Homework Assignment 1

Due: 11:59PM March 24, 2023

1. (1 pt. per blank) Fill in the blanks
   a. Underlying the structure of a database is the (**relational data model**): 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 (**tables**). A table has multiple (**columns**); Each column has a unique (**name**). Each (**row**) 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 (**instance**). The overall design of the database is called the (**schema**).
   d. A database system provides a/an (**data-definition language (DDL)**) to specify the database schema and a/an (**data-manipulation language (DML)**) 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 (**query**) is a statement requesting the retrieval of information.
   f. A/An (**transaction manager**) ensures that the database remains in a consistent state despite system failures, and that concurrent transaction executions proceed without conflicts.
   g. A/An (**file manager**) manages the allocation of space on disk storage and the data structures used to represent information stored on disk.
   h. A/An (**buffer manager**) is responsible for fetching data from disk storage into main memory, and deciding what data to cache in main memory.
   i. The (**null value**) is a special value that signifies the value is unknown or does not exist

2. What are the major disadvantages of keeping organizational information in a file-processing system.
   1. Data redundancy and inconsistency
   2. Difficulty in accessing data or slow access time.
   3. Data isolation
   4. Data integrity problems
   5. Atomicity problems
   6. Problem in concurrent access
   7. Unauthorized access (Security problems)

3. List four significant differences between a file-processing system and a DBMS
   1. A file-processing system refers to computer software that organizes and manages files stored on storage media. A DBMS is a piece of software that handles databases, which are collections of related data organized in tables.
   2. A file-processing system allows access to single file at a time. File systems accommodate flat files that have no relation to other files. A DBMS allows access to multiple tables at a time and supports relationships among data using primary keys and foreign keys.
   3. A file-processing system has no mechanism to enforce data integrity, security, concurrency control, or recovery. A DBMS provides various features to ensure data integrity, security, concurrency control, and recovery using constraints, encryption, locking, isolation levels, logging, backup, etc.
   4. A file-processing system has high data dependence as any change in the structure or format of a file requires changing all the programs that access that file. A DBMS has low data dependence as it provides an abstraction layer between the physical structure of data and the logical view of data using schemas.

4. 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) $\Pi_{person\_name}(\sigma_{city = \text{`Miami'}}(employee))$

b) $\Pi_{person\_name}(\sigma_{salary > 100000}(works))$

c) $\Pi_{person\_name}(\sigma_{city = \text{`Miami'}}(employee)) \cap \Pi_{person\_name}(\sigma_{salary > 100000}(works))$

d) $temp = employee \bowtie_{(employee.person\_name = works.person\_name)} (Works)$

$\Pi_{ID, person\_name}(\sigma_{company\_name \neq \text{`BigBank'}}(temp))$

e) $\Pi_{ID, person\_name}(temp) - \Pi_{temp.ID, temp.person\_name}(temp \bowtie_{(temp.salary \leq temp2.salary)} (\rho_{temp2}(temp)))$

f) $R2 = \sigma_{company\_name = \text{`BigBank'}}(temp)$

$\Pi_{ID, person\_name, city}(R2)$

g) $R3 = \sigma_{salary > 10000}(R2)$

$\Pi_{ID, person\_name, street, city}(R3)$

h) $\Pi_{ID, person\_name}(employee \bowtie_{employee.city = company.city} (company))$

5. Write the following queries in relational algebra, using the university schema (Figure 2.8, page 46).

a. $\Pi_{ID, name}(\sigma_{dept\_name = \text{'physics'}}(instructor))$

b. $\Pi_{ID, name}(instructor \bowtie_{instructor.dept\_name = department.dept\_name \wedge department.building = \text{'Watson'}} (department))$

c. $R1 = \Pi_{course\_id, dept\_name}(section \bowtie_{section.course\_id = course.course\_id} (course))$

$R2 = \Pi_{ID, name, course\_id}(student \bowtie_{student.ID = takes.ID} (takes))$

$\Pi_{ID, name}(R1 \bowtie_{R1.course\_id = R2.course\_id \wedge R1.dept\_name = \text{'Comp\_Sci'}} (R2))$

d. $\Pi_{ID, name}(student \bowtie_{student.ID = takes.ID \wedge takes.year = 2018} (takes))$

e. $\Pi_{ID, name}(student \bowtie_{student.ID = takes.ID \wedge takes.year \neq 2018} (takes))$

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.

1. List all instructor names in the Accounting department.
2.
    - Answer:

```
+-------+
|name   |
+-------+
|Lembr  |
|Moreira|
|Hau    |
|Ullman |
+-------+
```

    - SQL Query to obtain your answer:

```
use university;
SELECT ALL name FROM instructor WHERE dept_name='Accounting';
```

3. How many students are in the Statistics department?
    - Answer:

```
+-----+
|count|
+-----+
|85   |
+-----+
```

    - SQL Query to obtain your answer:

```
use university;
SELECT COUNT(*) AS count FROM student WHERE dept_name =
'Statistics';
```

4. How many unique student names are in the Astronomy department?
    - Answer:

```
+-----+
|count|
+-----+
|104  |
+-----+
```

    - SQL Query to obtain your answer:

```
SELECT COUNT(*) AS count
FROM (
        SELECT DISTINCT name FROM student WHERE dept_name =
'Astronomy'
     ) a;
```

5. Find all students who have “db” as a substring in their name.
    - Answer:

```
+-----+--------+----------+--------+
|ID   |name    |dept_name |tot_cred|
+-----+--------+----------+--------+
|2629 |Goldbu  |Languages |4       |
```

```
|67560|Sandberg|Geology   |63     |
|74070|Sandberg|Mech. Eng.|119    |
+-----+--------+----------+--------+
```

- SQL Query to obtain your answer:

```
use university;
SELECT * FROM student WHERE name LIKE '%db%';
```

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

- Answer:

```
+--------------------+
|Tables_in_university|
+--------------------+
|advisor             |
|classroom           |
|course              |
|department          |
|instructor          |
|prereq              |
|section             |
|student             |
|takes               |
|teaches             |
|time_slot           |
+--------------------+
```

- SQL Query to obtain your answer:

```
use university;
show tables;
```

8. (4 pt.) Execute and explain the differences among the results of the following queries.

1. SELECT * FROM instructor;
    a. This query will return all columns of all rows in the 'instructor' table.
2. SELECT 'Teacher' FROM instructor;
    a. This query will return a single column with the value 'Teacher' for each row in the 'instructor' table.
3. SELECT 'Teacher';
    a. This query will return a single column with the value 'Teacher' for a single row.
4. SELECT *, 'Teacher' FROM instructor;
    a. This query will return all columns of all rows in the 'instructor' table, with an additional column containing the value 'Teacher' for each row.