

Databases

Lecture 12

Written Exam - Problems

I

I Choose the correct answer(s) for the following multiple choice questions. Each question has at least one correct answer.

1. In a SELECT query:

FROM
WHERE
GROUP BY
HAVING
SELECT
DISTINCT
ORDER BY
TOP

- ☒ a. the SELECT clause can contain arithmetic expressions
- ☒ b. according to the conceptual evaluation strategy, ORDER BY is evaluated before GROUP BY
- ☒ c. HAVING can contain row-level qualification conditions → works in groups
- ☒ d. DISTINCT eliminates duplicates from the answer set
- ☐ e. none of the above answers is correct.

2. The natural join operator $R_1 * R_2$ in the relational algebra:

- ~~a.~~ returns a relation whose schema contains only the attributes in R_1 that don't appear in R_2 $R - R_2$
- b.** returns a relation whose schema contains all the attributes in R_1 and R_2 , with common attributes appearing only once
- c. returns 2 relation instances \rightarrow returns a single relation
- d. is not associative \rightarrow it is! $(A * B) * C = A * (B * C)$
- e. none of the above answers is correct.

3. In the ANSI-SPARC architecture (for a database system), a database can have:

a. exactly one **symbolic structure** *doesn't exist!!*

b. several conceptual structures (schema *just ONE*)

☒ c. several external structures → views

☐ d. several physical structures → indexes, files, *just one*

e. none of the above answers is correct.

4. Consider relation $S[\underline{A}, \underline{B}, \underline{C}, D, E, F, G, H]$ with:

- primary key $\{A, B, C\}$, no other candidate keys;
- functional dependencies that are known to hold over S : $\{F\} \rightarrow \{H\}$, $\{C\} \rightarrow \{E, G\}$; *redundancy! (dependent on a subset)*
- no repeating attributes.

a. S is not 1NF \rightarrow false, it is in 1NF since it doesn't have repeating attributes

~~b. S is 2NF~~ *not fully functionally dep!!*

c. S is not BCNF

d. S is 3NF *non-prime attr. H is transitively dep. on a key*
 $F \rightarrow H$, $C \rightarrow E, G$ \leftarrow these dependencies that make it not on

e. none of the above answers is correct.

$$F \rightarrow H$$
$$C \rightarrow E, G$$

$$\begin{aligned} [8] &= 4 \\ [4, 8] \end{aligned}$$

5. In a B-tree of order 8:

- a. a non-terminal node has at most 8 subtrees
- b. a non-terminal node with 7 values has 8 subtrees
- c. terminal nodes can be on different levels
- d. a non-terminal node with 7 values has 6 subtrees
- e. none of the above answers is correct.

6. Let α , β and γ be subsets of attributes in a relational schema. If $\alpha \rightarrow \beta$ and $\beta \rightarrow \gamma$, then by transitivity:

a. $\alpha \rightarrow \gamma$

b. $\gamma \rightarrow \alpha$

c. $\beta \rightarrow \alpha$

d. $\gamma \rightarrow \beta$

e. none of the above answers is correct.

7. Let RepairLog[RID, MechanicID, RollerCoasterID, RepairTime] be a table in a SQL Server database. RepairLog has 100.000 records and 2 indexes: a unique clustered index on RID and a non-clustered index on MechanicID.

Consider the following query:

```
SELECT RID, MechanicID, RepairTime  
FROM RepairLog  
WHERE MechanicID = 7
```

If the execution plan contains an Index Seek (NonClustered), it also contains a:

- a. *Clustered Index Scan*
 - b. *Index Scan (NonClustered)*
 - ☒ c. *Key Lookup (Clustered)*
 - d. *Index Trick (NonClustered)*
 - e. none of the above answers is correct.
- used because we make a search on mechanic ID*

8-10. Consider the relational schema $S[\underline{FK1}, \underline{FK2}, A, B, C, D, E]$, with primary key $\{FK1, FK2\}$. Answer questions 8-10 using the legal instance below:

↳ all integrity constraints

FK1	FK2	A	B	C	D	E
1	1	a1	b1	c1	7	2
1	2	a_	b3	c1	5	2
1	3	a2	b1	c2	Null	2
2	1	a3	b3	c2	Null	100
2	2	a3	b3	c3	Null	100

8. Consider queries Q_1 and Q_2 :

Q_1 :

SELECT *

FROM S s1 LEFT JOIN S s2 ON s1.FK1 = s2.E

Q_2 :

SELECT DISTINCT *

FROM S s1 INNER JOIN S s2 ON s1.FK1 = s2.E

The cardinality of the answer set of Q_i is denoted by $|Q_i|$.

$|Q_1| - |Q_2|$ is: $|Q_1| = 9$

a. 0

b. 3

c. 10

d. 2

e. none of the above answers is correct.

FK1	FK2	A	B	C	D	E
1	1	a1	b1	c1	7	2
1	2	a_	b3	c1	5	2
1	3	a2	b1	c2	Null	2
2	1	a3	b3	c2	Null	100
2	2	a3	b3	c3	Null	100

S[FK1, FK2, A, B, C, D, E]

Handwritten notes showing pairs of FK1, FK2 values and their corresponding D values:

- 2, 1 1, 1
- 2, 1 1, 2
- 2, 1 1, 3
- 2, 2 1, 1
- 2, 2 1, 2
- 2, 2 1, 3
- 1, 1 NULL
- 1, 2 NULL
- 1, 3 NULL

9. Regarding the functional dependencies of S:

a. at least one of the following dependencies is not satisfied by the instance: $\{A\} \rightarrow \{B\}$,

$\{FK1, FK2\} \rightarrow \{A, B\}$, $\{FK1\} \rightarrow \{A\}$

b. by examining the instance, we can conclude that at least one of the following

dependencies is specified on the schema S: $\{A\} \rightarrow \{B\}$, $\{FK1\} \rightarrow \{A, B\}$, $\{FK1\} \rightarrow \{A\}$

~~c.~~ at least two of the following dependencies are not satisfied by the instance:

$\{FK2\} \rightarrow \{A, B\}$, $\{A\} \rightarrow \{E\}$, $\{A, B\} \rightarrow \{E\}$, $\{FK1, FK2\} \rightarrow \{E\}$

d. by examining the instance, we can conclude that at least two of the following dependencies are specified on the schema S: $\{FK2\} \rightarrow \{A, B\}$, $\{A\} \rightarrow \{E\}$, $\{A, B\} \rightarrow \{E\}$, $\{B\} \rightarrow \{C, E\}$

e. none of the above answers is correct.

FK1	FK2	A	B	C	D	E
1	1	a1	b1	c1	7	2
1	2	a_	b3	c1	5	2
1	3	a2	b1	c2	Null	2
2	1	a3	b3	c2	Null	100
2	2	a3	b3	c3	Null	100

S[FK1, FK2, A, B, C, D, E]

10. Consider queries Q_1 and Q_2 :

Q_1 :

SELECT FK2, FK1, COUNT(DISTINCT B)

FROM S

GROUP BY FK2, FK1

HAVING FK1 = 0 ✓

HAVING A = ?... would be syntax error

Q_2 :

SELECT FK2, FK1, COUNT(C)

FROM S

GROUP BY FK2, FK1

HAVING MAX(E) < 0

The cardinality of the answer set of Q_i is denoted by $|Q_i|$.

FK1	FK2	A	B	C	D	E
1	1	a1	b1	c1	7	2
1	2	a_	b3	c1	5	2
1	3	a2	b1	c2	Null	2
2	1	a3	b3	c2	Null	100
2	2	a3	b3	c3	Null	100

S[FK1, FK2, A, B, C, D, E]

->

10.





$|Q_1| - |Q_2|$ is:

☒ a. 0

b. 2

c. 1

d. -1

e. none of the above answers is correct.

FK1	FK2	A	B	C	D	E
1	1	a1	b1	c1	7	2
1	2	a_	b3	c1	5	2
1	3	a2	b1	c2	Null	2
2	1	a3	b3	c2	Null	100
2	2	a3	b3	c3	Null	100

$S[\underline{FK1, FK2}, A, B, C, D, E]$

11. A secondary index: \Rightarrow SK doesn't include PK

- a. can contain duplicates
- b. cannot contain duplicates
- c. can be non-clustered \rightarrow same or close to the physical model
- d. cannot be non-clustered
- e. none of the above answers is correct.

12. For the relation $R[A, B, C]$ below, consider the 3 possible projections on 2 attributes: $AB[A, B]$, $BC[B, C]$, $AC[A, C]$. How many extra records does $AB * BC * AC$ contain (i.e., records that don't appear in R)?

A	B	C
a1	b2	c1
a1	b1	c2
a2	b1	c1

a. 0

b. 1

c. 2

d. 3

~~e. none of the above answers is correct.~~

$$\frac{4 \cdot 4 \cdot 4 = 64 - 3}{61}$$

$AB * BC$:

- a_1, b_2, c_1 ✓
- a_1, b_1, c_2 ✓
- a_1, b_2, c_1 ✓
- a_2, b_1, c_2 ✗
- a_1, b_1, c_1 ✓

AB :

- a_1, b_1
- a_1, b_2
- a_2, b_1
- ~~a_2, b_2~~

BC :

- b_1, c_1
- ~~b_2, c_2~~
- b_1, c_2
- b_2, c_1

AC :

- a_1, c_1
- a_1, c_2
- a_2, c_1
- ~~a_2, c_2~~

13. The cross-product operator $R_1 \times R_2$ in the relational algebra:

- ☒ a. returns a relation whose schema contains all the attributes in R_1 followed by all the attributes in R_2
- ☒ b. returns a relation whose schema contains only the attributes in R_1
- ☒ c. returns 3 relation instances \rightarrow returns 1
- ☒ d. is associative
- ☐ e. none of the above answers is correct.



II Answer the following questions / solve the following problems.

1. Rewrite the CREATE TABLE statements below such that the following restriction is enforced: one T1 entity can be associated with any number of T2 entities, and one T2 entity can be associated with at most one T1 entity. Don't add other SQL statements.

```
CREATE TABLE T1  
(IDT1 INT PRIMARY KEY,  
C1 VARCHAR(100))
```

```
CREATE TABLE T2
```

```
(IDT2 INT PRIMARY KEY,
```

```
C2 DATE), IDT1 INT REFERENCES T1 (IDT1)
```

2. Write the relational algebra expression below as a SQL query.

$$\pi_{\{A,B,C\}}(\underbrace{(\pi_{\{A,B,ID\}}(\sigma_{M=70}(R)))}_{\text{SELECT A, B, ID FROM R WHERE M=70}} \otimes_{ID=IDT1} \underbrace{(\pi_{\{C,IDT1\}}(\sigma_{N>5}(S)))}_{\text{SELECT C, IDT1 FROM S WHERE N>5}})$$

SELECT A, B, ID
FROM R
WHERE M = 70

SELECT C, IDT1
FROM S
WHERE N > 5

SELECT ABC

FROM

(SELECT A, B, ID
FROM R
WHERE M = 70) t1

INNER JOIN

(SELECT C, IDT1
FROM S
WHERE N > 5) t2

ON t1.ID = t2.IDT1

3-6. Consider the relational schema $R[\underline{RID}, A, B, C, D, E, F]$, with primary key $\{RID\}$. Answer questions 3-6 using the legal instance below:

RID	A	B	C	D	E	F
1	100	200	5	200	20	11
2	101	50	11	200	5	12
3	100	100	7	200	5	13
4	200	200	6	200	20	14
5	200	100	2	200	5	9
6	300	50	11	200	5	10

3. What's the result set returned by the following query? Write the tuples' values and the names of the columns.

SELECT r1.RID, r1.A + r2.A C2, r1.A * r2.A C3
 FROM R r1 LEFT JOIN R r2 ON r1.RID = r2.RID
 WHERE r1.A > ANY (SELECT B
 ^{MIN} FROM R
 WHERE C < 10)

911 > 100
 200
 100
 200
 100

RID	A	B	C	D	E	F
1	100	200	5	200	20	11
2	101	50	11	200	5	12
3	100	100	7	200	5	13
4	200	200	6	200	20	14
5	200	100	2	200	5	9
6	300	50	11	200	5	10

R[RID, A, B, C, D, E, F]

RID	C2	C3
2	202	10201
4	400	160000
5	400	160000
6	600	360000

4. Evaluate the expressions below. π doesn't eliminate duplicates. What's the cardinality of T ?

$= 24$ for sets = (4)

$S := \sigma_{F < 13}(R)$

$T := \pi_{\{S.RID, S.A\}}(S \otimes_{S.D=R.D} R)$

$\Rightarrow 4 * 6 = 24$ rows

RID	A	B	C	D	E	F
1	100	200	5	200	20	11
2	101	50	11	200	5	12
3	100	100	7	200	5	13
4	200	200	6	200	20	14
5	200	100	2	200	5	9
6	300	50	11	200	5	10

the other way $R[\underline{RID}, A, B, C, D, E, F]$

1 1
1 2
1 5
1 6
2 1
2 2
2 5
2 6
3 1
3 2
3 5
3 6
:

5. What's the result set returned by the following query? Write the tuples' values and the names of the columns.

SELECT R.*

RID A B C D E F
- - - - -

RID A B
N N N

select the eq.

FROM

(SELECT r1.RID, r2.A, r3.B

FROM R r1 INNER JOIN R r2 ON r1.A = r2.B

INNER JOIN R r3 ON r2.B > r3.D

WHERE r1.F > 10) t → 0 rows

RIGHT JOIN R ON t.RID = R.RID

RID	A	B	C	D	E	F
1	100	200	5	200	20	11
2	101	50	11	200	5	12
3	100	100	7	200	5	13
4	200	200	6	200	20	14
5	200	100	2	200	5	9
6	300	50	11	200	5	10

R[RID, A, B, C, D, E, F]

1 R.RID	2 R.RID	A=B	3 R.RID	R3.D
1	3	100	X	
1	5	100	X	
3	3	100	X	
3	5	100	X	
4	1	200	X	
4	4	200	X	
5	1	200	X	
5	4	200	X	

6. Write all functional dependencies F such that:

- F is satisfied by the current instance of R;
- the dependent of F is $\{D\}$;
- the determinant of F has a single column.

RID	A	B	C	D	E	F
1	100	200	5	200	20	11
2	101	50	11	200	5	12
3	100	100	7	200	5	13
4	200	200	6	200	20	14
5	200	100	2	200	5	9
6	300	50	11	200	5	10

R[RID, A, B, C, D, E, F]

$RID \rightarrow D$
 \vdots
All Cols $\rightarrow D$

7. Rewrite the expression below using only operators in the set $\{\sigma, \pi, \times, \cup, -\}$.

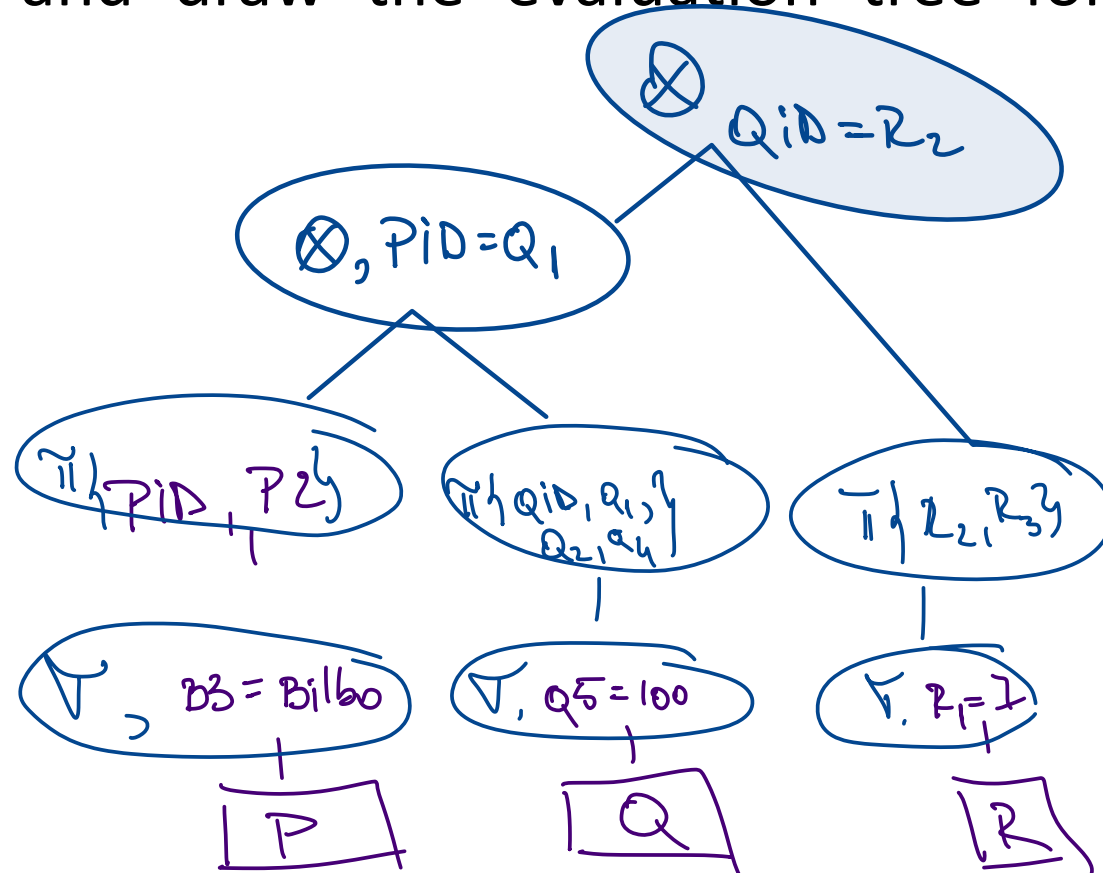
$$(R \bowtie_{R.ID=S.RID} S) \cap (T \bowtie_{T.ID=U.TID} U)$$

$$\bigvee_{R.ID=S.ID} (R \times S) - \left(\bigvee_{R.ID=S.ID} (R \times S) - \bigvee_{T.ID=U.ID} (T \times U) \right)$$

8. Let P, Q, R be 3 relations with schemas P[PID, P1, P2, P3], Q[QID, Q1, Q2, Q3, Q4, Q5], R[RID, R1, R2, R3], and E an expression in the relational algebra:

$$E = \pi_{\{P2, Q2, Q4, R3\}} (\sigma_{\{PID=Q1 \text{ AND } QID=R2 \text{ AND } P3='Bilbo' \text{ AND } Q5=100 \text{ AND } R1=7\}} (P \times Q \times R))$$

Optimize E and draw the evaluation tree for the optimized version of the expression.



Extra - Design problem

A musical instruments manufacturer relies on a relational database to support its activities. You are asked to design a part of the database schema. The manufacturer produces instruments of different categories. A **category** has: name (e.g., string, percussion, woodwind, etc), description, and several subcategories. A **subcategory** belongs to a category and has a name (e.g., violin, piano, contrabass, etc). An **instrument** currently in stock belongs to a subcategory and has: serial number, date of manufacture, color, and price. A **customer** has: name, score, and type (e.g., music academy, orchestra, etc). The company takes orders directly from customers, online or by phone. An **order** is placed by a customer and has: 2 dates – the date when the order was made and the date when it was honored (*null* for unfulfilled orders), a field indicating whether it's been placed online or by phone, and, for each subcategory of instruments in the order, the number of ordered instruments of each color (e.g., an order for 7 red violins, 3 white violins, 2 white pianos, and a yellow contrabass).

