Homework Assignment 1

Maximum earnable: 90 pt.
Due: 11:59PM March 27, 2025

- Read the assignment carefully. *You will need to **write and execute several SQL queries**; and **submit the results of your queries***.
- You are **allowed to re-use any of the queries from the lecture slides** while developing solutions to the problems.
- This is an individual work; Please be clear with HGU CSEE Standard:
  - Submitting assignments or program codes written by others or acquired from the internet without explicit approval of the professor is regarded as cheating.
  - Showing or lending one's own homework to other student is also considered cheating that disturbs fair evaluation and hinders the academic achievement of the other student.
  - It is regarded as cheating if two or more students conduct their homework together and submit it individually when the homework is not a group assignment.
- **Use of ChatGPT or similar AI tools**: Students are prohibited from using ChatGPT or similar AI platforms to directly obtain solutions for this assignment. The intent of the assignment is to exercise your understanding and application of the course material. Leveraging AI tools to bypass this learning process is considered a breach of academic integrity. Any evidence of such behavior will result in penalties.
- Your submission should be in English.
- When finished, submit your work to *LMS*.

Read Chapters 1-2 of Database System Concepts and answer the following questions.

1. (1 pt. per blank) Fill in the blanks.

(a) Underlying the structure of a database is the ( ): a collection of conceptual tools for describing data, data relationships, data semantics, and consistency constraints.

(b) In the relational model, data are represented in the form of ( ). A table has multiple ( ); Each column has a unique ( ). Each ( ) of the table represents one piece of information.

(c) The collection of information stored in the database at a particular moment is called a/an ( ). The overall design of the database is called the ( ).

(d) A database system provides a/an ( ) to specify the database schema and a/an ( ) to express database queries and updates. In practice, these are not two separate languages; instead, they simply form parts of a single language.

(e) A/An ( ) is a statement requesting the retrieval of information.

(f) A/An ( ) ensures that the database remains in a consistent state despite system failures, and that concurrent transaction executions proceed without conflicts.

(g) A/An ( ) manages the allocation of space on disk storage and the data structures used to represent information stored on disk.

(h) A/An ( ) is responsible for fetching data from disk storage into main memory, and deciding what data to cache in main memory.

(i) The ( ) is a special value that signifies the value is unknown or does not exist.

2. (5 pt.) What are the major disadvantages of keeping organizational information in a file-processing system?

3. (5 pt.; Exercise problem 1.7) List four significant differences between a file-processing system and a DBMS.

4. (3 pt. each; Exercise problems 2.6, 2.8, and 2.14) Consider the employee database of Figure 2.17 (page 60). Give an **expression in the relational algebra** to express each of the following queries:
a. Find the name of each employee who lives in city "Miami".
b. Find the name of each employee whose salary is greater than $100,000. *Note that the salary refers to the yearly wage*.
c. Find the name of each employee who lives in "Miami" and whose salary is greater than $100,000.
d. Find the ID and name of each employee who does not work for "BigBank".
e. Find the ID and name of each employee who earns at least as much as every employee in the database.
f. Find the ID, name, and city of residence of each employee who works for "BigBank".
g. Find the ID, name, street address, and city of residence of each employee who works for "BigBank" and earns more than $10,000.
h. Find the ID and name of each employee in this database who lives in the same city as the company for which s/he works.

5. (3 pt. each; Exercise problem 2.18) Write the following **queries in relational algebra**, using the university schema (Figure 2.8, page 46).
a. Find the ID and name of each instructor in the Physics department.
b. Find the ID and name of each instructor in a department located in the building "Watson".
c. Find the ID and name of each student who has taken at least one course in the "Comp. Sci." department.
d. Find the ID and name of each student who has taken at least one course section in the year 2018.
e. Find the ID and name of each student who has not taken any course section in the year 2018.

Warming up with MySQL:
Use the database client installed on your machine to connect to the database that will be introduced in the March 17 lecture. Below uses the "university" database (NOT *university_small*), which shares the same schemas with the database used in the lectures but contains a larger set of data records collected within a different period of time.

6. (5 pt. each) Find the answers to the following questions and provide the SQL queries showing how you find them. All queries should be complete to obtain the listed answers solely by themselves.

a. List all **instructor** names in the *Accounting* department.
Answer:

SQL Query to obtain your answer:

b. How many **students** are in the *Statistics* department?
Answer:

SQL Query to obtain your answer:

c. How many **unique student names** are in the *Astronomy* department?
Answer:

SQL Query to obtain your answer:

d. Find all students who have "**db**" as a substring in their name.
Answer:

SQL Query to obtain your answer:

7. (3 pt.) List the names of all tables that the "**university**" database has.
Answer:

SQL Query to obtain your answer:

8. (4 pt.) Execute and explain the differences among the results of the following queries.
(i) `SELECT * FROM instructor;`
(ii) `SELECT 'Teacher' FROM instructor;`
(iii) `SELECT 'Teacher';`
(iv) `SELECT *, 'Teacher' FROM instructor;`