Assignment-3 Solutions

Topics

1

If the cardinality of a relation R is a and the cardinality of a relation S is b, what is the cardinality of the Cartesian Product AXB?

✓ ab

- The Cartesian product operation of RR and SS, denoted by RXSRXS, involves combining every tuple in RR with every tuple in SS. By the multiplication principle of counting, we get the cardinality of the cartesian product as abab

✗ a+b

- Please rethink. Cartesian product does not just involve concatenation.

✗ ab

- Please rethink. Use the multiplication principle of counting.

2

Consider two sets A and B. In the lectures, we covered the basic relational operations such as union and intersection. Which of the following is/are true in general?

✓ If A⊂B, then A∪B=B and A∩B=A

- If A⊂B, then B already includes all elements from A. B also may have some extra elements A may not have. The union operation is selecting elements from A and then selecting elements from B. Since A is in B, it can be reduced to picking elements from B alone. The intersection operation is selecting elements which belong to both A and B. Since A is in B, it can be reduced to picking elements from A alone.

✓ $A \subset (A \cup B)$

- If $A \subset B$, then B already includes all elements from A. B also may have some extra elements A may not have. Union in this case, is just set B. Since $A \subset B$, it implies this option.

✓ $(A \cap B) \subset B$

- If $A \subset B$, then B already includes all elements from A. B also may have some extra elements that A may not have. Intersection in this case, is just set A. Since $A \subset B$, it implies this option.

3

Which of the following relational algebra operators can lead to a decrease in the cardinality of a relation?

✓ Projection (π)

- Projection can reduce the rows in case there are duplicate elements.

✗ Union ($\cup$)

- Union is a binary operation that involves selecting tuples from both relations. Even in the limiting case when one of the relations is a null relation, union would give a resultant relation that is having the same cardinality.

✓ Selection (σ)

- Selection is a unary operation that involves selecting rows from a table that satisfy a certain predicate. Thus, projection involves reducing rows.

4

Given two relations, R(A,B) and S(B,C,D), which of the following relational algebra expressions is equivalent to $\sigma_{A<C} (\sigma_{A>3} R \bowtie \sigma_{C<9} S)$?

✓ $\pi A,R.B,C,D\ \sigma R.B=S.B \text{ AND } A<C\ (\sigma A>3\ R \times \sigma C<9\ S)$

- A natural join is equivalent to a Cartesian product, followed by a selection to filter out tuples with non-matching values on common attributes, and then followed by a projection to merge common attributes.

✗ $R \bowtie A>3 \text{ AND } C<9 \text{ AND } A<C\ S$

- Theta condition missing R.B = S.B, and the final projection is missing as well

✗ $\pi A,R.B,C,D\ (R \bowtie A>3 \text{ AND } C<9 \text{ AND } A<C\ S)$

- Theta condition missing R.B = S.B

5

Let R and S be two relations with the schema R(P_,Q_,A,B,C), S(P_,Q_,D,E) where P and Q are the key of relation R and S respectively. Which of the I, II, III, IV expressions are equivalent?

✗ Only I and II

- In I, we first perform the natural join of R and S which involves selecting those tuples that have the same attributes (P and Q) in R and S. Then a selection of unique P values is performed. In II, we don’t look at attribute Q and only perform a projection on P attribute values. So, II and I are not identical

✗ Only I and III

- Correct, I and III are identical. Please rethink if another expression among the choices is equivalent to these as well.

✗ Only I, II and III

- In I, we first perform the natural join of R and S which involves selecting those tuples that have the same attributes (P and Q) in R and S. Then a selection of unique P values is performed. In II, we don’t look at attribute Q and only perform a projection on P attribute values. Thus we miss the R attribute in this case. So, II and I are not identical.

✓ Only I, III and IV

- In I, we first perform the natural join of R and S which involves selecting those tuples where they have the same attributes (P and Q) in R and S. Then a selection of unique P values are performed. In II, we don't look at attribute Q and only perform a projection on P attribute values. In III, the term in the parenthesis is equivalent to the natural join and thus I and II are identical. In IV, we use the set theory formula $A \cap BA \cap B = (A-(A-B))(A-(A-B))$, and thus it is equivalent to III

6

Given the two relations R and SS as shown in the figure, which tuple is in the result of R ⋈θ S, where θ is the condition R.B=S.B AND A<C? Assume the resultant relation has attributes in the order of A, R.B, S.B, C, D.

✓ (7, "b", "b", 8, 6)

- This is the only tuple in the join result

✕ (4, "c", "c", 3, 4)

- This violates A < C.

✕ (4, "c", "c", 8, 6)

- This is not even in the Cartesian product of R and S.

7

In the lecture, Professor Kevin discussed about natural join and theta join in relational algebra. Which of the following is/are true in general?

✓ Natural Join allows only equality conditions on all attributes which have the same names in relations R and S.

- Natural join does not allow conditions other than equality; for this, we have conditional joins.

✓ When applied on two relations, natural join will return tuples having identical attribute values only once.

- Natural joins don't allow for duplication of attribute values as it outputs.

✓ Theta join is the same as performing the corresponding selection over the Cartesian product of the two relations.

- Theta join tests for a condition over the attributes such as ≥, ≤ when performing join operation.

8

During a campus survey you asked a large number of students about their favorite film directors, and then you input the data into a relational database. The schema R(NetID, FavDirector) does not support lists of strings, that is, if a student has multiple favorite directors, that information will take multiple rows in the relation. Now a friend of yours wants you to find out who likes both Sergio Leone (SL) and Quentin Tarantino (QT). How would you formulate this query using a relational algebra expression?

✓ (π_{NetID} $\sigma_{FavDirector=SL}$ R) ∩ (π_{NetID} $\sigma_{FavDirector=QT}$ R)

- We want to find those who like SL, and those who like QT, via two selections on the same relation, and then take the intersection to get the set of those who like both.

✗ (π_{NetID} $\sigma_{FavDirector=SL}$ R) ∪ (π_{NetID} $\sigma_{FavDirector=QT}$ R)

- Union means "or", not "and".

✗ π_{NetID} ($\sigma_{FavDirector=SL}$ R ⋈ $\sigma_{FavDirector=QT}$ R)

- This natural join will result in an empty relation.

✗ π_{NetID} $\sigma_{FavDirector=SL\ AND\ FavDirector=QT}$ R

- FavDirector cannot take two different values at the same time, hence the selection will result in an empty table..

✗ $\pi\text{NetID } \sigma\text{FavDirector=SL OR FavDirector=QT R}$

- This produces a list of students who like SL or QT, but not necessarily both.

9

Recall the relational database "AcademicWorld" from class, in particular, consider the three relations in the database: Student(StudentID, Major, Birthday)Student(StudentID, Major, Birthday), Teaches(ProfessorID, CourseNumber)Teaches(ProfessorID, CourseNumber), Enrolls(StudentID, CourseNumber)Enrolls(StudentID, CourseNumber). Prof. Chang asks you to get the student IDs of all the statistics (stats) majors who are taking his courses. How would you formulate this query in relational algebra?

✓ $\pi\text{StudentID (Enrolls} \bowtie \sigma\text{ProfessorID=kcchang Teaches} \bowtie \sigma\text{Major=stats Student)}$

- The first natural join connects student enrollment to professor teaching via the common attribute CourseNumber; the second natural join enables major filtering on the previous join result.

✗ $\pi\text{StudentID (}\sigma\text{ProfessorID=kcchang Teaches} \bowtie \sigma\text{Major=stats Student)}$

- Missing the relationship between students and courses.

✗ $\pi\text{StudentID } \sigma\text{Major=stats (Enrolls} \bowtie \sigma\text{ProfessorID=kcchang Teaches)}$

- The one join result doesn’t have the Major attribute.

10

One popular application of the MapReduce framework is social network analysis. Let (u,v) denote that user u follows user v, and we want to, given a large set of such)(u,v) pairs as input, compute the number of followers for every user in the network. Try to write the map function and the reduce function to compute this knowledge. What is the input of your (working) reduce function?

✓ (v, [u1…un]), where v is followed by all ui in the array.

- The Map function should reverse the input key-value pair and output (followee, follower) pairs. The shuffling process will group these pairs according to the followee key, before assigning each such group to one reduce task, which then computes the length of the follower array as the desired output.

✗ (v,u) where v is followed by u.

✗ (u, [v1…vn]), where u follows all vi in the array.
