CSCB20 UTSC Worksheet 1-Winter 2025

1. List four applications you have used that most likely employed a database system to store persistent data.

Solution:

- Banking: For account information, transfer of funds, banking transactions.
- Universities: For student information, online assignment submissions, course registrations, and grades.
- Airlines: For reservation of tickets and schedule information.
- Online news sites: For updating news and maintaining archives.
- Online-trade: For product data, availability and pricing information, order-tracking facilities, and generating recommendation lists.

2. List at least three significant differences between a file-processing system and a DBMS.

Solution:

Some main differences between a database management system and a file-processing system are:

- Both systems contain a collection of data and a set of programs which access the data. A database management system coordinates both the physical and the logical access to the data, whereas a file-processing system coordinates only the physical access.
- A database management system reduces the amount of data duplication by ensuring that a physical piece of data is available to all programs authorized to have access to it, whereas data written by one program in a file-processing system may not be readable by another program.
- A database management system is designed to allow flexible access to data (i.e., queries), whereas a file-processing system is designed to allow predetermined access to data (i.e., compiled programs).
- A database management system is designed to coordinate multiple users accessing the same data at the same time. A file-processing system is usually designed to allow one or more programs to access different data files at the same time. In a file-processing system, a file can be accessed by two programs concurrently only if both programs have read-only access to the file.

3. Describe the differences in meaning between the terms *relation* and *relation schema.* Explain using an example.

Solution:

A relation schema is a type definition, and a relation is an instance of that schema. For example, *student* (*ss#, name*) is a relation schema and

123-456-222	John
234-567-999	Mary

is a relation based on that schema.

4. In the figure below, are instances of two relations that might constitute part of a banking database. Indicate the following:

acctNo	*type*	*balance*
12345	savings	12000
23456	checking	1000
34567	savings	25

The relation Accounts

firstName	*lastName*	*idNo*	*account*
Robbie	Banks	901-222	12345
Lena	Hand	805-333	12345
Lena	Hand	805-333	23456

The relation Customers

(a) The attributes of each relation

(b) The tuples of each relation

(c) The components of one tuple from each relation

(d) The relation schema for each relation

(e) The database schema

(f) A suitable domain for each attribute

(g) Another equivalent way to present each relation.

5. Consider the following bank database. Assume that branch names and customer names uniquely identify branches and customers, but loans and accounts can be associated with more than one customer.

branch(*branch_name*, *branch_city*, *assets*)
customer (*ID*, *customer_name*, *customer_street*, *customer_city*)
loan (*loan_number*, *branch_name*, *amount*)
borrower (*ID*, *loan_number*)
account (*account_number*, *branch_name*, *balance*)
depositor (*ID*, *account_number*)

(a) What are the appropriate primary keys?

(b) Given your choice of primary keys, identify appropriate foreign keys.

Solution:

(a) The primary keys of the various schemas are underlined. We allow customers to have more than one account, and more than one loan.

Branch (<u>*branch_name*</u>, *branch_city*, *assets*)

customer (<u>*ID*</u>, *customer_name*, *customer_street*, *customer_city*)

loan (<u>*loan_number*</u>, *branch_name*, *amount*)

borrower (<u>*ID*</u>, <u>*loan_number*</u>)

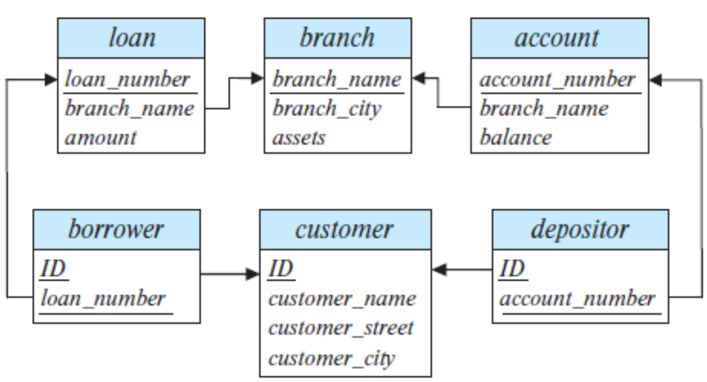
account (<u>*account_number*</u>, *branch_name*, *balance*)

depositor (<u>*ID*</u>, <u>*account_number*</u>)

(b) The foreign keys are as follows:

i. For *loan*: *branch_name* referencing *branch*.
ii. For *borrower*: Attribute *ID* referencing *customer* and *loan_number* referencing *loan*
iii. For *account*: *branch_name* referencing *branch*.
iv. For *depositor*: Attribute *ID* referencing customer and *account_number* referencing *account*

6. Construct a schema diagram for the bank database from Q.5.



7. Consider the bank database. Give an expression in the relational algebra for each of the following queries:

branch(*branch_name*, *branch_city, assets*)
customer (*ID*, *customer_name*, *customer_street, customer_city*)
loan (*loan_number*, *branch_name, amount*)
borrower (*ID*, *loan_number*)
account (*account_number*, *branch_name, balance*)
depositor (*ID*, *account_number*)

(a) Find each loan number with a loan amount greater than $10000.
(b) Find the ID of each depositor who has an account with a balance greater than $6000.
(c) Find the ID of each depositor who has an account with a balance greater than $6000 at the "Uptown" branch.

Solution:

(a) $\prod_{loan\_number} (\sigma_{amount > 10000}(loan))$
(b) $\prod_{ID} (\sigma_{\text{balance} > 6000}$
$(depositor \bowtie_{depositor.account\_number=account.account\_number} account))$
(c) $\prod_{ID} (\sigma_{\text{balance} > 6000 \wedge \text{branch\_name}=\text{"Uptown"}}$
$(depositor \bowtie_{depositor.account\_number=account.account\_number} account))$

8. Consider the employee database. Give an expression in the relational algebra to express each of the following queries:

employee (ID, person_name, street, city)
works (ID, person_name, company_name, salary)
company (company_name, city)

(a) Find the ID and name of each employee who works for "BigBank".
(b) Find the ID, name, and city of residence of each employee who works for "BigBank".
(c) Find the ID, name, street address, and city of residence of each employee who works for "BigBank" and earns more than $10000.
(d) Find the ID and name of each employee in this database who lives in the same city as the company for which she or he works.

Solution:

(a) $\prod_{ID,\ person\_name} (\sigma_{company\_name = \text{"BigBank"}} (works))$
(b) $\prod_{ID,\ person\_name,\ city} (employee \bowtie_{employee.id=works.id} (\sigma_{company\_name = \text{"BigBank"}} (works))$
(c) $\prod_{ID,\ person\_name,\ street,\ city} (\sigma_{(company\_name = \text{"BigBank"} \wedge salary > 10000)} (Works \bowtie_{employee.id=works.id} employee))$
(d) $\prod_{ID,\ person\_name}$
$(\sigma_{employee.city=company.city}(employee \bowtie_{employee.ID=works.ID} works \bowtie_{works.company\_name=company.company\_name}$ company))

9. Assume that the database schema consists of four relations, whose schemas are:

Product(maker, model, type)
PC(model, speed, ram, hd, price)

Laptop(model, speed, ram, hd, screen, price)
Printer(model, color, type, price)

We also give you a sample snapshot of these relations see below in figure. Write expressions of relational algebra to answer the following queries. Your answer should work for arbitrary data, not just the data of these figures. We also assume for convenience purposes that the model numbers are unique across all the different manufacturers and across all the product types.

maker	*model*	*type*
A	1001	pc
A	1002	pc
A	1003	pc
A	2004	laptop
A	2005	laptop
A	2006	laptop
B	1004	pc
B	1005	pc
B	1006	pc
B	2007	laptop
C	1007	pc
D	1008	pc
D	1009	pc
D	1010	pc
D	3004	printer
D	3005	printer
E	1011	pc
E	1012	pc
E	1013	pc
E	2001	laptop
E	2002	laptop
E	2003	laptop
E	3001	printer
E	3002	printer
E	3003	printer
F	2008	laptop
F	2009	laptop
G	2010	laptop
H	3006	printer
H	3007	printer

Figure 2: *
Sample data for Product

model	*speed*	*ram*	*hd*	*price*
1001	2.66	1024	250	2114
1002	2.10	512	250	995
1003	1.42	512	80	478
1004	2.80	1024	250	649
1005	3.20	512	250	630
1006	3.20	1024	320	1049
1007	2.20	1024	200	510
1008	2.20	2048	250	770
1009	2.00	1024	250	650
1010	2.80	2048	300	770
1011	1.86	2048	160	959
1012	2.80	1024	160	649
1013	3.06	512	80	529

Figure 3: *
Sample data for relation PC

model	*speed*	*ram*	*hd*	*screen*	*price*
2001	2.00	2048	240	20.1	3673
2002	1.73	1024	80	17.0	949
2003	1.80	512	60	15.4	549
2004	2.00	512	60	13.3	1150
2005	2.16	1024	120	17.0	2500
2006	2.00	2048	80	15.4	1700
2007	1.83	1024	120	13.3	1429
2008	1.60	1024	100	15.4	900
2009	1.60	512	80	14.1	680
2010	2.00	2048	160	15.4	2300

Figure 4: *
Sample data for relation Laptop

model	*color*	*type*	*price*
3001	true	ink-jet	99
3002	false	laser	239
3003	true	laser	899
3004	true	ink-jet	120
3005	false	laser	120
3006	true	ink-jet	100
3007	true	laser	200

Figure 5: *
Sample data for relation Printer

(a) What PC models have a speed of at least 3.00?

$R1 \coloneqq \sigma_{speed \geq 3.00}(PC)$

$R2 \coloneqq \pi_{model(R1)}$

(b) Which manufactures make laptops with a hard disk of at least 100GB?

$R1 \coloneqq \sigma_{hd \geq 100}(Laptop)$

$R2 \coloneqq R1 \bowtie Product$

$R3 \coloneqq \pi_{maker}(R2)$

(c) Find the model number and price of all products (of any type) made by manufacturer B?

$R1 \coloneqq \sigma_{maker=B}(Product \bowtie PC)$

$R2 \coloneqq \sigma_{maker=B}(Product \bowtie Laptop)$

$R3 \coloneqq \sigma_{maker=B}(Product \bowtie Printer)$

$R4 \coloneqq \pi_{model,price}(R1)$

$R5 \coloneqq \pi_{model,price}(R2)$

$R6 \coloneqq \pi_{model,price}(R3)$

$R7 \coloneqq R4 \cup R5 \cup R6$

(d) Find the model numbers of all color laser printers?

$R1 \coloneqq \sigma_{color=True\ AND\ type=laser}(Printer)$

$R2 \coloneqq \pi_{model}(R1)$

(e) Find those manufacturers that sell Laptops, but not PCs?

$R1 \coloneqq \sigma_{type=laptop}(Product)$

$R2 \coloneqq \sigma_{type=PC}(Product)$

$R3 \coloneqq \pi_{maker}(R1)$

$R4 \coloneqq \pi_{maker}(R2)$

$R4 \coloneqq R3 - R4$

(f) Find the manufacturers who sell exactly three different models of PC.

$R1 \coloneqq \pi_{maker,model}(Product \bowtie PC)$

$R2 \coloneqq \rho_{R2(maker2,model2)}(R1)$

$R3 \coloneqq \rho_{R3(maker3,model3)}(R1)$

$R4 \coloneqq \rho_{R4(maker4,model4)}(R1)$

$R5 \coloneqq R1 \bowtie_{(maker=maker2\ AND\ model<>model2)} R2$

$R6 \coloneqq R3 \bowtie_{(maker=maker3\ AND\ model3<>model2\ AND\ model3<>model)} R5$

$R7 \coloneqq R4 \bowtie_{(maker=maker4\ AND(\ model4=\ model\ OR\ model4=model2\ OR\ model4=model3\ )} R6$

$R8 \coloneqq \pi_{maker}(R7)$