Databases
Lecture 12 Written Exam - Problems
SOLUTIONS

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:
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
d. DISTINCT eliminates duplicates from the answer set
e. none of the above answers is correct.

Correct answer: AD

Lecture 4: the SELECT statement - logical processing (Transact-SQL)
FROM
WHERE
GROUP BY
HAVING
SELECT
DISTINCT
ORDER BY
TOP

HAVING is used to filter the results of GROUP BY and typically applies to aggregate functions rather than row-level conditions, which are normally specified in the WHERE clause.

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

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
d. is not associative
e. none of the above answers is correct.

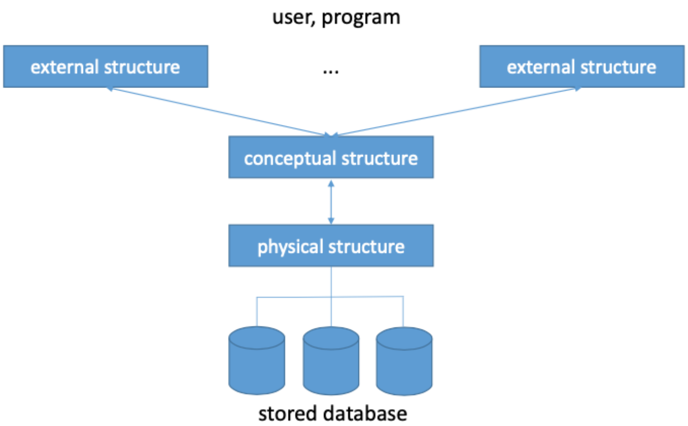
Correct answer: B

Lecture 6: *natural join*
•notation:$R *R$ 12
• resulting relation: • schema: the union of the attributes of the two relations (attributes with the same name in R and R appear once in the result)

3. In the ANSI-SPARC architecture (for a database system), a database can have: a. exactly one symbolic structure
b. several conceptual structures
c. several external structures
d. several physical structures
e. none of the above answers is correct.

Correct answer: C

Lecture 1:



4. Consider relation S[A, B, 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} → {H}, {C} → {E, G};
- no repeating attributes.

a. S is not 1NF
b. S is 2NF
c. S is not BCNF
d. S is 3NF
e. none of the above answers is correct.

Correct answer: C

Lecture 4 + Lecture 5:
First Normal Form (1NF): A relation is in 1NF if all the attributes contain atomic values. / A relation is in the first normal form (1NF) if it doesn't have any repeating attributes.
Since it is mentioned that there are no repeating attributes, we can conclude that S is in 1NF.

Second Normal Form (2NF): A relation is in 2NF if it is in 1NF and all non-prime attributes are fully functionally dependent on the primary key. The primary key is {A, B, C}.
For the functional dependencies:
{F} → {H}: H is not dependent on the primary key, but rather on F which is not part of the primary key; thus, it violates 2NF.
{C} → {E, G}: E and G are dependent on part of the primary key (C), also violating 2NF.
Therefore, S is not in 2NF.

Boyce-Codd Normal Form (BCNF): A relation is in BCNF if for every non-trivial functional dependency (X → Y), (X) is a super key.
{F} → {H} does not meet the criteria since F is not a super key. Consequently, S is not in BCNF.

Third Normal Form (3NF): A relation is in 3NF if it is in 2NF and there are no transitive dependencies.
Since S is not in 2NF, it cannot be in 3NF.

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.

Correct answer: AB

Lecture 9: B-tree of order m 1. if the root is not a terminal, it has at least 2 subtrees 2. all terminal nodes – same level 3. every **non-terminal node – at most m subtrees** 4. a node with **p subtrees has p-1 ordered values** (ascending order): $K_1<K_2<...<K_{p-1}$ ● A_1: values less than K_1 ● A_i: values between K_{i-1} and K_i, i=2,...,p-1 ● A_p: values greater than K_{p-1} • obs. limits on number of subtrees (and values) / node result from the manner in which inserts / deletes are performed so that the second requirement in the definition is met 5. every **non-terminal node – at least [m/2] subtrees** • **a node can have at most m subtrees**, (a maximum of m-1 values), and **at least [m/2] subtrees**, (at least [m/2]-1=[(m-1)/2] values).

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.

Correct answer: A

Lecture 5: 4. If $\alpha \rightarrow \beta$ and $\beta \rightarrow \gamma$, then $\alpha \rightarrow \gamma$ - *transitivity*.

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.

Correct answer: C

Lecture 8 / 9 / 10 + Seminar 5:

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

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 Q1 and Q2:

Q1:
SELECT *
FROM S s1 LEFT JOIN S s2 ON s1.FK1 = s2.E

Q2:
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:

a. 0
b. 3
c. 10
d. 2
e. none of the above answers is correct.

Correct answer: B

Q1:

```
SELECT * FROM S s1 LEFT JOIN S s2 ON s1.FK1=s2.E
```

150 %

Results | Messages

	FK1	FK2	A	B	C	D	E	FK1	FK2	A	B	C	D	E
1	1	1	a1	b1	c1	7	2	NULL	NULL	NULL	NULL	NULL	NULL	NULL
2	1	2	a_	b3	c1	5	2	NULL	NULL	NULL	NULL	NULL	NULL	NULL
3	1	3	a2	b1	c2	NULL	2	NULL	NULL	NULL	NULL	NULL	NULL	NULL
4	2	1	a3	b3	c2	NULL	100	1	1	a1	b1	c1	7	2
5	2	1	a3	b3	c2	NULL	100	1	2	a_	b3	c1	5	2
6	2	1	a3	b3	c2	NULL	100	1	3	a2	b1	c2	NULL	2
7	2	2	a3	b3	c3	NULL	100	1	1	a1	b1	c1	7	2
8	2	2	a3	b3	c3	NULL	100	1	2	a_	b3	c1	5	2
9	2	2	a3	b3	c3	NULL	100	1	3	a2	b1	c2	NULL	2

Q2:

```
SELECT DISTINCT * FROM S s1 INNER JOIN S s2 ON s1.FK1=s2.E
```

150 %

Results | Messages

	FK1	FK2	A	B	C	D	E	FK1	FK2	A	B	C	D	E
1	2	1	a3	b3	c2	NULL	100	1	1	a1	b1	c1	7	2
2	2	2	a3	b3	c3	NULL	100	1	1	a1	b1	c1	7	2
3	2	1	a3	b3	c2	NULL	100	1	2	a_	b3	c1	5	2
4	2	2	a3	b3	c3	NULL	100	1	2	a_	b3	c1	5	2
5	2	1	a3	b3	c2	NULL	100	1	3	a2	b1	c2	NULL	2
6	2	2	a3	b3	c3	NULL	100	1	3	a2	b1	c2	NULL	2

9. Regarding the functional dependencies of S:

a. at least one of the following dependencies is not satisfied by the instance: {A} → {B}, {FK1, FK2} → {A, B}, {FK1} → {A}

b. by examining the instance, we can conclude that at least one of the following dependencies is specified on the schema S: {A} → {B}, {FK1} → {A, B}, {FK1} → {A}

c. at least two of the following dependencies are not satisfied by the instance: {FK2} → {A, B}, {A} → {E}, {A, B} → {E}, {FK1, FK2} → {E}

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

e. none of the above answers is correct.

Correct answer: ABD

{A} → {B}

When **A is unique (Primary key)** then the **functional dependency is fulfilled.**

When **A has values that appear more than once** should have **the same correspondent value in B** such that the **functional dependency is fulfilled**.

a. {A} → {B} true, {FK1, FK2} → {A, B} true, {FK1} → {A} false

b. {A} → {B} true , {FK1} → {A, B} false, {FK1} → {A} false

c. {FK2} → {A, B} false, {A} → {E} true, {A, B} → {E} true, {FK1, FK2} → {E} true

d. {FK2} → {A, B} false, {A} → {E} true, {A, B} → {E} true, {B} → {C, E} false

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

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 $|Q_i|$.

$|Q_1|$ - $|Q_2|$ is:

a. 0
b. 2
c. 1
d. -1
e. none of the above answers is correct.

Correct answer: A

Q1:

```sql
-- Q1
SELECT FK2, FK1, COUNT(DISTINCT B)
FROM S
GROUP BY FK2, FK1
HAVING FK1=0
```

150 %

Results | Messages

FK2	FK1	(No column name)

Q2:

```sql
-- Q2
SELECT FK2, FK1, COUNT(C)
FROM S
GROUP BY FK2, FK1
HAVING MAX(E)<0
```

150 %

Results | Messages

FK2	FK1	(No column name)

11. A secondary index:
a. can contain duplicates
b. cannot contain duplicates
c. can be non-clustered
d. cannot be non-clustered
e. none of the above answers is correct.

Correct answer: AC

Lecture 9:
secondary index
• index that is not primary
• secondary indexes can contain duplicates

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.

Correct answer: B

```sql
-- AB * BC
SELECT DISTINCT A, AB.B, C FROM AB INNER JOIN BC ON AB.B=BC.B

-- AB * BC * AC
--SELECT DISTINCT AB.A, BC.B, AC.C FROM AB INNER JOIN BC ON AB.B=BC.B

-- (AB*BC)*AC
SELECT DISTINCT F.A, F.B, AC.C
FROM (SELECT AB.A, AB.B, BC.C FROM AB INNER JOIN BC ON AB.B=BC.B
INNER JOIN AC ON AB.A=AC.A) F INNER JOIN AC ON F.A=AC.A AND F.C=AC.C
```

165 %

Results | Messages

	A	B	C
1	a1	b1	c1
2	a1	b1	c2
3	a1	b2	c1
4	a2	b1	c1
5	a2	b1	c2

	A	B	C
1	a1	b1	c1
2	a1	b1	c2
3	a1	b2	c1
4	a2	b1	c1

By 'hand':
R:

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

AB:

A	**B**
a1	b2
a1	b1
a2	b1

BC:

B	**C**
b2	c1
b1	c2
b1	c1

AB*BC: column B is the column that satisfies the equality AB.**B**=BC.**B**

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

AB*BC*AC = (AB*BC)*AC: column A is the column that satisfies the equality AB*BC.**A**=AC.**A** AND AB*BC.**C**=AC.**C**

AB*BC:

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

AC:

A	**C**
a1	c1
a1	c2
a2	c1

⇨ AB*BC*AC = (AB*BC)*AC

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

So, 1 extra record.

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

d. is associative

e. none of the above answers is correct.

Correct answer: AD

Lecture 6:
cross-product
•notation:*R* ×*R*
• resulting relation:
• schema: the attributes of *R* followed by the attributes of *R*
• tuples: every tuple *r* in *R* is concatenated with every tuple *r* in *R* 11 22
• equivalent SELECT statement
SELECT *
FROM R1 CROSS JOIN R2
Lecture 11:
associativity for the cross-product and the natural join

II

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)

Answer:

```
CREATE TABLE T2
(IDT2 INT PRIMARY KEY,
C2 DATE,
IDT1 INT FOREIGN KEY REFERENCES T1(IDT1)
)
```

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

$$\pi_{\{A,B,C\}}((\pi_{\{A,B,ID\}}(\sigma_{M=70}(R))) \otimes_{ID=IDT1} (\pi_{\{C,IDT1\}}(\sigma_{N>5}(S))))$$

Answer:

```
SELECT A, B, C
FROM
    (SELECT A, B, ID FROM R WHERE M=70)
    INNER JOIN
    (SELECT C, IDT1 FROM S WHERE N>5)
    ON ID=IDT1
```

3-6. Consider the relational schema R[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
FROM R WHERE C < 10)

Answer:

```sql
SELECT r1.RID, r1.A+r2.A C2, r1.A*r2.A C3
FROM RR r1 LEFT JOIN RR r2 ON r1.RID = r2.RID
WHERE r1.A>ANY (SELECT B FROM RR WHERE C<10)
```

150 %

Results | Messages

	RID	C2	C3
1	2	202	10201
2	4	400	40000
3	5	400	40000
4	6	600	90000

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

$$S := \sigma_{F<13}(R)$$
$$T := \pi_{\{S.RID,S.A\}}(S \otimes_{S.D=R.D} R)$$

Answer:

SELECT S.RID, S.A FROM
(SELECT * FROM R WHERE F<13) INNER JOIN R ON R.D=S.D

```sql
SELECT S.RID, S.A FROM
(SELECT * FROM RR WHERE F<13) S
INNER JOIN RR ON S.D=RR.D
```

150 %

Results | Messages

	RID	A
1	1	100
2	1	100
3	1	100
4	1	100
5	1	100
6	1	100
7	2	101
8	2	101
9	2	101
10	2	101
11	2	101
12	2	101
13	5	200
14	5	200
15	5	200
16	5	200
17	5	200
18	5	200
19	6	300
20	6	300
21	6	300
22	6	300
23	6	300
24	6	300

(24 rows)

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

SELECT R.*

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
RIGHT JOIN R ON t.RID = R.RID

Answer:

```
-- 5
SELECT RR.*
FROM
(SELECT r1.RID, r2. A, r3.B
FROM RR r1 INNER JOIN RR r2 ON r1.A=r2.B
INNER JOIN RR r3 ON r2.B>r3.D
WHERE r1.F>10) t
RIGHT JOIN RR ON t.RID=RR.RID
```

150 %

Results Messages

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

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.

Answer:

Lecture 4:
In the dependency $\alpha \rightarrow \beta$, α is the *determinant*, and β is the *dependent*

- So, β={D} (from the dependent of F is *{D}*;)
- And, α has a single column, and all the attributes will fit (from the determinant of F has a single column)

RID → D
A → D
B → D
C → D
E → D
F → D

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

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

Answer:

Lecture 7: $R_1 \cap R_2 = R_1 - (R_1 - R_2)$

$$R_1 = \sigma_{R.ID=S.RID}(R \times S);\ R_2 = \sigma_{T.ID=U.TID}(T \times U)$$

So,

$$\sigma_{R.ID=S.RID}(R \times S) - (\sigma_{R.ID=S.RID}(R \times S) - \sigma_{T.ID=U.TID}(T \times U))$$

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.

Answer:

Lecture 11:

P(<u>PID</u>, P1, P2, P3); P3='Bilbo'
Q(<u>QID</u>, *Q1*, Q2, Q3, Q4, Q5); PID=Q1 (PK=FK); Q5=100
R(<u>RID</u>, R1, *R2*, R3); QID=R2 (PK=FK); R1=7

Denote by C=(PID=Q1 AND QID=R2 AND P3='Bilbo' AND Q5=100 AND R1=7) - the WHERE conditions
and B={P2, Q2, Q4, R3}- the list of attributes to display

So, $E=\pi_B(\sigma_C(P \times Q \times R))$

P×Q×R = (P×Q)×R OR
P×Q×R = P×(Q×R).

Commute σ with ×; the transitivity of the equality operator

PID=Q1	QID=R2	P3='Bilbo'	Q5=100	R1=7
C1	C2	C3	C4	C5
in (P, Q)	in (Q,R)	in P	in Q	in R

So, $\sigma_B(P \times Q \times R) = \sigma_{C1 \text{ AND } C2}((\sigma_{C3}(P)\times\sigma_{C4}(Q))\times\sigma_{C5}(R))$, OR

$\sigma_{C1 \text{ AND } C2}(\sigma_{C3}(P)\times(\sigma_{C4}(Q)\times\sigma_{C5}(R)))$

Replace selection and cross product with condition join:
$= (\sigma_{C3}(P)\otimes_{C1}\sigma_{C4}(Q))\otimes_{C2}\sigma_{C5}(R)$, OR

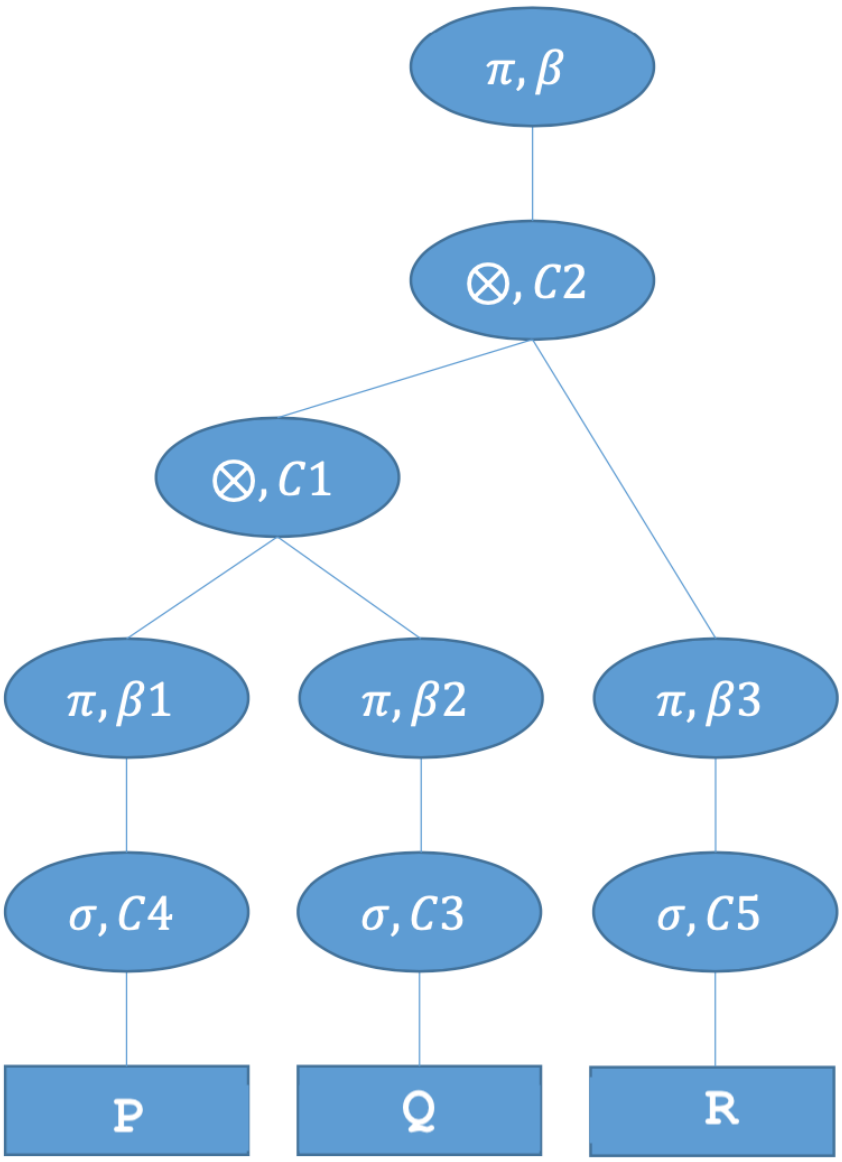
$\sigma_{C3}(P)\otimes_{C1}(\sigma_{C4}(Q))\otimes_{C2}\sigma_{C5}(R))$

Now, we chose, one of the previous versions:
$E=\pi_B((\sigma_{C3}(P)\otimes_{C1}\sigma_{C4}(Q))\otimes_{C2}\sigma_{C5}(R))$
Commute projection (π) with join:
B1={P2}
B2={Q2, Q4}

B3={R3}
useful for B and join.

So, E=$\pi_B(\pi_{B1}(\sigma_{C3}(P))\otimes_{C1}\pi_{B2}(\sigma_{C4}(Q)))\otimes_{C2}\pi_{B3}(\sigma_{C5}(R)))$

An evaluation tree can be constructed for this last version of the relational algebra expression using information from the system catalog and possibly statistical information, an execution plan can be generated from the last version of the expression; every relational operator is replaced by an evaluation algorithm.



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).

Answer:

```
Categories - Subcategories: 1-n
Companies - Customers: m-n (through DirectOrders)
Manufactures - Instruments: 1-n
Subcategories - Instrument: 1-n
Customers - instruments: m-n (through Orders)

-- drop database Lecture12_CS
create database Lecture12_CS
go
use Lecture12_CS
go

create table Manufacturers(
Mid int primary key identity,
MName varchar(50))

create table Categories(
Catid int primary key identity,
Name varchar(50),
Description varchar(50))

create table Subcategories(
Sid int primary key identity,
Name varchar(50),
Catid int foreign key references Categories(Catid))

create table Customers(
Cid int primary key identity,
Name varchar(50),
Score int,
Type varchar(50))

create table Companies(
CompId int primary key identity,
Ttile varchar(50))
```

```
create table DirectOrders(
CustomerId int foreign key references Customers(Cid),
CompanyId int foreign key references Companies(CompId),
TypeOP VARCHAR(50) CHECK(TypeOP='Online' OR TypeOP='Phone')
primary key(CustomerId, CompanyId))

create table Instruments(
Iid int primary key identity,
SerialNumber varchar(50) UNIQUE,
ManufactureDate date,
Color varchar(30),
Price int,
Sid int foreign key references Subcategories(Sid),
Mid int foreign key references  Manufacturers(Mid))

create table Orders(
Cid int foreign key references Customers(Cid),
Iid int foreign key references Instruments(Iid),
DateOrderMade date,
DateOrderHonored date,
OnlinePhone varchar(20),
NumberOfInstruments int,
primary key(Cid, Iid))
```