Cartesian Product X

A	B
1	2
3	4

(a) Relation R

B	C	D
2	5	6
4	7	8
9	10	11

(b) Relation S

A	R.B	S.B	C	D
1	2	2	5	6
1	2	4	7	8
1	2	9	10	11
3	4	2	5	6
3	4	4	7	8
3	4	9	10	11

(c) Result $R \times S$

Question 1: Cartesian Product

Given the following relations R and S, compute the Cartesian product $\mathrm{R} \times \mathrm{S}$.

Relation R

A B
1 2
3 4

Relation S

B C D
5 6 7
8 9 10

Result $\mathrm{R} \times \mathrm{S}$

Compute the Cartesian product and display the resulting relation.

Question 2: Cartesian Product with Different Attributes

Given the following relations R and S, compute the Cartesian product $\mathrm{R} \times \mathrm{S}$.

Relation R

X Y
1 2
3 4

Relation S

YZW
2 5 6
4 7 8

Result $R \times S$

Compute the Cartesian product and display the resulting relation.

Question 3: Cartesian Product with Larger Relations

Given the following relations R and S, compute the Cartesian product $R \times S$.

Relation R

AB
1 2
3 4
5 6

Relation S

BCD
2 5 6
4 7 8
6 9 10

Result $R \times S$

Compute the Cartesian product and display the resulting relation.

Explanation of Concepts in the Questions

1. **Cartesian Product (×)**: Combines each tuple from the first relation with each tuple from the second relation, resulting in a new relation with all possible combinations.
2. **Attributes**: The columns in relations. The resulting relation from a Cartesian product will have attributes from both input relations.
3. **Tuples**: The rows in relations. The number of tuples in the resulting relation is the product of the number of tuples in the input relations.

Question 1: Natural Join

Given the following relations $\mathrm{R}R$ and $\mathrm{S}S$, compute the natural join $\mathrm{R} \bowtie \mathrm{S} R \bowtie S$.

Relation $\mathrm{R}R$

A	B
1	2
3	4

Relation $\mathrm{S}S$

B	C	D
2	5	6
4	7	8
9	10	11

Result $\mathrm{R} \bowtie \mathrm{S}$

Compute the natural join and display the resulting relation. Explain why the third tuple of $\mathrm{S}S$ does not appear in the result.

Question 2: Natural Join with Different Attributes

Given the following relations $\mathrm{R}R$ and $\mathrm{S}S$, compute the natural join $\mathrm{R} \bowtie \mathrm{S} R \bowtie S$.

Relation R

X	Y
1	2
3	4

Relation S

Y	Z	W
2	5	6
4	7	8
6	9	10

Result $\mathrm{R} \bowtie \mathrm{S}$

Compute the natural join and display the resulting relation. Explain why the third tuple of $\mathrm{S}S$ does not appear in the result.

Question 3: Natural Join with Larger Relations

Given the following relations R and S, compute the natural join $\mathrm{R} \bowtie \mathrm{S}$.

Relation $\mathbf{R}$

A	B
1	2
3	4
5	6

Relation $\mathbf{S}$

B	C	D
2	5	6
4	7	8
6	9	10

Result $\mathbf{R} \bowtie \mathbf{S}$

Compute the natural join and display the resulting relation. Explain why the third tuple of $\mathrm{S}S$ does not appear in the result.

Explanation of Concepts in the Questions

1. **Natural Join ($\bowtie\bowtie$)**: Combines tuples from two relations based on common attributes, and the resulting relation includes only the tuples that have matching values in these common attributes.
2. **Resultant Schema**: The schema of the resulting relation is the union of the schemas of the input relations, with common attributes appearing only once.
3. **Tuple Matching**: Only tuples with matching values in the common attributes are included in the result. Tuples without matches are excluded.

Joins $\bowtie$

A	B	C
1	2	3
6	7	8
9	7	8

(a) Relation U

B	C	D
2	3	4
2	3	5
7	8	10

(b) Relation V

A	B	C	D
1	2	3	4
1	2	3	5
6	7	8	10
9	7	8	10

(c) Result $U \bowtie V$

First tuple of U has two matching tuples in V.

Third tuple of B has two matching tuples in U.

Examples

Is the following statement correct:

a) R∩S = R - (R-S)

---

1. Is the following statement correct?

   a) (R ∪ S) - S = R - (S - R)

2. Determine whether the following equation holds:

   a) (R - S) ∩ (S - R) = Ø

3. Verify if the following identity is true:

   a) (R ∩ S) ∪ (R - S) = R

Let's prove each of the statements with set theory principles and provide examples.

Statement 1: (R ∪ S) - S = R - (S - R)

Proof:

1. The left-hand side:
   - $(R \cup S) - S$ represents all elements in $R \cup S$ but not in S. Since S is subtracted, only elements from R remain.
   - So, $(R \cup S) - S = R - S$.
2. The right-hand side:
   - $S - R$ consists of elements that are in S but not in R.
   - $R - (S - R)$ means taking all elements of R and removing any that are in $S - R$.
   - But $S - R$ has no elements from R, so subtracting it does not change R, meaning $R - (S - R) = R - S$.

Since both sides simplify to $R - S$, the statement is **true**.

Example:

Let $R = \{1, 2, 3\}$ and $S = \{3, 4, 5\}$:

- $R \cup S = \{1, 2, 3, 4, 5\}$
- $(R \cup S) - S = \{1, 2\}$
- $S - R = \{4, 5\}$
- $R - (S - R) = \{1, 2\}$

Since both sides are equal, the statement is **true**.

Statement 2: (R - S) ∩ (S - R) = ∅

Proof:

1. $R - S$ consists of elements in R that are not in S.
2. $S - R$ consists of elements in S that are not in R.
3. The intersection of these two sets would contain elements that belong to both $R - S$ and $S - R$.
   - But by definition, $R - S$ contains elements **not in** S, and $S - R$ contains elements **not in** R.
   - There cannot be an element common to both sets.
   - Hence, their intersection must be **empty**.

Thus, the statement is **true**.

Example:

Let $R = \{1, 2, 3\}$ and $S = \{3, 4, 5\}$:

- $R - S = \{1, 2\}$
- $S - R = \{4, 5\}$
- $(R - S) \cap (S - R) = \emptyset$

Since there are no common elements, the statement is **true**.

Statement 3: (R ∩ S) ∪ (R - S) = R

Proof:

1. $R \cap S$ consists of elements common to both R and S.
2. $R - S$ consists of elements in R but not in S.
3. The union $(R \cap S) \cup (R - S)$ contains all elements that are either:
   - In both R and S (from $R \cap S$), or
   - In R but not in S (from $R - S$).
   - Since these two sets together cover all elements in R, the union must be **equal to** R.

Thus, the statement is **true**.

Example:

Let $R = \{1, 2, 3, 4\}$ and $S = \{3, 4, 5\}$:

- $R \cap S = \{3, 4\}$
- $R - S = \{1, 2\}$
- $(R \cap S) \cup (R - S) = \{1, 2, 3, 4\} = R$

Since both sides are equal, the statement is **true**.

Exercise 1

R(A, B) and *S*(B, C). Which two are equivalent. Give an example of how the different one can be different.

a. $\pi_{A,C}(R \bowtie \sigma_{B=1} S)$

b. $\pi_A(\sigma_{B=1} R) \times \pi_C(\sigma_{B=1} S)$

c. $\pi_{A,C}(\pi_A R \times \sigma_{B=1} S)$

Similar Questions:

1. **Given relations R(A, B) and S(B, C), determine which of the following are equivalent:** a. πA,C(σB=2(R ⋈ S))
   b. πA(σB=2(R)) × πC(σB=2(S))
   c. πA,C(πA R ⋈ σB=2(S))
   Provide an example where one differs from the others.
2. **Consider the relations R(A, B) and S(B, C). Which of the following expressions are equivalent?** a. πA,B(R ⋈ σB=10(S))
   b. πA(σB=10(R)) ⋈ πB(σB=10(S))
   c. πA,B(πB(R) ⋈ σB=10(S))
   Justify your answer with an example.
3. **For relations R(A, B) and S(B, C), determine if the following equivalences hold:** a. πA,C(σB=5(R ⋈ S))
   b. πA,C(σB=5(R)) × πC(σB=5(S))
   c. πA,C(R ⋈ πB,C(σB=5(S)))
   Prove or provide a counterexample.

Questions with Proof and Examples:

1. **Prove or disprove the equivalence of the following expressions:**
    - a. $\pi_{A,C}(\sigma_{B=4}(R \bowtie S))$
    - b. $\pi_A(\sigma_{B=4}(R)) \bowtie \pi_C(\sigma_{B=4}(S))$
    - c. $\pi_{A,C}(R \bowtie \pi_{B,C}(\sigma_{B=4}(S)))$

      Proof:
        - The projection operation $\pi_{A,C}$ removes B, so (a) and (c) are equivalent. However, (b) decomposes the selection before joining, which may cause loss of information if B is required.
        - **Example:** If R = {(1,4), (2,4), (3,5)} and S = {(4,10), (5,20)}, then (a) and (c) both return {(1,10), (2,10)} while (b) may lose the relation context.
2. **Show whether the following expressions are equivalent:**
    - a. $\pi_{A,C}(\sigma_{B=3}(R \bowtie S))$
    - b. $\pi_A(\sigma_{B=3}(R)) \bowtie \pi_C(\sigma_{B=3}(S))$
    - c. $\pi_{A,C}(R \bowtie \pi_{B,C}(\sigma_{B=3}(S)))$

      Proof:
        - If B is a join key, removing it before the join (as in (b)) may lead to incorrect results.
        - **Example:** If R = {(1,3), (2,3), (3,4)} and S = {(3,7), (4,8)}, (a) and (c) return {(1,7), (2,7)} but (b) fails.

---

3. **Consider relations R(A, B) and S(B, C). Are the following transformations valid?**
    - a. $\pi_{A,C}(\sigma_{B=6}(R \bowtie S))$
    - b. $\pi_A(\sigma_{B=6}(R)) \times \pi_C(\sigma_{B=6}(S))$
    - c. $\pi_{A,C}(R \bowtie \pi_{B,C}(\sigma_{B=6}(S)))$

      Proof:
        - (b) represents a Cartesian product instead of a join, potentially introducing spurious tuples.
        - **Example:** If R = {(1,6), (2,6), (3,7)} and S = {(6,9), (7,11)}, (a) and (c) return {(1,9), (2,9)}, while (b) might incorrectly generate extra tuples.

Exercise 2

Consider a relation *R*(A, B) that contains r tuples, and a relation *S*(B, C) that contains s tuples; and r > 0 and s > 0.

In terms of R and S the minimum and maximum number of tuples that could be in the result of each expression:

a. $\pi_{A,C}(R \bowtie S)$

b. $\pi_B R - (\pi_B R - \pi_B S)$

c. $(R \bowtie R) \bowtie R$

d. $\sigma_{A>B} R \cup \sigma_{A<B} R$

Similar Questions:

1. **Consider a relation R(A, B) with r tuples and a relation S(B, C) with s tuples. If both relations are stored using a hash-based indexing scheme, determine the minimum and maximum number of tuples that could be produced by the following queries:**

   a. π_A,C(R ⋈ S)

   b. π_B R − (π_B R ∩ π_B S)

   c. (R ⋈ S) ⋈ T, where T(C, D) has t tuples.

   d. σ_A>5 R ∪ σ_B<10 S

2. **Given two relations R(A, B) and S(B, C) with r and s tuples respectively, compute the minimum and maximum possible number of tuples in the output for:**

   a. π_A,B(R ⋈ S)

   b. π_C (σ_B>10(R ⋈ S))

   c. (R ⋈ S) − (σ_B=5 R)

   d. σ_A=1 R ∪ σ_C=2 S

3. **Let R(A, B) and S(B, C) be two relations with r and s tuples respectively. Assuming that attribute B is a foreign key in R referencing S, determine the result sizes for:**

   a. R ⋈ S

   b. π_A,C (σ_B=1(R ⋈ S))

   c. (R × S) − (R ⋈ S)

   d. (σ_A>100 R) ∩ (σ_B<50 S)

Questions with Proof and Examples:

1. **Prove or disprove:** $\pi\_A,C (R \bowtie S) = \pi\_A,C (\pi\_A R \bowtie S)$
   - **Proof:** The projection before the join may remove attributes necessary for the join condition, leading to different results. Example:
     - **R(A, B)**: {(1,2), (3,4)}
     - **S(B, C)**: {(2,5), (4,6)}
     - Applying π_A before join removes B, making the join impossible.
2. **Prove:** $\pi\_A (\sigma\_B=10 R) \times \pi\_C (\sigma\_B=10 S) = \pi\_A,C (\sigma\_B=10(R \bowtie S))$
   - **Proof:** Since selection on B ensures that only tuples with B=10 remain, performing a join and then projecting A and C is equivalent to performing projections on each relation separately and then taking the Cartesian product.
3. **Prove or give a counterexample:** $(R \bowtie S) \bowtie T = R \bowtie (S \bowtie T)$
   - **Proof:** This holds due to the associative property of joins.
   - **Example:**
     - **R(A, B)**: {(1,2), (3,4)}
     - **S(B, C)**: {(2,5), (4,6)}
     - **T(C, D)**: {(5,7), (6,8)}
     - $R \bowtie S$ gives {(1,2,5), (3,4,6)}, then joining with T results in {(1,2,5,7), (3,4,6,8)}.
     - $S \bowtie T$ first gives {(2,5,7), (4,6,8)}, then joining with R produces the same result.

Exercise 3

Calculate the number of Tuples for 1 and 2.
$T(R)=100$, $V(R,A) =5$

1. $\sigma_{A=10}(R)$

2. $\sigma_{A\leq 10}(R)$

Similar Questions:

Question 1:

Calculate the number of tuples for the given conditions:

T(R) = 200, V(R, B) = 10

1. **σ_B=5(R)**
2. **σ_B>5(R)**

Question 2:

Given T(R) = 500 and V(R, C) = 20, determine the number of tuples for:

1. **σ_C=15(R)**
2. **σ_C≤15(R)**

Question 3:

For the relation R(A, D) with T(R) = 1000 and V(R, A) = 25, calculate the tuples in the result of:

1. **σ_A=7(R)**
2. **σ_A≥7(R)**

Three More with Proof and Examples:

Question 4:

Prove that for a relation R(A, B) with total tuples **T(R)** and distinct values **V(R, A)**:

$$\sigma_{A=x}(R) = \frac{T(R)}{V(R, A)}$$

Example:

If T(R) = 300 and V(R, A) = 10, then:

$$\sigma_{A=5}(R) = \frac{300}{10} = 30$$

Question 5:

Prove that for a range selection where A ≤ x, the expected number of tuples is:

$$\sigma_{A\leq x}(R) = \frac{x}{V(R, A)} \times T(R)$$

Example:

If T(R) = 500, V(R, A) = 50, and we want A ≤ 10:

$$\sigma_{A\leq 10}(R) = \frac{10}{50} \times 500 = 100$$

Question 6:

Show that if attribute A follows a uniform distribution in relation R(A, B), then:

$$\sigma_{A>x}(R) = \left(1 - \frac{x}{V(R,A)}\right) \times T(R)$$

Example:

Given T(R) = 400, V(R, A) = 40, find tuples for A > 30:

$$\sigma_{A>30}(R) = \left(1 - \frac{30}{40}\right) \times 400 = (1 - 0.75) \times 400 = 100$$

Exercise 4

Calculate the number of Tuples for 1.
T(R)=100. Values of A are uniformly distributed over [1,20]

1. $\sigma_{A=10}(R)$
2. $\sigma_{A\leq 10}(R)$
3. $\sigma_{A>10}(R)$

Similar Questions:

Exercise 5

Calculate the number of Tuples for 1 and 2.

T(R) = 200, V(R, A) = 10

1. $\sigma_{A=5}(R)$
2. $\sigma_{A\geq 8}(R)$

Exercise 6

Calculate the number of Tuples for 1, 2, and 3.

T(R) = 500. Values of A are uniformly distributed over [1, 50].

1. $\sigma_{A=25}(R)$
2. $\sigma_{A\leq 25}(R)$
3. $\sigma_{A>25}(R)$

Exercise 7

Consider relation R(A, B, C), where T(R) = 150, and attribute A is uniformly distributed over [1, 30].

1. $\sigma_{A=15}(R)$
2. $\sigma_{A\leq 15}(R)$
3. $\sigma_{A>15}(R)$

Examples with Solutions

Example 1

T(R) = 100, A uniformly distributed over [1, 20].

1. $\sigma_{A=10}(R)$
    - Probability of any single value of A: **1/20**
    - Expected tuples: **(1/20) * 100 = 5**
2. $\sigma_{A \leq 10}(R)$
    - Probability: **10/20 = 1/2**
    - Expected tuples: **(1/2) * 100 = 50**
3. $\sigma_{A>10}(R)$
    - Probability: **10/20 = 1/2**
    - Expected tuples: **(1/2) * 100 = 50**

Example 2

T(R) = 200, A uniformly distributed over [1, 50].

1. $\sigma_{A=25}(R)$
    - Probability: **1/50**
    - Expected tuples: **(1/50) * 200 = 4**
2. $\sigma_{A \leq 25}(R)$
    - Probability: **25/50 = 1/2**
    - Expected tuples: **(1/2) * 200 = 100**
3. $\sigma_{A>25}(R)$
    - Probability: **25/50 = 1/2**
    - Expected tuples: **(1/2) * 200 = 100**

Example 3

T(R) = 500, A uniformly distributed over [1, 10].

1. σ_A=5(R)
   - Probability: **1/10**
   - Expected tuples: **(1/10) * 500 = 50**
2. σ_A≤5(R)
   - Probability: **5/10 = 1/2**
   - Expected tuples: **(1/2) * 500 = 250**
3. σ_A>5(R)
   - Probability: **5/10 = 1/2**
   - Expected tuples: **(1/2) * 500 = 250**

Exercise 5

$R(A, B, C)$ is a relation. $T(R) = 10000$. $V(R, A) = 50$.
W = $\sigma_{A=10 \text{ AND } B<10}(R)$. What will be T(W)?

Similar Questions:

Question 3:

Relation **T(M, N, P, Q)** has:

- $T(T) = 15000$
- $V(T, M) = 75$
- $V(T, P) = 50$

Define **Z** as:

$$Z = \sigma_{M=40 \text{ AND } P \geq 10}(T)$$

What is $T(Z)$?

Question 1:

Relation **R(A, B, C)** is given with:

- $T(R) = 5000$ tuples
- $V(R, A) = 25$

Define **W** as:

$$W = \sigma_{A=5 \text{ AND } B>20}(R)$$

What is $T(W)$?

Question 2:

Relation **S(X, Y, Z)** has:

- $T(S) = 20000$
- $V(S, X) = 100$
- $V(S, Y) = 40$

Define **Q** as:

$$Q = \sigma_{X=30 \text{ AND } Y<15}(S)$$

What is $T(Q)$?

Examples with Solutions:

Example 1:

Given relation **R(A, B, C)** with:

- $T(R) = 10000$
- $V(R, A) = 50$
- $V(R, B) = 20$

Find **T(W)** for:

$$W = \sigma_{A=10 \text{ AND } B<10}(R)$$

Solution:

1. The selection $A = 10$ reduces the number of tuples to:

$$\frac{T(R)}{V(R, A)} = \frac{10000}{50} = 200$$

2. The second condition $B < 10$ means that we consider only part of the values of B. Assuming uniform distribution, approximately half of the values of B are less than 10, so:

$$T(W) = \frac{200}{2} = 100$$

Example 2:

Given relation **S(X, Y, Z)** with:

- $T(S) = 8000$
- $V(S, X) = 40$
- $V(S, Y) = 10$

Find **T(Q)** for:

$$Q = \sigma_{X=5 \text{ AND } Y=3}(S)$$

Solution:

1. The selection $X = 5$ reduces the tuples to:

$$\frac{T(S)}{V(S, X)} = \frac{8000}{40} = 200$$

2. The selection $Y = 3$ further reduces it:

$$\frac{200}{V(S, Y)} = \frac{200}{10} = 20$$

Thus, $T(Q) = 20$.

Example 3:

Given relation **T(M, N, P, Q)** with:

- $T(T) = 12000$
- $V(T, M) = 60$
- $V(T, P) = 30$

Find **T(Z)** for:

$$Z = \sigma_{M=12 \text{ AND } P>15}(T)$$

Solution:

1. The selection $M = 12$ reduces the tuples to:

$$\frac{T(T)}{V(T, M)} = \frac{12000}{60} = 200$$

2. The selection $P > 15$ means we consider a portion of the P values. Assuming uniform distribution, half of the P values satisfy the condition:

$$T(Z) = \frac{200}{2} = 100$$

Thus, $T(Z) = 100$.

Preservation of value sets

For R(A,B) and S(B,C)

1. $W(A,B) = \sigma_{A=10}(R)$
   a. $V(W,A) = 1$
   b. $V(W,B) = V(R,B)$
2. $W(A,B,C) = R \bowtie S$
   a. $V(W,A) = V(R,A)$
   b. $V(W,B) = \min\{V(R,B), V(S,B)\}$
   c. $V(W,B) = V(R,B)$

After a selection, the values for attributes other than the one on which selection was performed retain their distinct values.

Similarly after a Join, attributes other than ones on which Join was performed retain their distinct values.

1. Selection: $W(A,B) = \sigma_{A=10}(R)$

- **Selection (σ)** filters tuples based on a condition, but it does not change the values of attributes that are not part of the selection condition.
- The image provides:
  - $V(W,A) = 1$ → Since selection **fixes** the value of A to 10, it has only **one distinct value**.
  - $V(W,B) = V(R,B)$ → The attribute B remains unchanged, meaning it retains all distinct values from R.
- **Example:**
  - Assume **R(A,B)** contains:

```scss
(10, X1), (10, X2), (15, X3), (10, X4), (20, X5)
```

  - If we apply **σ_{A=10}(R)**, only tuples where $A = 10$ remain:

```scss
(10, X1), (10, X2), (10, X4)
```

  - The distinct values of **B** are still **{X1, X2, X4}**, meaning the values for B are **preserved**.

2. Join: $W(A, B, C) = R \bowtie S$

- **Join (⋈)** is performed on the common attribute **B**, which means:
    - $V(W, A) = V(R, A)$ → The attribute **A** retains all distinct values from **R**.
    - $V(W, B) = \min\{V(R, B), V(S, B)\}$ → Since the join matches tuples using B, the number of distinct values for B is **limited by the smaller set** from R and S.
    - $V(W, C) = V(S, C)$ → The attribute **C** retains its distinct values from **S**.
- **Example:**
    - Assume:

```mathematica
R(A,B) = {(1, X), (2, Y), (3, Z)}
S(B,C) = {(X, M), (Y, N), (Y, O), (Z, P)}
```

    - The **join on B** results in:

```scss
(1, X, M), (2, Y, N), (2, Y, O), (3, Z, P)
```

    - $V(W, A) = V(R, A) = \{1, 2, 3\}$ (Preserved)
    - $V(W, B) = \min(V(R, B), V(S, B))$ → **{X, Y, Z}** (since Y appears twice)
    - $V(W, C) = V(S, C) = \{M, N, O, P\}$ (Preserved)

Conclusion

The **Preservation of Value Sets** states that:

- **Selection retains all distinct values** for attributes not used in the condition.
- **Join preserves distinct values** for non-join attributes but **reduces** values of the join attribute to the intersection of both relations.

Exercise 6

For R(A,B), S(B,C) and T(D,E) we have the following specs:

$T(R) = 1000$, $V(R,A) = 1$, $V(R,B) = 30$

$T(S) = 200$, $V(S,B) = 40$, $V(S,C) = 10$

$T(T) = 10000$, $V(T,D) = 10$, $V(T,E) = 20$

Estimate the number of Tuples for the results.

a) $R \bowtie S$

b) $R \bowtie \sigma_{C=10}(S)$

c) $R \bowtie \sigma_{B=5 \text{ AND } C=10}(S)$

d) $\sigma_{A=10}(R) \bowtie S$ Assume result of $\sigma_{A=10}(R)$ is not empty.

e) $\pi_A(R) \bowtie S$

   i) Assuming projection will not eliminate duplicates

   ii) Assuming projection will eliminate duplicates

f) $R \bowtie S \bowtie T$

Exercise 7

For relations R(A,B), S(B,C), and T(C,D), we have the following specifications:

- $T(R) = 800$, $V(R,A) = 2$, $V(R,B) = 40$
- $T(S) = 300$, $V(S,B) = 30$, $V(S,C) = 20$
- $T(T) = 5000$, $V(T,C) = 25$, $V(T,D) = 50$

Estimate the number of tuples for the results:

a) $R \bowtie S$

b) $R \bowtie \sigma_{C=5}(S)$

c) $R \bowtie \sigma_{B=20 \text{ AND } C=5}(S)$

d) $\sigma_{A=2}(R) \bowtie S$ (Assume $\sigma_{A=2}(R)$ is not empty.)

e) $\pi_A(R) \bowtie S$

- i) Assuming projection will not eliminate duplicates
- ii) Assuming projection will eliminate duplicates

  f) $R \bowtie S \bowtie T$

Exercise 8

For relations P(X,Y), Q(Y,Z), and R(Z,W), we have the following specifications:

- T(P) = 600, V(P,X) = 5, V(P,Y) = 50
- T(Q) = 400, V(Q,Y) = 40, V(Q,Z) = 30
- T(R) = 7000, V(R,Z) = 35, V(R,W) = 70

Estimate the number of tuples for the results:

a) $P \bowtie Q$
b) $P \bowtie \sigma_{Z=15}(Q)$
c) $P \bowtie \sigma_{Y=10 \text{ AND } Z=15}(Q)$
d) $\sigma_{X=3}(P) \bowtie Q$ (Assume $\sigma_{X=3}(P)$ is not empty.)
e) $\pi_X(P) \bowtie Q$

- i) Assuming projection will not eliminate duplicates
- ii) Assuming projection will eliminate duplicates
  f) $P \bowtie Q \bowtie R$

Exercise 9

For relations A(M,N), B(N,O), and C(O,P), we have the following specifications:

- T(A) = 1200, V(A,M) = 10, V(A,N) = 60
- T(B) = 500, V(B,N) = 50, V(B,O) = 40
- T(C) = 9000, V(C,O) = 45, V(C,P) = 90

Estimate the number of tuples for the results:

a) $A \bowtie B$
b) $A \bowtie \sigma_{O=8}(B)$
c) $A \bowtie \sigma_{N=12 \text{ AND } O=8}(B)$
d) $\sigma_{M=7}(A) \bowtie B$ (Assume $\sigma_{M=7}(A)$ is not empty.)
e) $\pi_M(A) \bowtie B$

- i) Assuming projection will not eliminate duplicates
- ii) Assuming projection will eliminate duplicates
  f) $A \bowtie B \bowtie C$

Exercise 7

UPDATED

For R(X,Y) we have the following specs:

T(R) = 20000, V(R,X) =500, V(R,Y) = 10

Assume Attribute X = 20bytes, Y = 30bytes. And there is a clustering index on attribute X.

With block size 4096 bytes and a block header of 96bytes.

a) What is B(R)?
   i) For Spanned
   ii) Unspanned
b) How many tuples will the following queries return? What will be the size of the result in bytes?
   i) $\sigma_{X=10}(R)$
   ii) $\pi_Y(R)$ (Assume projection will not eliminate duplicates.)
   iii) How many I/O would you roughly need to retrieve the records for (i)? (Assume index to be in memory.)

Exercise 8

For **S(A,B,C)**, we have the following specifications:

- T(S) = 50000, V(S,A) = 1000, V(S,B) = 500, V(S,C) = 50
- Assume **Attribute A = 16 bytes**, **B = 24 bytes**, **C = 40 bytes**.
- There is a **clustering index** on attribute **A**.
- Block size is **8192 bytes**, with a block header of **128 bytes**.

Questions:

a) What is **B(S)**?

i) For **Spanned** storage

ii) For **Unspanned** storage

b) How many tuples will the following queries return? What will be the size of the result in bytes?

i) $\sigma_{A=50}(S)$

ii) $\pi_B(S)$ (**Assume projection will not eliminate duplicates.**)

iii) How many **I/O operations** are needed to retrieve the records for (i)? (**Assume the index is in memory.**)

Exercise 9

For **T(D,E,F,G)**, we have the following specifications:

- **T(T) = 150000, V(T,D) = 2000, V(T,E) = 1000, V(T,F) = 500, V(T,G) = 200**
- Assume **Attribute D = 12 bytes, E = 20 bytes, F = 28 bytes, G = 36 bytes**.
- There is a **clustering index** on attribute **D**.
- Block size is **4096 bytes**, with a block header of **64 bytes**.

Questions:

a) What is **B(T)**?
i) For **Spanned** storage
ii) For **Unspanned** storage

b) How many tuples will the following queries return? What will be the size of the result in bytes?
i) $\sigma_{E=100}(T)$
ii) $\pi_F(T)$ (**Assume projection will not eliminate duplicates.**)
iii) How many **I/O operations** are needed to retrieve the records for (i)? (**Assume the index is in memory.**)

Exercise 10

For **U(X,Y,Z)**, we have the following specifications:

- **T(U) = 100000, V(U,X) = 5000, V(U,Y) = 2000, V(U,Z) = 1000**
- Assume **Attribute X = 18 bytes, Y = 22 bytes, Z = 32 bytes**.
- There is a **clustering index** on attribute **X**.
- Block size is **8192 bytes**, with a block header of **128 bytes**.

Questions:

a) What is **B(U)**?
i) For **Spanned** storage
ii) For **Unspanned** storage

b) How many tuples will the following queries return? What will be the size of the result in bytes?
i) $\sigma_{X=200}(U)$
ii) $\pi_Y(U)$ (**Assume projection will not eliminate duplicates.**)
iii) How many **I/O operations** are needed to retrieve the records for (i)? (**Assume the index is in memory.**)

Here are **three additional exercises** similar to the one you provided, covering **spanned vs. unspanned storage, query result estimation, and I/O cost analysis**.

Exercise 1

For relation **S(A, B, C)**, we have the following specifications:

- **T(S) = 50000, V(S, A) = 1000, V(S, B) = 2500, V(S, C) = 500**
- Assume attribute sizes:
  - **A = 15 bytes, B = 25 bytes, C = 40 bytes**
- **Block size = 4096 bytes** with a **block header of 96 bytes**
- There is a **clustering index on attribute B**.

Questions

(a) Compute **B(S)** (number of blocks needed) for:

- **Spanned storage**
- **Unspanned storage**

(b) How many tuples will the following queries return? What will be the **size of the result in bytes**?

1. **σB=500(S)**
2. **πA(S)** (Assume projection does not eliminate duplicates.)

(c) Estimate the **I/O cost** to retrieve the tuples for **σB=500(S)** (Assume the index is in memory).

Exercise 2

For relation **T(P, Q, R)**, we have the following specifications:

- **T(T) = 250000, V(T, P) = 5000, V(T, Q) = 10000, V(T, R) = 2000**
- Assume attribute sizes:
  - **P = 10 bytes, Q = 35 bytes, R = 45 bytes**
- **Block size = 8192 bytes** with a **block header of 128 bytes**
- There is a **secondary index on attribute P** (not clustering).

Questions

(a) Compute **B(T)** for:

- **Spanned storage**
- **Unspanned storage**

(b) Estimate the number of tuples and result size for:

1. **σP=1000(T)**
2. **πQ(T)** (Assume projection does not eliminate duplicates.)

(c) Estimate the **I/O cost** to retrieve the tuples for **σP=1000(T)** (Assume index is in memory, but records are scattered across blocks).

Exercise 3

For relation **R(X, Y, Z, W)**, we have the following specifications:

- **T(R) = 120000, V(R, X) = 8000, V(R, Y) = 3000, V(R, Z) = 1200, V(R, W) = 600**
- Assume attribute sizes:
  - **X = 12 bytes**, **Y = 18 bytes**, **Z = 20 bytes**, **W = 30 bytes**
- **Block size = 4096 bytes** with a **block header of 64 bytes**
- There is a **clustering index on attribute Z**.

Questions

(a) Compute **B(R)** for:

- **Spanned storage**
- **Unspanned storage**

(b) Estimate the number of tuples and result size for:

1. **σZ=300(R)**
2. **πX,Y(R)** (Assume projection does not eliminate duplicates.)

(c) Estimate the **I/O cost** to retrieve the tuples for **σZ=300(R)** (Assume index is in memory).

Exercise 8

UPDATED

For R(A,B,C,D) with following specs:

T(R) = 1000, V(R,A) =1, V(R,B) = 30, V(R,C)=10, V(R,D)=20

Estimate the number of Tuples for the results and the V(W,X), W=result, X={A,B,C,D}. For parts c and d assume projection to eliminate duplicates.

a) $\sigma_{A=10}(R)$
b) $\sigma_{B=5 \text{ AND } C<10}(R)$
c) $\pi_{A}(R)$
d) $\pi_{A,B}(R)$

Question 1: Tuple and Distinct Value Estimation

For relation $R(A, B, C, D)$ with the following specifications:

- $T(R) = 2000$
- $V(R, A) = 1$
- $V(R, B) = 50$
- $V(R, C) = 20$
- $V(R, D) = 40$

Estimate the number of tuples for the results and the $V(W, X)$, where W is the result and $X = \{A, B, C, D\}$. For parts c and d, assume projection to eliminate duplicates.

a) $\sigma_{A=1}(R)$
b) $\sigma_{B=10}(R)$
c) $\pi_{A,B}(R)$
d) $\pi_{C,D}(R)$

Question 2: Tuple and Distinct Value Estimation

For relation $R(A, B, C, D)$ with the following specifications:

- $T(R) = 1500$
- $V(R, A) = 2$
- $V(R, B) = 40$
- $V(R, C) = 15$
- $V(R, D) = 30$

Estimate the number of tuples for the results and the $V(W, X)$, where W is the result and $X = \{A, B, C, D\}$. For parts c and d, assume projection to eliminate duplicates.

a) $\sigma_{A=2}(R)$
b) $\sigma_{C=5}(R)$
c) $\pi_{A,C}(R)$
d) $\pi_{B,D}(R)$

Question 3: Tuple and Distinct Value Estimation

For relation $R(A, B, C, D)$ with the following specifications:

- $T(R) = 3000$
- $V(R, A) = 3$
- $V(R, B) = 60$
- $V(R, C) = 25$
- $V(R, D) = 50$

Estimate the number of tuples for the results and the $V(W, X)$, where W is the result and $X = \{A, B, C, D\}$. For parts c and d, assume projection to eliminate duplicates.

a) $\sigma_{B=20}(R)$
b) $\sigma_{D=10}(R)$
c) $\pi_{A,D}(R)$
d) $\pi_{B,C}(R)$

Exercise 9

For S(A,B) with following specs:
T(S) = 40000, B(S) = 500
V(S,A) =4000, V(S,B) = 2000
Assume S has a primary dense index on A and secondary index on B.
How many IOs would you expect the following operations to take if the indices are in memory.

a) $\sigma_{A=10}(S)$
b) $\sigma_{B=5}(S)$

Exercise 10

For S(A, B) with the following specs:

T(S) = 50,000, B(S) = 600

V(S, A) = 5,000, V(S, B) = 2,500

Assume S has a primary dense index on A and a secondary index on B.

How many IOs would you expect the following operations to take if the indices are in memory?

(a) $\sigma_{A=20}(S)$
(b) $\sigma_{B=15}(S)$

Exercise 11

For S(A, B, C) with the following specs:

T(S) = 75,000, B(S) = 800

V(S, A) = 7,500, V(S, B) = 3,000, V(S, C) = 1,500

Assume S has a primary dense index on A and a secondary index on B.

How many IOs would you expect the following operations to take if the indices are in memory?

(a) $\sigma_{A=30}(S)$
(b) $\sigma_{B=10}(S)$

Exercise 12

For S(A, B) with the following specs:

T(S) = 100,000, B(S) = 1,000

V(S, A) = 10,000, V(S, B) = 5,000

Assume S has a primary dense index on A and a secondary index on B.

How many IOs would you expect the following operations to take if the indices are in memory?

(a) $\sigma_{A=50}(S)$
(b) $\sigma_{B=25}(S)$

Exercises with Solutions

Exercise 13 (With Solution)

For S(A, B) with the following specs:
T(S) = 80,000, B(S) = 900
V(S, A) = 8,000, V(S, B) = 4,000

Assume S has a primary dense index on A and a secondary index on B.
How many IOs would you expect the following operations to take if the indices are in memory?

Solution

(a) $\sigma_{A=40}(S)$

- A has a **primary dense index**, so we can directly access the tuples.
- The number of tuples matching $A = 40$ is:

$$\frac{T(S)}{V(S,A)} = \frac{80,000}{8,000} = 10$$

- Since the index is in memory, the IO cost is **10** (one per tuple).

(b) $\sigma_{B=20}(S)$

- B has a **secondary index**, so we must retrieve **all** tuples matching $B = 20$.
- The number of tuples matching $B = 20$ is:

$$\frac{T(S)}{V(S,B)} = \frac{80,000}{4,000} = 20$$

- Since the secondary index is in memory, each tuple access requires an additional IO.
- Total IOs = **20**.

Exercise 14 (With Solution)

For S(A, B) with the following specs:
T(S) = 60,000, B(S) = 700
V(S, A) = 6,000, V(S, B) = 3,000

Assume S has a primary dense index on A and a secondary index on B.
How many IOs would you expect the following operations to take if the indices are in memory?

Solution

(a) $\sigma_{A=15}(S)$

- A has a **primary dense index**, so direct access is possible.
- Number of matching tuples:

$$\frac{T(S)}{V(S,A)} = \frac{60,000}{6,000} = 10$$

- IO cost: **10**.

(b) $\sigma_{B=30}(S)$

- B has a **secondary index**, so we must scan all matching tuples.
- Number of matching tuples:

$$\frac{T(S)}{V(S,B)} = \frac{60,000}{3,000} = 20$$

- IO cost: **20**.