Homework Assignment 1
Due: 11:59PM March 00, 2025

1.

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.

Answer 1.

> a. Data model
> b. Table, column, attribute, tuple
> c. Instance , schema
> d. DDL(data definition language), DML(data manipulation language)
> e. Query
> f. Transaction management
> g. Storage management
> h. Buffer management
> i. Null

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

Answer 2

> 1. Data redundancy and inconsistency; in the file system stored duplicate same data through several files.
> 2. Integrity constraints enforcement issues; file system is hard to modify data at once.
> 3. Concurrency problems; file system possibly to crash when the user modify same data at the same time.
> 4. Security and access control issues; file system hard to access role-based access control

3. List four significant differences between a file-processing system and a DBMS.

> 1. Data independence
> 2. Data redundancy and inconsistency
> 3. Concurrency Control
> 4. Security and access Control

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. 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.

> a. Π person_name (σ city = 'Miami' (employee))
>
> b. Π person_name (σ salary > 100000 (works))
>
> c. Π person_name (σ city = 'Miami' (employee) ⋈ employee.person_name = works.person_name (σ salary > 100000 (works)))
>
> d. Π person_name (employee) - Π person_name (σ company_name = 'BigBank' (works))
>
> e. Π person_name (works) - Π w1.person_name (works w1 ⋈ w2.salary > w1.salary (works w2))
>
> f. Π person_name, city (σ company_name = 'BigBank' (works) ⋈ works.person_name = employee.person_name (employee))
>
> g. Π person_name, street, city (σ company_name = 'BigBank' ∧ salary > 10000 (works) ⋈ works.person_name = employee.person_name (employee))
>
> h. Π person_name (employee ⋈ employee.person_name = works.person_name (works) ⋈ works.company_name = company.company_name (company) ⋈ employee.city = company.city)

5. 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.

Answer 5
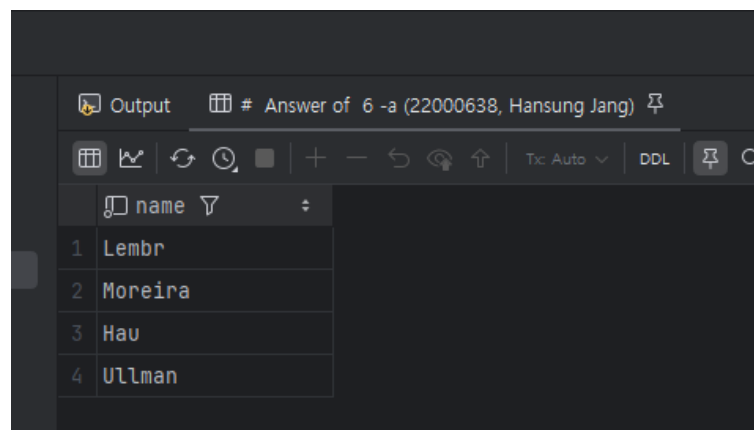
> a. Π ID, name (σdept_name = 'Physics'(instructor))
> b. Π ID, name (σ building = 'Watson' (department) ⋈ department.dept_name = instructor.dept_name
>   (instructor))
> c. Π ID, name (σ dept_name = 'Comp. Sci.' (course) ⋈ course.course_id = takes.course_id (takes) ⋈
>   takes.ID = student.ID (student))
> d. Π ID, name (σ year = 2018 (section) ⋈ section.course_id = takes.course_id ∧ section.sec_id =
>   takes.sec_id ∧ section.semester = takes.semester (takes) ⋈ takes.ID = student.ID (student))
> e. Π ID, name (student) - Π ID, name (Π ID (σ year = 2018 (section) ⋈ section.course_id = takes.course_id
>   ∧ section.sec_id = takes.sec_id ∧ section.semester = takes.semester (takes) ⋈ takes.ID = student.ID
>   (student)))

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:
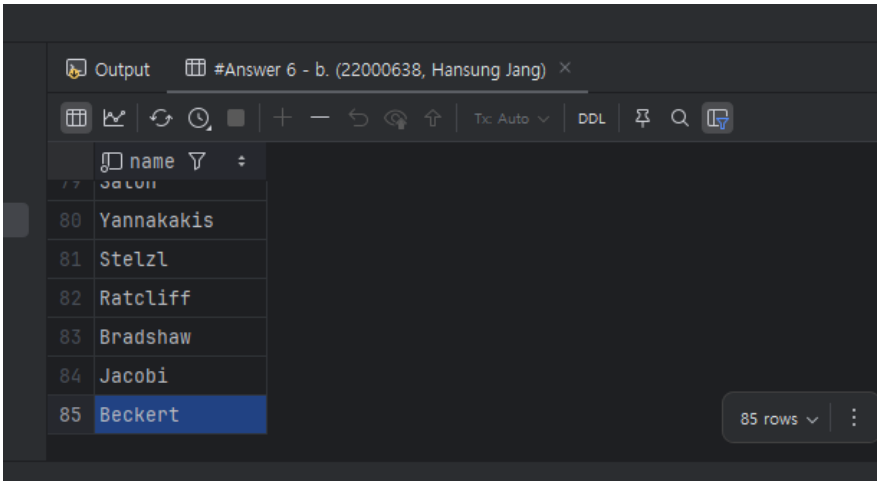
[Query Result]	[SQL query answer]
	SELECT ALL name from instructor Where dept_name = 'Accounting';

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

<mark>Answer</mark>: **85 students**
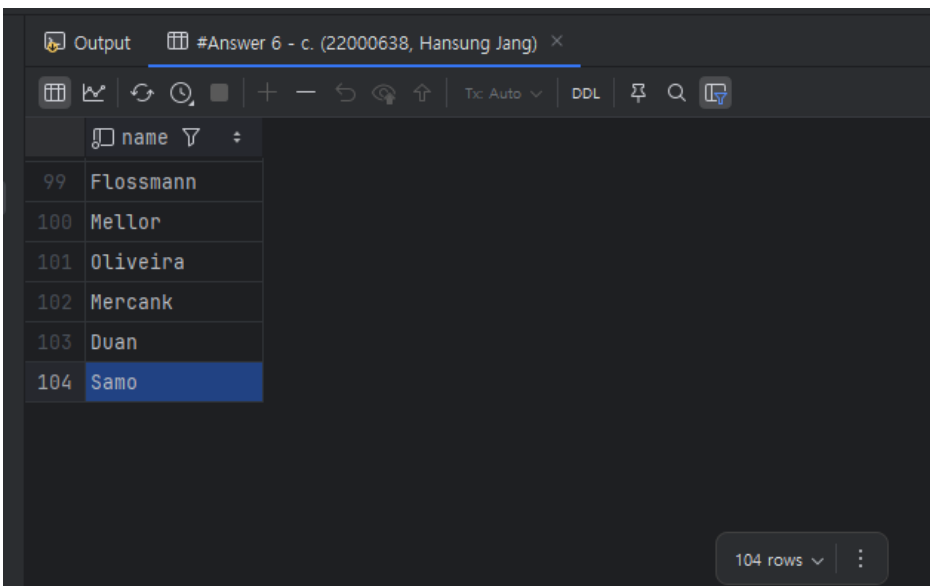
[Query Result]	[SQL query answer]
	SELECT ALL name from student Where dept_name = 'Statistics';

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

<mark>Answer</mark>: **104 names**
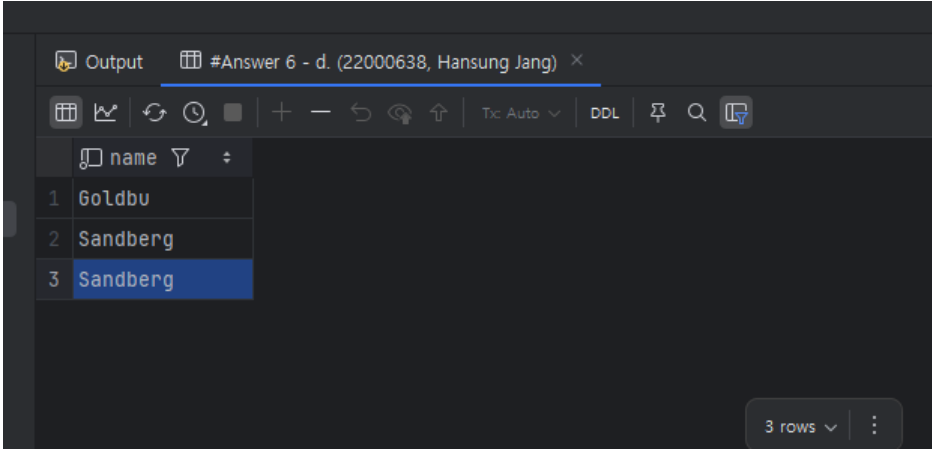
[Query Result]	[SQL query answer]
	SELECT **Distinct** name from student Where dept_name = 'Astronomy';
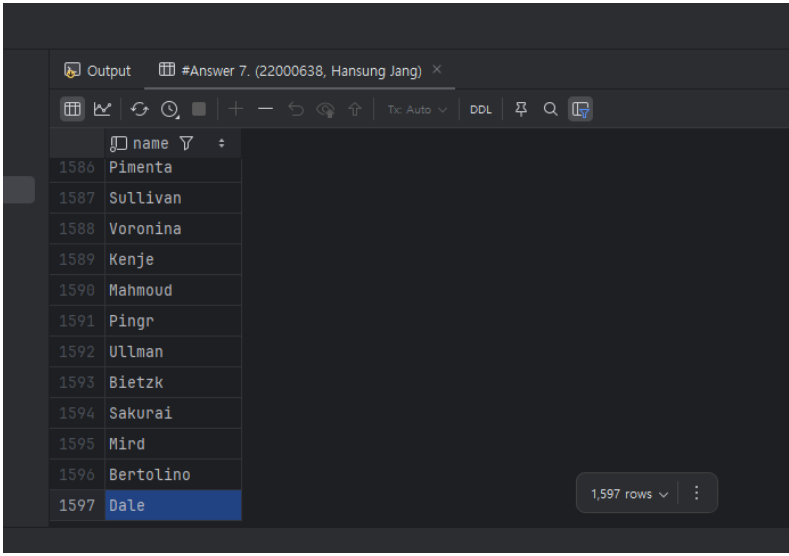
d.   Find all students who have "**db**" as a substring in their name.

<mark>Answer</mark>: **3 students**

[Query Result]	[SQL query answer]
	SELECT ALL name from student where name **Like** '%db%';

7. (3 pt.) List the names of all tables that the "**university**" database has.
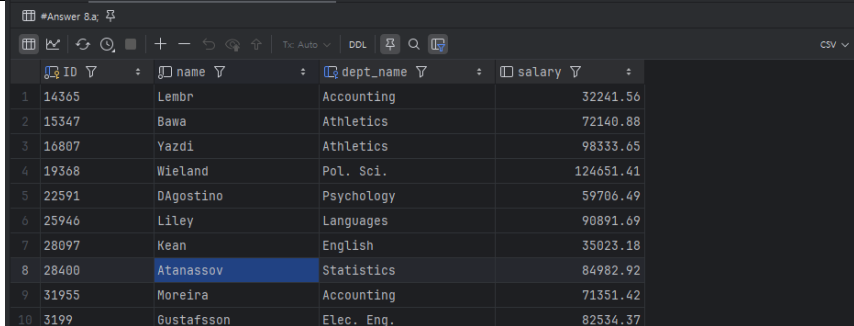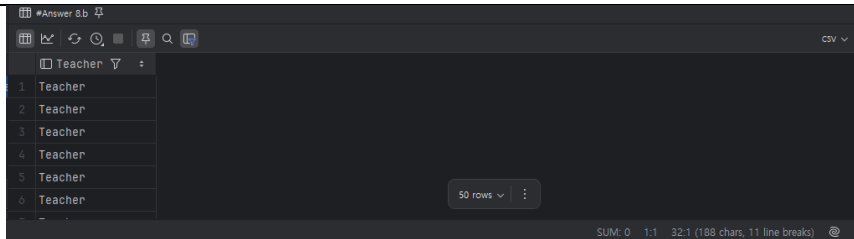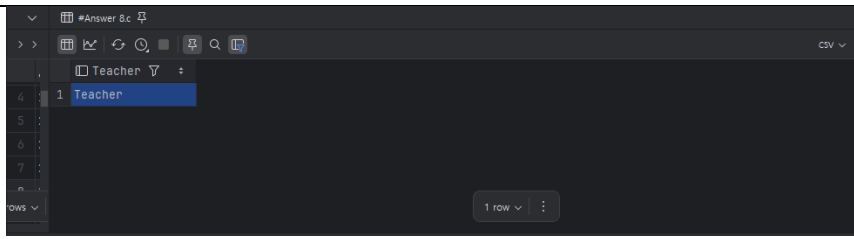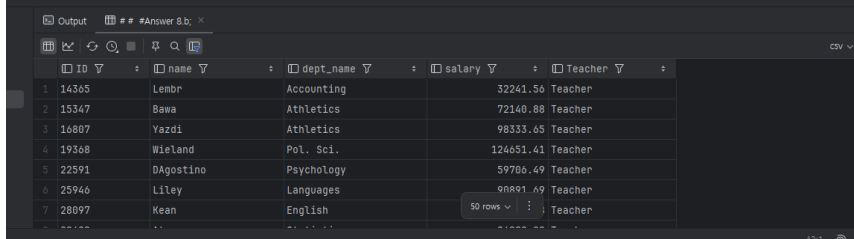
<mark>Answer</mark>: **1,597 names**

[Query Result]	[SQL query answer]
	(SELECT ALL name from student) **UNION** (SELECT ALL name from instructor); ;

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;

Answer:

Query Result	Explain
	(8.a.) is **a query that uses the *** **operator to display all tuples** **from the instructor table.**
	(8.b.) is **a query that uses a literal** **operator with FROM** **instructor, which outputs the** **text 'Teacher' once for each** **tuple in the instructor table.**
	(8.c.) is **similar to (8.b), this query** **uses a literal operator, but** **without a FROM clause, so it** **only outputs 'Teacher' once.**
	(8.d.) is **a query that uses both the** **selection operator * and a** **literal operator to output all** **tuples from the instructor** **table along with an additional** **column containing the text** **'Teacher'.**