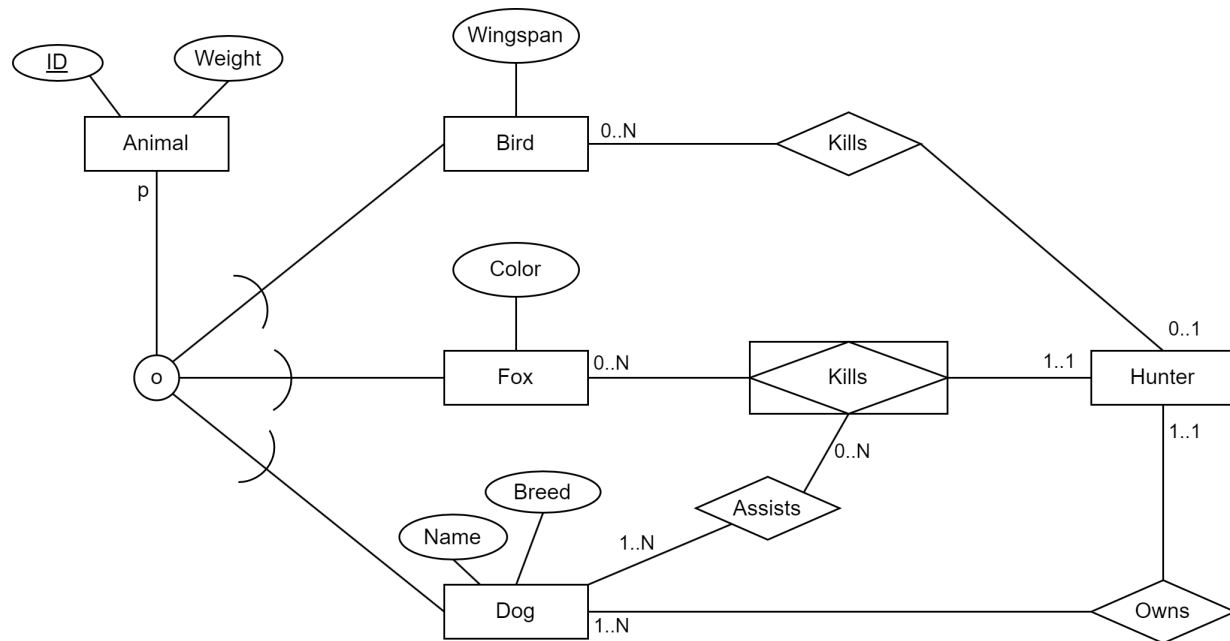
1) In this part, consider ONLY the ER diagram above. Choose the correct statements.

(a) All foxes have a weight.

(b) A hunter can kill a fox without the assistance of a dog.

(c) A dog can kill a bird.

(d) Hunters are animals.

(e) All animals are dead.

(f) A hunter can kill a fox with the assistance of at least one dog.

(g) All foxes are dead.

(h) A bird can be killed by many hunters.

(i) All birds are dead.

(j) Given a hunter, it is possible to find the total weight of the animals he/she has killed.

(k) A bird can be a dog.

2 ER Diagram Design (original by Hildur Davíðsdóttir)

A hiking association has decided to create a database to support its operations. In collaboration with hired consultants, the association has outlined the requirements detailed below. Furthermore, they have chosen you as their database designer. Your first task is to create an ER diagram.

The following requirements were determined through interviews with the hiking association. Your database design should strictly follow the requirements definition. For example, ID columns should only be used where explicitly mentioned. When ID columns are mentioned, you can assume that they are keys.

1. The database must keep track of all people who are to be registered in the system. It should store their ID, name (divided into first and last name), address, SSN, and unique phone number.

2. In this database, each person is a hiker and/or a guide. For guides, their specialization is registered. For hikers, their highest peak is registered (if they have hiked any peaks).

3. Each guide works for at least one company. The start_date and end_date are stored for the hiring. Guides can resign from a company and later be rehired by the same company. The ID and name of companies are stored. A company can be established without having any guides working for it.

4. For guides, a list of their certifications is maintained. Each certification is identified by its name. The database also records the issue date of the certification, the name of the academy that issued it, and the corresponding grade. The grade must be assigned as a letter ranging from A to F.

5. Companies can sell products they create that have an ID and a price. All products are of either of two types: Destinations and Packages.

6. Destinations are organized journeys to a particular destination (example destination: Laugavegur). They have a name and a track name that together form a unique identifier, distance (in km), elevation (in m), and a difficulty rating. Destinations can have other prerequisite destinations, i.e. it is recommended that people have finished certain destinations before pursuing some other destination (e.g., you should have completed Preikestolen and Kjeragbolten before Trolltunga).

7. Packages are package deals that contain multiple destinations (example package: An Icelandic Hiking Adventure, containing Esjan, Vífilsfell and Akrafjall). Destinations do not necessarily have to be part of any package. Packages have a name and a difficulty rating, in addition to their ID and price.

8. Hikers may join hiking clubs. Clubs have an ID, name, and date of establishment. Remember, a club is not a club if it has no members!

9. Both individual hikers and hiking clubs may be representatives of one brand. The database must keep track of these brand representatives and store their ID and the payment they receive. The database must also keep track of brands represented and should store their ID and name.

10. The database should keep track of trips. Trips have an ID and a starting date. Hikers registered in the database have all participated in at least one trip. At least four hikers must participate in each trip, and all trips must have at least one guide. Each trip has one destination, but initially, created destinations have no trips. After a trip, hikers can rate the guidance provided by each guide on the trip. The rating is stored along with a timestamp. Note that the database keeps no track of which hikers buy which products (packages/destinations), it only keeps track of the trip participation.

---

2) Draw an ER diagram reflecting the requirements above, using the ER notation of the PDBM textbook, with amendments made in the lecture. You may not extend the notation to fit the requirements; when the ER diagram cannot implement (part of) a requirement, it should not do so. We strongly recommend using software to create the diagram (for example, draw.io); if it is hand-drawn, please ensure it is sufficiently readable.

3   SQL

In this homework, you will work with a fictional database of countries, cities, and languages. To start working with the database, run the commands in `HW2-DB.sql` found in LearnIT using the PostgreSQL DBMS on your laptop. It is recommended to use `psql` for this purpose.

The database contains a variety of information on countries in the following schema:

```
continents(Continent)
countries(Code, Name, Region, ..., Population, ...)
countries_continents(CountryCode, Continent, Percentage)

cities(ID, Name, CountryCode, District, Population)
empires(CountryCode, Empire)
countries_languages(CountryCode, Language, IsOfficial, Percentage)
```

- `continents` has **7** entries in the database.

- `countries` has **239** entries in the database.

- `countries_continents` has **243** entries in the database.

- `cities` has **4079** entries in the database.

- `empires` has **13** entries in the database.

- `countries_languages` has **986** entries in the database.

Most attributes are self-explanatory. The primary and foreign keys are correctly defined, but you must study the DDL commands to understand the details of these. Some additional notes are in order:

- Some countries are present on more than one continent, and therefore have two entries in `countries_continents`; the `Percentage` attribute refers to the percentage of the population that lives on that continent.

- The table `empires` lists the constituent countries of some (fictional) empires. Countries that are not present in this table are not considered part of any empire.

- The `Percentage` data for languages in `countries_languages` also refers to the percentage of the population that speaks the language. The data is not complete, as the sum of percentages for countries is not 100.0 in all cases; this may be due to rounding errors or due to missing data.

- The data has various other errors, partly by design and partly because it is based on a publicly available dataset that has some errors in it.

5

- In `cities`, the `District` attribute refers to the region of the country where the city is located (rather than a district of the city).

---

3) Answer each of the following questions using a single SQL query on the homework database. Enter the numerical answer of each query into the LearnIT quiz. Finally, upload a script with your queries. Queries should still adhere to the detailed guidelines given in Homework 1.

(a) The empire 'Great Britain' consists of 4 countries. How many countries does the empire 'Iberian' consist of?

(b) There are 4 countries that are present on more than one continent. How many of these countries are partially in Asia?

(c) In the countries of North America that have more than 80 million inhabitants, there are a total of 111,946,176 people who speak Spanish, according to the statistics in the database. How many people who speak Spanish exist in the countries of Europe that have more than 50 million inhabitants?

(d) According to the database, two languages are spoken in all countries of 'Benelux'. How many languages are spoken in all countries of 'Danish Empire'?

Note: This is a *division* query; points will only be awarded if division is attempted.

4 (BSc) Functions & Triggers

4) Choose the correct statements

(a) A row-level AFTER trigger on INSERT can modify the newly inserted row.

(b) In PostgreSQL, a trigger needs to be associated with a function.

(c) A function can only be executed using a SELECT statement.

(d) If a function fails, its changes will be reverted.

5 (MSc) SQL in Python

5) Assume `conn` is a connection object from psycopg. Choose the correct statements

(a) A cursor can only execute one query.

(b) If you perform `conn.execute(`*`query`*`).fetchmany(4)` twice you get the first 8 rows of the *query*.

(c) Prepared Statements can only be used for an SQL query's variables.

(d) If `autocommit=False`, then changes from cursors will only take effect if `conn.commit()` is called.

6 Passing grade

The maximum grade is 70 points. You get 20 points for the ER Diagram Interpretation, 20 points for the ER Diagram Design and Implementation, 4 points for each correct SQL query, 1 point for each correct numerical answer, and 5 points for the BSc or MSc sections. To pass the homework, you need to have *at least 20/70* points. However, the more you do, the more feedback you get and the more you learn.

Homework 2

1. **1. ER Diagram interpretation**

   [MULTI] [20 points] [0.10 penalty] [Multiple]

   1) Choose the correct statements

   (a) All foxes have a weight. (20%)
   (b) A hunter can kill a fox without the assistance of a dog. (-16.66667%)
   (c) A dog can kill a bird. (-16.66667%)
   (d) Hunters are animals. (-16.66667%)
   (e) All animals are dead. (-16.66667%)
   (f) A hunter can kill a fox with the assistance of at least one dog. (20%)
   (g) All foxes are dead. (20%)
   (h) A bird can be killed by many hunters. (-16.66667%)
   (i) All birds are dead. (-16.66667%)
   (j) Given a hunter it is possible to find the total weight of the animals he/she has killed. (20%)
   (k) A bird can be a dog. (20%)

2. **2. ER diagram**

   [ESSAY] [20 points] [0.10 penalty] [monospaced]

   2) Submit a file with an ER diagram.

   Notes for grader:

   - See ER diagram in HW2-ER.pptx. Note that the slide show has an animation, that will show each item in the project description in its own step.

3. **3a) Numerical answer**

   [NUMERICAL] [5 points] [0.10 penalty]

   3a) Run the query and paste the result here (an integer):

   - 3 ✓

4. **3b) Numerical answer**

   [NUMERICAL] [5 points] [0.10 penalty]

   3b) Run the query and paste the result here (an integer):

   - 4 ✓

5. **3c) Numerical answer**

   [NUMERICAL] [5 points] [0.10 penalty]

   3c) Run the query and paste the result here (an integer):

- 236902.8 ✓

6. **3d) Numerical answer**

NUMERICAL | 5 points | 0.10 penalty

3d) Run the query and paste the result here (a floating point):

- 1 ✓

7. **(BSc) Functions & Triggers**

MULTI | 20 points | 0.10 penalty | Multiple

4) Choose the correct statements

(a) A row-level AFTER trigger on INSERT can modify the newly inserted row. (-50%)
(b) In PostgreSQL, a trigger needs to be associated with a function. (50%)
(c) A function can only be executed using a SELECT statement. (-50%)
(d) If a function fails, its changes will be reverted. (50%)

8. **(MSc) SQL in Python**

MULTI | 20 points | 0.10 penalty | Multiple

5) Choose the correct statements

(a) A cursor can only execute one query. (-50%)
(b) If you perform `conn.execute(`*query*`).fetchmany(4)` twice, you get the first 8 rows of the *query*. (-50%)
(c) Prepared Statements can only be used for an SQL query's variables. (50%)
(d) If `autocommit=False`, then changes from cursors will only take effect if `conn.commit()` is called. (50%)

Total of marks: 100