Fundamentals of Database Systems

Assignment: 1

Due Date: 8th August, 2017

Instructions

This question paper contains 15 questions in 5 pages.

Q1: The users are allowed to access different parts of data differently. They have editing rights for some data and only viewing rights for some and no right for other data. This is an example of _____ feature in DBMS.

A. data integrity

B. data security

C. data isolation

D. atomic operations

> **Explanation:** Data security property in DBMS gives the power to provide different levels of access controls to different people.

Q2: Simplify the following.

$$r(A, B, C)$$

$$s(A, B, C)$$

$$\sigma_{A>C}((r \cup s) - (r - s) - (s - r))$$

A. $\sigma_{A>C}(r \bowtie s)$

B. $\sigma_{A>C}(r \cup s)$

C. $\sigma(r \cap s)$

D. $\sigma_{A>C}(r \cap s)$

> **Explanation:** $((r \cup s) - (r - s) - (s - r)) = r \cap s$

Q3: Choose the correct option.

$$t1 \leftarrow \sigma_\theta((r \times s) \div r)$$

$$t2 \leftarrow \sigma_\theta(s)$$

$$t3 \leftarrow \sigma_\theta(r)$$

A. $t1 = t3$

B. $t2 = t3$

C. $t1 = t2$

D. None of the above

Explanation: $((r \times s) \div r) = s$

Q4: Consider the following relation, where the primary key is underlined.

Flights(Fno, Src, Dst, Dep, Arr)

The set of attributes is denoted by F.
Write a relational algebra query with **basic relational algebra operators** to enlist all possible connecting flights from one city to another, having one stop and proper flight arrival and departure timing.

A. $\sigma_{f1.dst=f2.src}(\rho_{f1(F)}(Flights) \times \rho_{f2(F)}(Flights))$

B. $\Pi_{f1.dst=f2.src \wedge f2.dep>f1.arr}(\rho_{f1(F)}(Flights) \times \rho_{f2(F)}(Flights))$

C. $\sigma_{f1.dst=f2.src \wedge f2.dep>f1.arr}(\rho_{f1(F)}(Flights) \bowtie \rho_{f2(F)}(Flights))$

D. $\mathbf{\Pi_{f1.Fno,f2.Fno}\sigma_{f1.dst=f2.src \wedge f2.dep>f1.arr}(\rho_{f1(F)}(Flights) \times \rho_{f2(F)}(Flights))}$

Explanation: The table Flights needs to be joined with itself but needs to be renamed for the distinction for specifying join condition and while projecting, the ideal join conditions for feasible flights require the destination of 1st flight and source of 2nd flight to be same and the departure time of 2nd flight to be after the arrival of first.

Q5: What does data integrity mean in the DBMS context?

A. Keeping all data in one place.

B. Making sure that data is not lost by taking regular backups.

C. Adding constraints to ensure semantic data correctness.

D. Being able to easily merge two databases.

Explanation: Data integrity constraints ensure semantic correction.

Q6: Which of the following statements are **not** true?

A. Superset of a superkey is also a superkey.

B. Superset of a candidate key is a superkey.

C. Subset of a candidate key can be a superkey.

D. Proper subset of a candidate key is also a candidate key.

Explanation: A and B are true from definition. For C, Candidate key itself is a subset of a candidate key, thus can be a superkey. D is false follows from definition.

Q7: In a bank, each customer is given a unique identification number, *cid*. All the current account holders have different (Firstname, Lastname) pair. (Some of them may have same first names, or same lastnames.)
As per the RBI regulations, the bank also collects everyone's Adhaar Card number. The company's database stores the following fields: (*cid, firstname, lastname, age, adhaar*)

Which of the rows in the following do **not** contain **wrong** information?

No.	Candidate Keys	Super Keys	Primary Keys
1	(*cid*), (*adhaar*)	(*cid*), (*adhaar, age*) (*cid, adhaar*)	(*cid*), (*adhaar*)
2	(*adhaar*)	(*cid*), (*firstname, lastname*)	(*cid*)
3	(*cid*), (*adhaar*)	(*cid, firstname*), (*adhaar, firstname*)	(*adhaar*)
4	(*firstname, lastname*), (*adhaar*)	(*cid, adhaar*), (*adhaar, firstname, lastname*)	(*cid*)

Note:

- Each key is a tuple, enclosed by round brackets.
- Each cell may contain multiple keys, but does not list all possible values.

A. Row 1 and 2

B. Row 2 and 3

C. Only Row 3

D. Only Row 4

Explanation: Only (*cid*) and (*adhaar*) are candidate keys. Any superset of those are superkeys and any ONE of those can be chosen as primary key.

Q8: If a database has two tables T_1, T_2 and both of these tables have a same column C, then C is a foreign key.

A. True

B. False

Explanation: Just having a same column does not make it a foreign key. It needs to be a primary key for one table, and then it is called a foreign key for the other table.

Q9: An operation on the relation A outputs B such that B contains only selected attributes of A. Operation would be called:

A. Selection

B. Difference

C. Projection

D. Intersection

Explanation: Follows from the definition.

Q10: Following sequence of operators are used on a schema. In which of the following, would the database schema be unchanged?

A. Select, Union, Select, Difference

B. Select, Project, Union, Select

C. Select, Difference, Cartesian Product, Union

D. Project, Difference, Union, Select, Rename, Select

Explanation: Project, Cartesian Product and Rename operators change the schema, others don't.

Q11: Why are *additional operators* used in relational algebra?

I Set Intersection operator increases the expressive power of 6 basic operators.
II Assignment operator increases the epxressive power of 6 basic operators.
III Join operator increases the expressive power of 6 basic operators.
IV Divison operator increases the expressive power of 6 basic operators.
V Addition of these operators simplifies writing queries.

A. Only I, II and V
B. Only III and IV
C. Only IV
D. Only V

Explanation: These operators do not add any expressive power, only make it easier to write complex queries.

Q12: Given relations r and s. Function C denotes number of results in the output. Which of the following statements **can be** correct?

I $C(r \bowtie s) > C(r \rightarrowtriangle s)$
II $C(r \rightarrowtriangle s) > C(r \rightarrowtriangle\!\!\!\!\sqsubset s)$
III $C(r \rightarrowtriangle\!\!\!\!\sqsubset s) > C(r \bowtie\!\!\!\!\sqsubset s)$
IV $C(r \rightarrowtriangle s) > C(r \bowtie\!\!\!\!\sqsubset s)$

A. I and II
B. II and III
C. Only III
D. III and IV

Explanation: Left and right outer joins can yield result set of sizes both greater or smaller than each other depending on the data. Full outerjoin will always have a set size greater than (or equal to) the result of left or right joins.

Q13: Consider the following schema for an office payroll system, where primary keys are underlined and foreign keys are italicized.

Person(pid, fname, lname)
Employee(*pid*, desig, salary)

Write a query to display full name and designation of employees who have salary greater than average salary given for their designation.

A. $\Pi_{fname,\ desig}\sigma_{salary>avg(salary)}(Person * ({}_{desig}\mathcal{G}_{avg(salary)}(Employee)))$
B. $\mathbf{\Pi_{fname,\ lname,\ desig}\sigma_{salary>avg(salary)}(Person * (Employee * ({}_{desig}\mathcal{G}_{avg(salary)}(Employee))))}$
C. $\Pi_{fname,\ desig}\sigma_{salary>avg(salary),\ pid,\ lname}(Person * (Employee * ({}_{desig}\mathcal{G}_{avg(salary)}(Employee))))$

D. $\Pi_{fname,\ lname,\ desig}\sigma_{salary>avg(salary)}(Person * ({}_{desig}\mathcal{G}_{avg(salary)}(Employee)))$

Explanation: Options A and D have wrong joins as they lack Employee table. Option C lacks full name. Option B joins *Employee* table with the groups of average salary on designation, joins it with *Person* table to fetch full name, selects rows with salary greater than average salary and projects appropriate attributes.

Q14: For the above payroll system, now assume that the *Employee* table has the following schema.

Employee(*pid*, desig, cur_sal, init_sal)

cur_sal and *init_sal* respectively represent current salary and initial salary of an employee. Write a relational algebra query to calculate average rise in salary for each designation.

A. $\sigma_{desig,\ avg(curr\_sal-init\_sal)}(Employee)$

B. $\sigma_{desig}\mathcal{G}_{curr\_sal-init\_sal}(Employee)$

C. $\Pi_{desig,\ avg(curr\_sal-init\_sal)}(Employee)$

D. $\boldsymbol{_{desig}\mathcal{G}_{avg(curr\_sal-init\_sal)}(Employee)}$

Explanation: Aggregate operation *avg*() is required to be applied on generalised projection and on tuples grouped by designation. Hence, options A and C are wrong, Option B is wrong because σ is neither applied correctly nor required.

Q15: Consider the following relation schema and a query that uses additional operators of relational algebra.

$r(A, B, C)$; $s(D, A, E)$; $t(A, B, C, D, E)$; $u(C, D, E)$

$$((r * s) \cap t) \div u$$

What can be said about the result set if we write this query using only the 6 basic operators of relational algebra. Select all correct options.

A. Result set of the basic-operator-query will be larger than result set of given query.

B. Result set will have attributes A and B only.

C. Some of the operations in query cannot be performed due to incompatible relation schemas.

D. Query cannot be written using only basic operators.

Explanation: Additional operators do not add any power over basic relational algebra, thus the query **can** be written using basic operators, however the result set will be exactly the same since it would be the exact same query. It is evident from definition of the involved operators that all relevant schemas are compatible. Schema of natural join will be (A, B, C, D, E) since the join will happen on A, and division with (C, D, E) will result in schema (A, B).