

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.

Answer 3

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.

Answer 4

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

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

- Find the ID and name of each instructor in the Physics department.
- Find the ID and name of each instructor in a department located in the building "Watson".
- Find the ID and name of each student who has taken at least one course in the "Comp. Sci." department.
- Find the ID and name of each student who has taken at least one course section in the year 2018.
- Find the ID and name of each student who has not taken any course section in the year 2018.

Answer 5

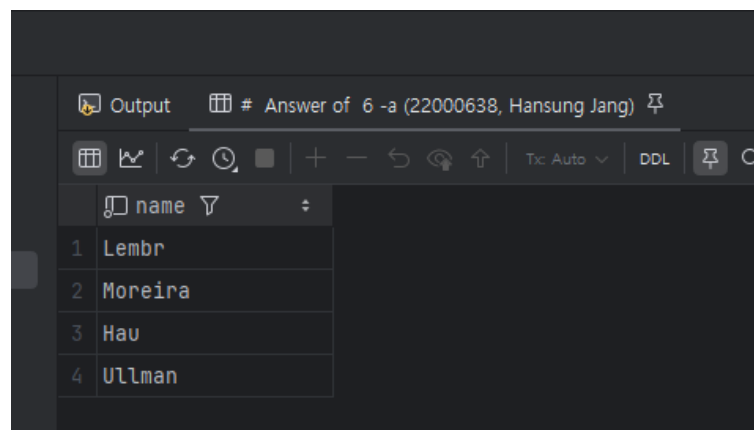
- π ID, name (σ dept_name = 'Physics'(instructor))
- π ID, name (σ building = 'Watson' (department) \bowtie department.dept_name = instructor.dept_name (instructor))
- π ID, name (σ dept_name = 'Comp. Sci.' (course) \bowtie course.course_id = takes.course_id (takes) \bowtie takes.ID = student.ID (student))
- π ID, name (σ year = 2018 (section) \bowtie section.course_id = takes.course_id \wedge section.sec_id = takes.sec_id \wedge section.semester = takes.semester (takes) \bowtie takes.ID = student.ID (student))
- π ID, name (student) - π ID, name (π ID (σ year = 2018 (section) \bowtie section.course_id = takes.course_id \wedge section.sec_id = takes.sec_id \wedge section.semester = takes.semester (takes) \bowtie 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.

- List all **instructor** names in the *Accounting* department.

Answer:

[Query Result]



The screenshot shows a database interface with a query result table. The table has a single column labeled 'name' and four rows of data. The names listed are Lembr, Moreira, Hau, and Ullman. The interface includes a toolbar with various icons for editing and viewing the query.

	name
1	Lembr
2	Moreira
3	Hau
4	Ullman

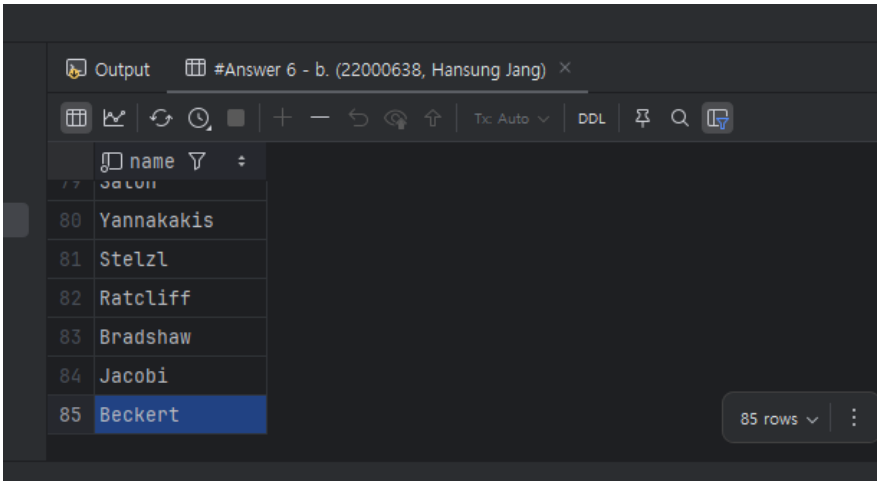
[SQL query answer]

```
SELECT ALL name
from instructor
Where dept_name = 'Accounting';
```

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

Answer: 85 students

[Query Result]



The screenshot shows a database query result window titled "#Answer 6 - b. (22000638, Hansung Jang)". The window displays a table with a single column named "name". The table contains 85 rows of student names. The first row is partially visible as "Salvi", and the last row is "Beckert". The window also shows a toolbar with various icons and a status bar indicating "85 rows".

name
Salvi
Yannakakis
Stelzl
Ratcliff
Bradshaw
Jacobi
Beckert

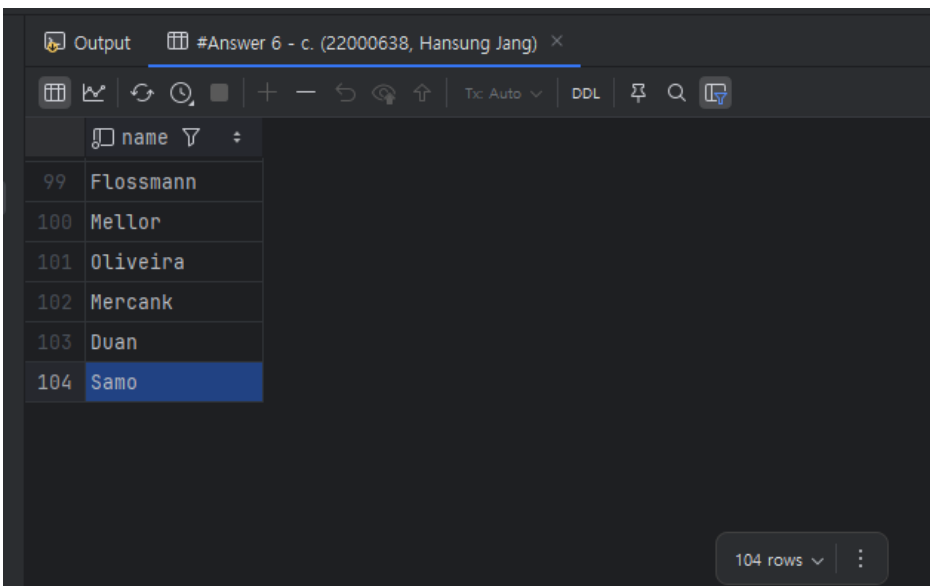
[SQL query answer]

```
SELECT ALL name  
from student  
Where dept_name = 'Statistics';
```

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

Answer: 104 names

[Query Result]



The screenshot shows a database query result window titled "#Answer 6 - c. (22000638, Hansung Jang)". The window displays a table with a single column named "name". The table contains 104 rows of unique student names. The first row is "Flossmann", and the last row is "Samo". The window also shows a toolbar with various icons and a status bar indicating "104 rows".

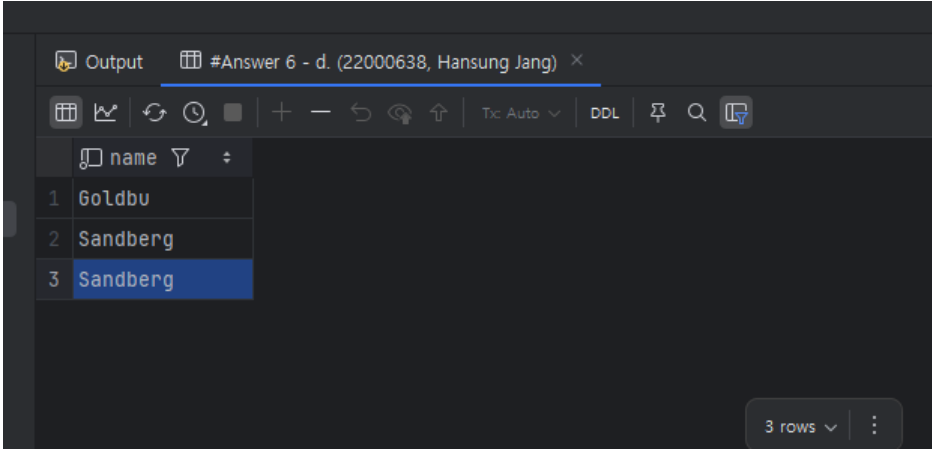
name
Flossmann
Mellor
Oliveira
Mercank
Duan
Samo

[SQL query answer]

```
SELECT Distinct name  
from student  
Where dept_name =  
'Astronomy';
```

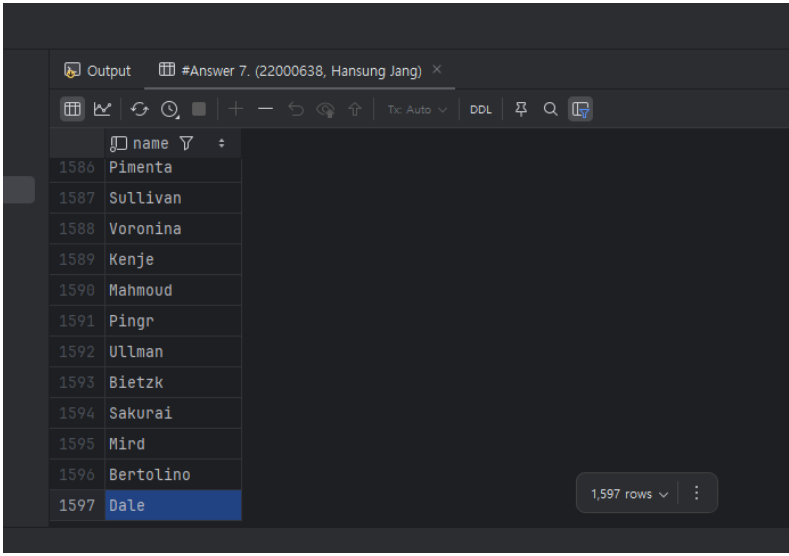
- d. Find all students who have **"db"** as a substring in their name.

Answer: 3 students

[Query Result]	[SQL query answer]
 <p>The screenshot shows a database query result window titled '#Answer 6 - d. (22000638, Hansung Jang)'. The window displays a table with one column named 'name'. There are three rows of data: 'Goldbu', 'Sandberg', and 'Sandberg'. The third row is highlighted. At the bottom right, it indicates '3 rows'.</p>	<pre>SELECT ALL name from student where name Like '%db%';</pre>

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

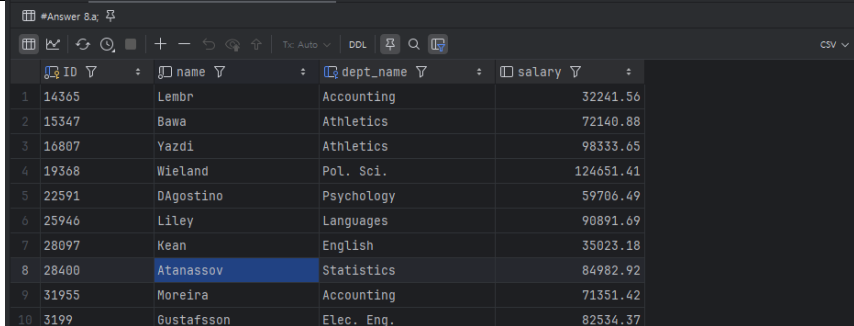
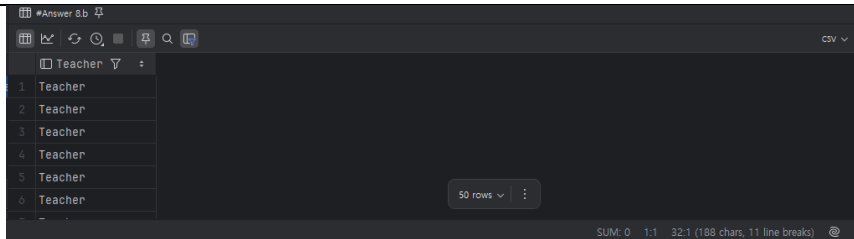
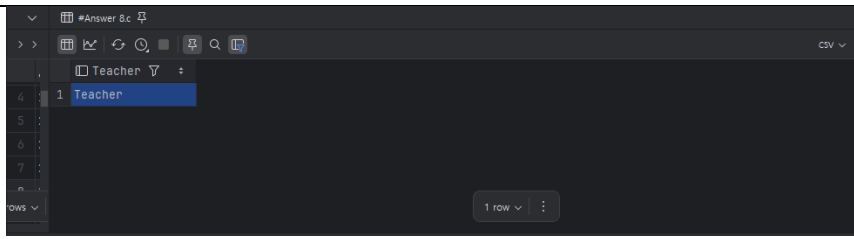
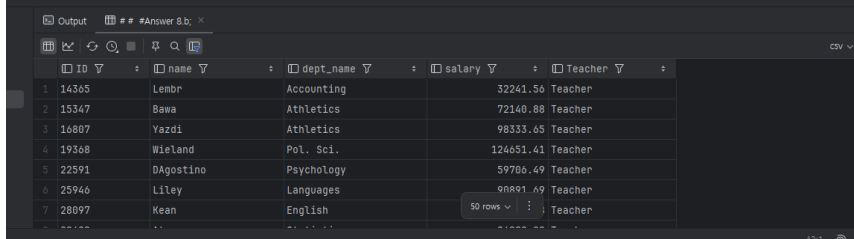
Answer: 1,597 names

[Query Result]	[SQL query answer]
 <p>The screenshot shows a database query result window titled '#Answer 7. (22000638, Hansung Jang)'. The window displays a table with one column named 'name'. The first few rows are: 1586 Pimenta, 1587 Sullivan, 1588 Voronina, 1589 Kenje, 1590 Mahmoud, 1591 Pingr, 1592 Ullman, 1593 Bietzk, 1594 Sakurai, 1595 Mird, 1596 Bertolino, and 1597 Dale. The last row is highlighted. At the bottom right, it indicates '1,597 rows'.</p>	<pre>(SELECT ALL name from student) UNION (SELECT ALL name from instructor); ;</pre>

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																																																							
 <table><thead><tr><th>ID</th><th>name</th><th>dept_name</th><th>salary</th></tr></thead><tbody><tr><td>14365</td><td>Lembr</td><td>Accounting</td><td>32241.56</td></tr><tr><td>15347</td><td>Bawa</td><td>Athletics</td><td>72140.88</td></tr><tr><td>16807</td><td>Yazdi</td><td>Athletics</td><td>98333.65</td></tr><tr><td>19368</td><td>Wieland</td><td>Pol. Sci.</td><td>124651.41</td></tr><tr><td>22591</td><td>D'Agostino</td><td>Psychology</td><td>59786.49</td></tr><tr><td>25946</td><td>Liley</td><td>Languages</td><td>90891.69</td></tr><tr><td>28097</td><td>Kean</td><td>English</td><td>35023.18</td></tr><tr><td>28400</td><td>Atanassov</td><td>Statistics</td><td>84982.92</td></tr><tr><td>31955</td><td>Moreira</td><td>Accounting</td><td>71351.42</td></tr><tr><td>3199</td><td>Gustafsson</td><td>Elec. Eng.</td><td>82534.37</td></tr></tbody></table>	ID	name	dept_name	salary	14365	Lembr	Accounting	32241.56	15347	Bawa	Athletics	72140.88	16807	Yazdi	Athletics	98333.65	19368	Wieland	Pol. Sci.	124651.41	22591	D'Agostino	Psychology	59786.49	25946	Liley	Languages	90891.69	28097	Kean	English	35023.18	28400	Atanassov	Statistics	84982.92	31955	Moreira	Accounting	71351.42	3199	Gustafsson	Elec. Eng.	82534.37	(8.a.) is a query that uses the * operator to display all tuples from the instructor table.											
ID	name	dept_name	salary																																																					
14365	Lembr	Accounting	32241.56																																																					
15347	Bawa	Athletics	72140.88																																																					
16807	Yazdi	Athletics	98333.65																																																					
19368	Wieland	Pol. Sci.	124651.41																																																					
22591	D'Agostino	Psychology	59786.49																																																					
25946	Liley	Languages	90891.69																																																					
28097	Kean	English	35023.18																																																					
28400	Atanassov	Statistics	84982.92																																																					
31955	Moreira	Accounting	71351.42																																																					
3199	Gustafsson	Elec. Eng.	82534.37																																																					
 <table><thead><tr><th>Teacher</th></tr></thead><tbody><tr><td>Teacher</td></tr><tr><td>Teacher</td></tr><tr><td>Teacher</td></tr><tr><td>Teacher</td></tr><tr><td>Teacher</td></tr><tr><td>Teacher</td></tr></tbody></table>	Teacher	Teacher	Teacher	Teacher	Teacher	Teacher	Teacher	(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.																																																
Teacher																																																								
Teacher																																																								
Teacher																																																								
Teacher																																																								
Teacher																																																								
Teacher																																																								
Teacher																																																								
 <table><thead><tr><th>Teacher</th></tr></thead><tbody><tr><td>Teacher</td></tr></tbody></table>	Teacher	Teacher	(8.c.) is similar to (8.b), this query uses a literal operator, but without a FROM clause, so it only outputs 'Teacher' once.																																																					
Teacher																																																								
Teacher																																																								
 <table><thead><tr><th>ID</th><th>name</th><th>dept_name</th><th>salary</th><th>Teacher</th></tr></thead><tbody><tr><td>14365</td><td>Lembr</td><td>Accounting</td><td>32241.56</td><td>Teacher</td></tr><tr><td>15347</td><td>Bawa</td><td>Athletics</td><td>72140.88</td><td>Teacher</td></tr><tr><td>16807</td><td>Yazdi</td><td>Athletics</td><td>98333.65</td><td>Teacher</td></tr><tr><td>19368</td><td>Wieland</td><td>Pol. Sci.</td><td>124651.41</td><td>Teacher</td></tr><tr><td>22591</td><td>D'Agostino</td><td>Psychology</td><td>59786.49</td><td>Teacher</td></tr><tr><td>25946</td><td>Liley</td><td>Languages</td><td>90891.69</td><td>Teacher</td></tr><tr><td>28097</td><td>Kean</td><td>English</td><td>35023.18</td><td>Teacher</td></tr><tr><td>28400</td><td>Atanassov</td><td>Statistics</td><td>84982.92</td><td>Teacher</td></tr><tr><td>31955</td><td>Moreira</td><td>Accounting</td><td>71351.42</td><td>Teacher</td></tr><tr><td>3199</td><td>Gustafsson</td><td>Elec. Eng.</td><td>82534.37</td><td>Teacher</td></tr></tbody></table>	ID	name	dept_name	salary	Teacher	14365	Lembr	Accounting	32241.56	Teacher	15347	Bawa	Athletics	72140.88	Teacher	16807	Yazdi	Athletics	98333.65	Teacher	19368	Wieland	Pol. Sci.	124651.41	Teacher	22591	D'Agostino	Psychology	59786.49	Teacher	25946	Liley	Languages	90891.69	Teacher	28097	Kean	English	35023.18	Teacher	28400	Atanassov	Statistics	84982.92	Teacher	31955	Moreira	Accounting	71351.42	Teacher	3199	Gustafsson	Elec. Eng.	82534.37	Teacher	(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'.
ID	name	dept_name	salary	Teacher																																																				
14365	Lembr	Accounting	32241.56	Teacher																																																				
15347	Bawa	Athletics	72140.88	Teacher																																																				
16807	Yazdi	Athletics	98333.65	Teacher																																																				
19368	Wieland	Pol. Sci.	124651.41	Teacher																																																				
22591	D'Agostino	Psychology	59786.49	Teacher																																																				
25946	Liley	Languages	90891.69	Teacher																																																				
28097	Kean	English	35023.18	Teacher																																																				
28400	Atanassov	Statistics	84982.92	Teacher																																																				
31955	Moreira	Accounting	71351.42	Teacher																																																				
3199	Gustafsson	Elec. Eng.	82534.37	Teacher																																																				