CSC540 Homework 4 Solution

Question 1

Suppose relation R(X, Y, Z) has the tuples:

X	Y	Z
2	6	5
5	6	1
9	0	1
9	0	7
5	6	5

Compute the bag union of the following four expressions, each of which is the bag projection (PI) of a grouping (GAMMA) operation using renaming (RHO):

1. PI_A(RHO_R{X, Y, A}(GAMMA_{X, Y, AVG(Z)}(R)))
2. PI_A(RHO_R{Y, A}(GAMMA_{Y, SUM(Z)}(R)))
3. PI_A(RHO_R{X, A}(GAMMA_{X, MIN(Z)}(R)))
4. PI_A(RHO_R{Y, A}(GAMMA_{Y, MAX(X)}(R)))

Demonstrate that you have computed this bag correctly by identifying, from the list below, the correct count of occurrences for one of the elements.

Question Explanation

Let us consider PI_A(RHO_R{X, Y, A}(GAMMA_{X, Y, AVG(Z)}(R))) first. The GAMMA operation groups by X and Y; that is, tuples are in the same group if and only if they agree on both X and Y. Thus, (5, 6, 1) and (5, 6, 5) are in one group, (9, 0, 1) and (9, 0, 7) are in another group, but (2, 6, 5) is in a group by itself. We take the AVG of Z for each group to get the value of A. Their AVG's are, respectively, 3, 4, and 5. Thus, we have {{3, 4, 5}} so far for our bag result.
Now, consider PI_A(RHO_R{Y, A}(GAMMA_{Y, SUM(Z)}(R))). Grouping by only Y happens to place the tuples of R in two groups {(2, 6, 5), (5, 6, 1), (5, 6, 5)} and {(9, 0, 1), (9, 0, 7)}. We need to take the sums of the Z values for each groups, which are, respectively 11 and 8. These are also values of A for the bag being assembled, which now has {{3, 4, 5, 11, 8}}.
Next, consider PI_A(RHO_R{Y, A}(GAMMA_{Y, MAX(X)}(R))) , We need to take the max of X, which are 5 and 9 respectively. As for PI_A(RHO_R{X, A}(GAMMA_{X, MIN(Z)}(R))), grouping by X places tuples {(2, 6, 5)}, {(5, 6, 1), (5, 6, 5)}, and {(9, 0, 1), (9, 0, 7)}. To get values of A, we need to take the MIN's of Z, which are 5, 1, and 1.
Thus, the complete bag for the union of these four expressions is {{3, 4, 5, 11, 8, 5, 9, 5, 1, 1}}. There is two 1's, one 3, one 4, three 5's, one 8, one 9 and one 11.

Correct Answer

5 appears exactly three times.
8 appears exactly once.
1 appears exactly two times.
11 appears exactly once.

Incorrect Answer

1 appears exactly three times.
2 appears exactly once.
3 appears exactly three times.
4 appears exactly two times.
5 appears exactly two times.
7 appears exactly once.

8 appears exactly three times.
10 appears exactly once.

Question 2

Suppose relation R(A, B, C, D) has the tuples:

A	B	C	D
2	2	3	3
3	2	4	4
3	3	5	2
3	3	2	5
4	4	3	3
4	4	4	4
5	3	5	2
3	2	2	5
5	2	3	3
5	2	4	4

Using bag projection and difference, compute $\pi_{A,B}(R) - \rho_{S(A,B)}(\pi_{C,D}(R))$. Note that the remaining is only to give the two projections the same schema. Which of the following is true about the tuples that appear in the result?

Question Explanation

The bag projection of R onto attributes A and B is {{(2, 2), (3, 2), (3, 3), (3, 3), (4, 4), (4, 4), (5, 3), (3, 2), (5, 2), (5, 2)}} and the bag projection of R onto C and D is {{(3, 3), (4, 4), (5, 2), (2, 5), (3, 3), (4, 4), (5, 2), (2, 5), (3, 3), (4, 4)}}. The renaming gives both relations the schema (A,B), so we can take the difference legally. In the bag difference, a tuple appears a number of times equal to the difference of the number of times it appears in the first argument minus the number of times it appears in the second. However, that difference can never be less than 0; that is, a tuple that appears at least as many times in the second argument as it does in the first does not appear in the result at all. (3, 3) appears twice in the first projection and three times in the second. Since 3 > 2, (3, 3)appears zero times in the result. (4, 4) also appears zero times in the result. Likewise, (5, 2),which appears twice in both arguments does not appear in the result. (2, 2), (3, 2), (3, 2), and (5, 3) will exist in the result. So, the bag difference is thus: {{(2, 2), (3, 2), (3, 2), (5, 3)}}.

Correct Answer

(2, 2) appears once in the result.
(3, 2) appears twice in the result.
(5, 3) appears once in the result.
(3, 3) does not appear in the result.

Incorrect Answer

(2, 2) does not appear in the result.
(3, 2) appears once in the result.
(4, 4) appears -1 times in the result.
(5, 3) appears twice in the result.
(5, 2) appears once in the result.
(3, 3) appears once in the result.

(5, 2) appears twice in the result.

Question 3

Consider the relational database shown below:

student(student-name, street, city)
study(student-name, university-name, SAT)
university(university-name, city)
tutor(tutor-name, person-name)

Identify the correct relational algebra expression for the queries shown below. Assume the following notations:

Π - Projection

∞ - Natural Join

σ - Selection

× - Products

Question Explanation

1. First, we find all the students who study at NC State University; we do it by using the study relation. The information about a student's city is stored in the student relation. Hence a join of the first result with the student relation and then a projection on the student-name and city gives us the desired result.
2. First, we find all the students whose SAT scores are less than or equal to those for every student in "NC State University". The result of this step includes even the students studying at "NC State University". In order to find the solution to the problem, we subtract the above result from the study relation. A projection of the student-name yields the desired result.
3. The information about a student's city is stored in the student relation, and the information about a university's city is stored in the university relation. The mapping between the university and the student-name is stored in the study relation. Thus a join of these 3 relations yields the desired result.

Correct Answer

1. Find the names and cities of residence of all students who study at NC State University.

$\Pi_{\text{student-name, city}}(\text{student} \infty (\sigma_{\text{university-name= "NC State University"}} (\text{study})))$

2. Find the names of all students whose SAT score is greater than the SAT score of every student of NC State University.

$\Pi_{\text{student-name}}(\text{study}) - (\Pi_{\text{study.student-name}}(\text{study} \infty_{\text{study.SAT <= study2.SAT and study2.university-name = "NC State University"}} \rho_{\text{study2}}(\text{study})))$

3. Find the names of all students in this database who live in the same city as the university for which they study.

$\Pi_{\text{student-name}}$ (student ∞ study ∞ university)

Incorrect Answer

1. Find the names and cities of residence of all students who study at NC State University.

$\Pi_{\text{student-name, city}}(\text{student} \infty (\sigma_{\text{university-name = "NC State University"}} (\text{university})))$

2. Find the names and cities of residence of all employees who work for First Bank Corporation.

$\Pi_{\text{student-name, city}}(\text{student} \times (\sigma_{\text{university-name= "NC State University"}}(\text{study})))$

3. Find the names of all students whose SAT score is greater than the SAT score for every student of NC State University.

$\Pi_{\text{student-name}}(\text{study}) - (\sigma_{\text{study.student-name}}(\text{study} \infty_{\text{study.SAT <=study2.SAT and study2.university-name = "NC State University"}} \rho_{\text{study2}}(\text{study})))$

4. Find the names of all students whose SAT score is greater than the SAT score for every student of NC State University.

$\Pi_{\text{student-name}}(\text{study}) - (\sigma_{\text{study.student-name}}(\text{study} \infty_{\text{study.SAT <=study2.SAT}}) \rho_{\text{study2}}(\text{study}))$

5. Find the names of all students in this database who live in the same city as the university for which they study.

$\Pi_{\text{student-name}}(\text{student} \times (\text{study} \infty \text{university}))$

6. Find the names of all students in this database who live in the same city as the university for which they study.

$\sigma_{\text{student-name}}(\text{student} \infty (\text{study} \times \text{university}))$

Question 4

Here are three relations, R1(m, n), R2(m, n), and R3(m, n). Their current values are:

R1

m	n
a	a
a	b
b	a
b	b

R2

m	n
a	a
a	b
b	a
b	b

R3

m	n
a	a
a	b
b	a
b	b

Compute the result of the following query:

```
SELECT R1.m, R1.n, R2.n, R3.n
FROM R1, R2, R3
WHERE R1.n = R2.m AND R2.n <> R3.n AND R3.m <> b;
```

Identify in the list below the true statement about whether or not a tuple appears in the output and how many times it appears in the output.

Question Explanation

If you look at the data carefully, there are only two values (a and b) in the three tables. So, you can consider the values as binary values (0 as a and 1 as b). The product of R1, R2, and R3 contains 64 tuples --- all sequences of six a's and b's. The WHERE condition R1.n = R2.m forces the second and third components to be the same, thus eliminating half the tuples and leaving 32. The condition R2.n <> R3.n eliminates half the remaining --- those that do not differ in the 4th and 6th components. By the above process, we get 16 tuples of the form (p, q, q, r, s, r'), where r' is the complement of r, that is, one of r and r' is a and the other is b.

The last condition, R3.m <> b, enforces the s component to be a. Therefore, the result will be 8 tuples of the form (p, q, q, r, 'a', r').

The SELECT clause produces from each of these 8 tuples a row of the result: (p, q, r, r').

Correct Answer

(a, a, a, b) appears once.
(b, a, a, b) appears once.
(a, b, a, b) appears once.
(a, a, a, a) does not appear.
(b, b, b, b) does not appear.
(b, b, a, a) does not appear.

Incorrect Answer

(a, a, a, b) appears twice.
(b, b, b, a) appears twice.
(a, b, a, b) appears twice.
(a, a, b, b) appears once.
(a, b, b, b) appears once.
(a, b, b, a) does not appear.
(b, b, a, b) does not appear.

Question 5

Suppose relation R(a, b, c, d, e) currently has the tuples:

R

a	b	c	d	e
1	4	3	7	3
2	1	4	3	3
5	3	1	2	2
3	8	5	1	7

Which of the following tuples is in the generalized projection PROJ_{b, d - a, 3 * e}(R)?

Question explanation
None

Correct Answers

(4, 6, 9)
(1, 1, 9)
(3, -3, 6)
(8, -2, 21)

Incorrect Answers

(4, 6, 6)
(4, 2, 9)
(1, 2, 3)
(2, 1, 4, 3, 3)

(3, 3, 4)
(8, -4, 7)

Question 6

Suppose relation R1(L, M) has the tuples:

R1

L	M
7	f
3	d
4	e
6	d
1	a
9	b
3	j

and suppose relation R2(M, N) has the tuples:

R2

M	N
c	3
d	2
b	6
i	5
e	3

Identify which of the following (L, M, N) tuples can result from the left natural outer-join of R1 and R2.

Question Explanation

After the left natural outer-join, we can see the following result:

L	M	N
7	f	null
3	d	2
4	e	3
6	d	2
1	a	null
9	b	6
3	j	null

Correct Answers

(7, f, null)
(3, d, 2)
(4, e, 3)
(6, d, 2)
(1, a, null)
(9, b, 6)
(3, j, null)

Incorrect Answers

(null, null, null)
(7, f, 2)
(9, null, 6)
(null, b, 6)
(6, d, 3)
(4, e, null)
(3, null, null)
(1, a, 6)
(null, null, 3)

Question 7

Suppose relation R(L, M, N) has the tuples:

L	M	N
1	1	2
2	1	1
2	3	2
1	1	1
3	2	1
1	1	3

Using bag projection and intersection, compute $\Pi_{(L,M)}(R) \cap \rho_{S(L,M)}(\Pi_{(M,N)}(R))$.
Note that the remaining is only to give the two projections the same schema. Which of the following is true about the tuples that appear in the result?

Question Explanation

The bag projection of R onto attributes L and M is {{(1, 1), (2, 1), (2, 3), (1, 1), (3, 2), (1, 1)}} and the bag projection of R onto M and N is {{(1, 2), (1, 1), (3, 2), (1, 1), (2, 1), (1, 3)}}. The renaming gives both relations the schema (L, M), so we can take the intersection legally. In the bag intersection, a tuple appears the minimum of the number of times it appears in the two relations whose intersection is taken. Thus, (1, 1) appears twice since it appears three times in the first projection and twice in the second. Both (2, 1) and (3, 2) appear once since they appear only once in the first projection and in the second. (2, 3), (1, 3), and (1, 2) do not appear since each of them appears in one of the two projections but not both.

Correct Answers

(1, 1) appears twice in the result.
(2, 1) appears once in the result.
(3, 2) appears once in the result.
(2, 3) does not appear in the result.

(1, 3) does not appear in the result.
(1, 2) does not appear in the result.

Incorrect Answers

(1, 1) appears three times in the result.
(2, 1) appears twice in the result.
(3, 2) appears twice in the result.
(2, 3) appears once in the result.
(1, 3) appears once in the result.
(1, 2) appears once in the result.
(1, 1) appears five times in the result.
(3, 2) does not appear in the result.