Introduction to Database Systems

Fall 2018

Instructor: Mas-ud Hussain

TAs: Yijiao Chen, Guan Guan, Weizheng Xu

Instructions Name'. Netid'.

Your midterm should have 10 problems distributed on 9 pages (including this front page), for a total of 110 points including 10 bonus points. Please state clearly any assumptions you think you need to make in order to solve a particular problem. You have 80 minutes to complete the quiz, and you should feel free to write on the back of each page. Plan your time wisely. Good luck!!

Problem 1. Assume we have two relations R(A,B) and S(B,C). All three attributes (A, B, and C) are integer attributes. Assume that Relation R contains the following tuples: (1,2), (2,3), and (3,4). Assume that Relation S contains the following tuples (2,2), (2,3) and (5,1).

(a) (*1 pts*) Is A a possible key for R? (yes/no):

(b) (*1 pts*) Is (B, C) a candidate key for S? (yes/no): (Because, (B,C) is not minimal, i.e., {C} itself is a key)

(c) (*1 pts*) How many tuples are in the result of the natural join between R and S? Answer (a number): 2

(d) (*2 pts*) Propose a valid primary key for each of the relation, R and S.

Answer: R(A), S(C)

Problem 2. (*5 pts*) Given two relations R1 and R2, where R1 contains N1 tuples, R2 contains N2 tuples, and N2 > N1 > 0, give the minimum and maximum possible sizes (in tuples) for the resulting relation produced by each of the following relational algebra expressions in the table.

Expression	Assumption	Min	Max
$R1 \cup R2$	$R1$ and $R2$ are union-compatible	N2	N1 + N2
$R1 - R2$	$R1$ and $R2$ are union-compatible	0	N1
$R1 \times R2$		N1 X N2	N1 X N2
$\sigma_{a=5}(R1)$	$R1$ has an attribute named a	0	N1
$\pi_a(R1)$	$R1$ has attribute a, N1>0	1	N1

Problem 3. (*9 pts*) Assume, we have two relations R(A,B) and S(B C). Also, assume all join operations are natural joins. Please select one answer from multiple options.

(a) Q1: $\sigma_{A=3}(R \bowtie S)$; Q2:$(\sigma_{A=3}(R)) \bowtie S$

(i) Q1 and Q2 produce the same answer.

(ii) The answer to Q1 is always contained in the answer to Q2.

(iii) The answer to Q2 is always contained in the answer to Q1.

(iv) Q1 and Q2 produce completely different answers.

In both cases, JOIN is done based on column 'B', so, selection of A=3 in either way produces same result.

(b) Q1: SELECT DISTINCT * FROM R; Q2: (SELECT * FROM R) INTERSECT (SELECT * FROM R);

(i) Q1 and Q2 produce the same answer. [They both produce SELECT * FROM R]

(ii) The answer to Q1 is always contained in the answer to Q2.

(iii) The answer to Q2 is always contained in the answer to Q1.

(iv) Q1 and Q2 produce completely different answers.

(c) In the SQL 3-valued logic, what is the value of the following expression:

R.a > R.b OR R.a ≤ 0 OR R.b ≥ 0

(i) Only TRUE or FALSE.

(ii) Only FALSE or UNKNOWN.

(iii) Only TRUE or UNKNOWN.

(iv) Any of TRUE, FALSE, or UNKNOWN.

To be false, all must be false.
If, R.a > R.b is false, then
R.a ≤ R.b —(1)
If Ra ≤ 0 is false, then R.a > 0 (2)
But, then R.b ≥ 0 is true from (1) and (2).

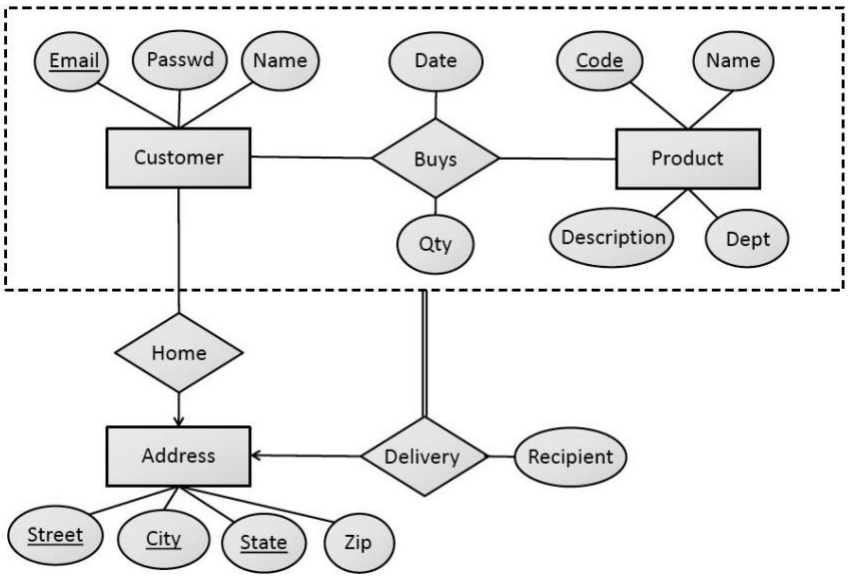


Figure 1: ERD for Problem 4.

Problem 4. (10 pts) Consider the ER diagram presented in Figure 1. Write the schemas of the corresponding relations using the relational model representation. Please explicitly indicate/mention the foreign keys, and any other constraints such as Not Null.

Address(Street, City, State, Zip)

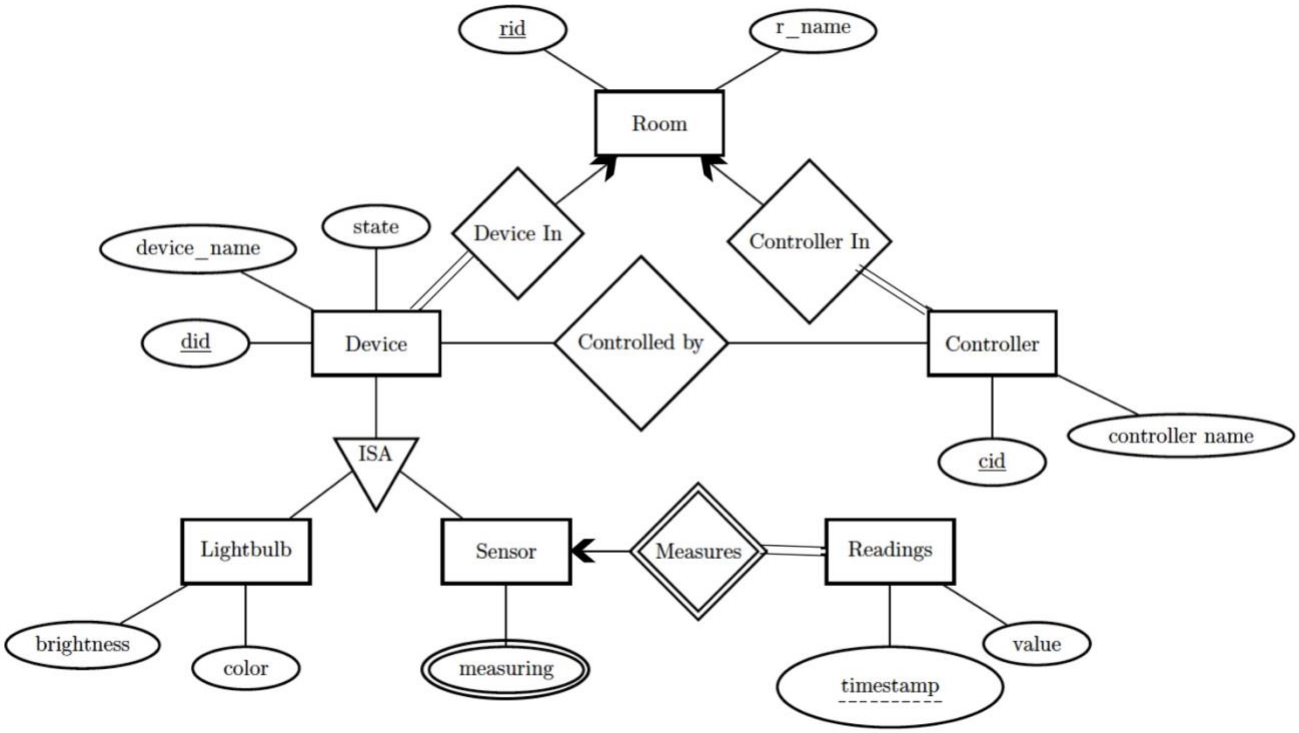
Customer(Email, Passwd, Name, Street, City, State)
Foreign key: Street, City, State (Address)

Product(Code, Name, Dept, Description)

Buys(CustEmail, ProdCode, Date, Qty, Recipient, Street, City, State)
Foreign keys: CustEmail (Customer), ProdCode (Product), Street, City, State (Address)
Not NULL: Street, City, State, Recipient

Problem 5.(20 pts) You are designing a database for your smart home. Your home has devices and controllers, each of which must be located in a particular room, and each room may have multiple devices and controllers. All devices have an ID, name and record their state as being either on or o . A light bulb is a specialized device that also keeps track of its brightness and color. Your home also supports specialized sensor devices, and their entities tell the user what they are measuring (possibly multiple values) currently. Sensor devices write to a measurement table with their device ID, a timestamp, and the measurement. Controllers command devices and they have a name and identifier. Individual devices may be controlled by one or more controllers, such as Amazon Echo, or a third-party app like IFTTT. Draw the ER-diagram, including relationship cardinalities and participation constraints.

ERD SOLUTION:



Problem 6. (*9 pts*) Short answers.

(a) List at least four advantages of using a database management system over writing a C-style program for a data-intensive application.

1. Data Independence
2. Security
3. Data Administration
4. Crash Recovery
5. Concurrency
6. Efficient Queries

} Any 4

(b) State the purpose of using ALL in SQL after SELECT/EXCEPT/UNION/INTERSECT.

To make the query a "bag" operation.

(c) Suppose, we have a functional dependency f: $X \rightarrow Y$ for a relation R. Write down the three conditions, any of which if true, ensures that f does not violate 3NF over R.

(1) f is a trivial FD; or

(2) X contains a key; OR

(3) Y is part of some key.

Problem 7. (*15 pts*) Consider the following relational schema:

`Product(pid, name, price, mfgr) [Primary Key: pid]`

`Buys(cid, pid) [Primary Key: (pid, cid)]`

`Customer(cid, cname, age) [Primary Key: cid]`

You can assume that no two products have the same price. Write (or, rewrite) the following queries in relational algebra.

(a) *Find the names of all customers who have purchased all products that are not manufactured by Sears.*

$\rho(NotSears, \sigma_{mfgr \neq 'Sears}(Product))$ (find Product not made by Sears)
$\rho(disqualified, \pi_{cid}((\pi_{cid}(Buys) \times NotSears) - Buys))$ (find tuples disqualified by division)
$\rho(division, \pi_{cid}(Buys) - disqualified))$ (find cids that match all in NotSears)
$\pi_{name}(Customers \bowtie_{cid} division)$ (display customer names who match all in NotSears)

(You can Use direct division as well)

(b) *Find the names and cids of all customers who have purchased the most expensive product.*

$\rho(p1, Product)$
$\rho(p2, Product)$
$\rho(selfJoin, p1 \bowtie_{p1.price<p2.price} p2)$
$\rho(max, \pi_{p2.price}(Product) - \pi_{p1.price}(selfJoin))$
$\pi_{Customers.cid,name}(max \bowtie_{price} Product \bowtie_{pid} Buys \bowtie_{cid} Customers)$

(c)

```
SELECT C.cid FROM Customer C, Buys B, Product P
WHERE C.cid = B.cid and B.pid = P.pid GROUP BY C.cid HAVING count(*) > 3
```

$\sigma_{cc>3}\left({}_{C.cid}\mathcal{G}_{count(*)\ as\ cc}\left(\rho_c(Customer) \bowtie Buys \bowtie Product\right)\right)$

Problem 8. (*15 pts*) Write the following queries in SQL using the given relational schema:

`Cocktails (cname, price) [Primary Key: cname]`

`Ingredients (iname, unit_cost, ABV) [Primary Key: iname]`

`Recipes (cname, iname, units) [Primary Key: (cname, iname)]`

(a) *List the names of ingredients that either cost less than $1 or are not used in any recipes.*

```
select iname
from   Ingredients
where  unit_cost < 1
UNION
(select iname
 from   Ingredients
 EXCEPT
 select I.iname
 from   Ingredients I, Recipes R
 where   I.iname = R.iname );
```

(b) *List pairs of ingredients that are used in the same cocktail.* Result should have the schema (iname1, iname2, cname). Do not list the same pair more than once, i.e., you should not list both (gin, lemon juice, Gimlet) and (lemon juice,gin, Gimlet).

```
select R1.iname, R2.iname, R1.cname
from   Recipes R1, Recipes R2
where  R1.cname = R2.cname
and    R1.iname < R2.iname;
```

(c) *List names of cocktails together with the profit made on the sale of each of them.* For a cocktail, we define profit as the difference between its price and the combined cost of their ingredients. Order results by profit from higher to lower.

```
select C.cname, C.price - SUM (R.units * I.unit_cost) as profit
from   Cocktails C, Recipes R, Ingredients I
where  C.cname = R.cname
and    R.iname = I.iname
group by C.cname, C.price
order by profit desc;
```

R:

A	B	C
Anne	04/93	A
Bob	09/71	A
Carol	12/77	D
Dave	05/90	X
Erin	04/85	A
Frank	08/78	B
Grace	10/83	A

and P:

C	D
A	401k
B	403b
C	401a
D	457

Figure 2: Relation instances of R and P.

Problem 9. (*10 pts*) Assume that you are given the instances of the two relations, R and P, as shown in Figure 2. What is the effect (i.e., the output) of the following operations:

(a) $\pi_C(\sigma_{A \leq \text{'Drew'}}(R)) \ltimes P$

C	D
A	401K
D	457

(b) SELECT * FROM (SELECT C FROM R WHERE A < 'Frank') as Q LEFT OUTER NATURAL JOIN P ORDER BY C DESC:

C	D
X	NULL
D	457
A	401K
A	401K
A	401K

Problem 10.

(a) (*6 points*) Consider relation R (A<u>B</u>CD<u>E</u>). For each of the following functional dependencies, state whether or not it violates BCNF and 3NF.

(i) A → B

Violates BCNF? (yes/no): yes

Violates 3NF? (yes/no): no

(ii) B → D

Violates BCNF? (yes/no): yes

Violates 3NF? (yes/no): yes

(iii) DBE → DA

Violates BCNF? (yes/no): no

Violates 3NF? (yes/no): no

(b) (*6 points*) Let relation R(A,B,C,D,E,F,G,H) satisfy the following functional dependencies, $\mathbb{F}$: {A → B, CH → A, B → E, BD → C, EG → H, BDE → F}. Show that AD → F is in $\mathbb{F}^+$.

1. A → B (GIVEN)
2. B → E (GIVEN)
3. A → BE (TRANSITIVITY, 1&3)
4. AD → BDE (AUGMENTATION)
5. BDE → F (GIVEN)
6. AD → F (TRANS. / ✱ 485)