VIGNESH KUMAR KARTHIKSYAN RAJALAKSHMI	20424508
Name	CWID

Quiz
2

November 25-29, 2018, 11:59PM

CS525 - Quiz 2

Please leave this empty!	1	2	3	Sum

Instructions

- **Due on Blackboard by November 29, 2018 11:59PM Chicago Time**
- This is an individual and not a group assignment. Fraud will result in **0** points
- Things that you are allowed to use
  - Textbook
  - Lecture notes (electronic or printed)
  - Personal notes
- For your convenience the number of points for each part and questions are shown in parenthesis.
- There are **3** parts in this exam
  1. Relational Algebra
  2. Result Size / I-O Cost Estimation
  3. Index Structures

- `I affirm my awareness of the standards of the Illinois Institute of Technology Honor Code`

`Sign` K.R. Vignesh Kumar .....

Part 1 Relational Algebra (Total: 15 Points)

Question 1.1 (15 Points)

Consider relations $R(A, B)$, $S(B, C, D)$ and $T(C, D)$. The following sub-problems ask you to rewrite relational algebra expressions. You can assume that the relations contain sets (not bags). If the requested rewrite is not feasible, state so and briefly explain why. Also, make sure there are no unneeded expressions in your rewrite, e.g., $\pi_{CD}(T)$ and $\sigma_{A \neq A}(R)$ are unneeded.

(a) State whether the following expression is feasible. If so, rewrite the following expression (including the projection, if necessary) by pushing the projection as far down as possible:

$\pi_{AD}[\sigma_{C=5}(R \bowtie S)]$

$\pi_{AD}[(\sigma_{C=5} R) \bowtie (\sigma_{C=5} S)]$

Yes the expression is feasible

∴ the rewritten expression is

$\pi_{AD}[\sigma_{C=5}[\pi_{ABC}(R) \bowtie \pi_{DBC}(S)]]$

(b) State whether the following expression is feasible. If so, rewrite the following expression so it does not contain a union operator and contains one selection operator (instead of two):

$[R \bowtie (\sigma_{C=2} S)] \cup [(\sigma_{A=1} R) \bowtie S]$

Yes the expression is feasible and the rewritten expression is

$\sigma_{(C=2 \vee A=1)}[R \bowtie S]$

(c) State whether the following expression is feasible. If so, What is the minimum number of operators required to express the query represented by the following expression?

$\sigma_{D=2}[(\sigma_{C=2} S) \bowtie (\sigma_{C=1} T)]$

~~No~~ ~~Yes~~ the expression is feasible and the rewritten expression is $[\sigma_{(C=2 \wedge D=2)} \bowtie \sigma_{(C=1 \wedge D=2}(T)]$

∴ Minimum number of operator required = 3

Part 2 Result Size Estimations (Total: 16 Points)

Consider two relations R(A, B, C) and S(B, C, D). We want to estimate the number of tuples and the size of the following expression: $U = \pi_{ACD}[(\sigma_{A=3 \wedge B=5} R) \bowtie S]$

We are given the following information:

- T(R) = 100000; V(R, A) = 20; V(R, B) = 50; V(R, C) = 150.
- T(S) = 5000; V(S, B) = 100; V(S, C) = 200; V(S, D) = 30.
- All attributes are 10 bytes in size.
- We assume query values are selected from values in the relations.

Question 2.1 Estimate Result Size (5 Points)

First consider the innermost select $W = \sigma_{A=3 \wedge B=5} R$. Compute the following values.

1. T(W) = $\frac{T(R)}{V(R,A) * V(R,B)} = \frac{100,000}{20 * 50} = \frac{100,000}{1000} = 100$

   $= 100$

2. S(W) = Size of the W can be calculated by the attributes W(A, B, C)

   $\therefore 10 + 10 + 10 = 30$

3. V(W, A) = We have mentioned A = 3 in the condition, so only one unique value is possible $\therefore V(W, A) = 1$

4. V(W, B) = as we have mentioned B = 5 in Condition

   $V(W, B) = 1$

5. V(W, C) = Since we do not know the unique values of C in W it is possible that all the tuples might have unique values of C.

   $\therefore V(W, C) = 100$

Question 2.2 Estimate Result Size (6 Points)

Next consider the join $\mathbf{Y} = \mathbf{W} \bowtie \mathbf{S}$. Compute the following values

1. T(Y) = $T(W_{(a,b,c)} \bowtie S_{(b,c,d)}) = \frac{T(W) * T(S)}{\max(V(W,b), V(S,b)) * \max(V(W,c), V(S,c))}$

$= \frac{100 * 5000}{\max(1,100) * \max(100,200)} = \frac{500,000}{100 * 200} = \frac{500,000}{20000} = 25$

2. S(Y) =

$S(A, B, C, D)$ since there are 4 attributes

$\therefore$ 40 bytes.

3. V(Y, A) =

We have mentioned $A = 3$ in condition of W

$\therefore V(Y, A) = 1$

4. V(Y, B) =

We have mentioned $B = 5$ in condition of W

$\therefore V(Y, B) = 1$

5. V(Y, C) =

As we dont know the unique value of C it might have all the tuples of Y a unique value of C

$\therefore V(Y, C) = 25$

6. V(Y, D) =

As you dont know the unique values of D

We assume that Y might contain unique value of D in all of the tuples

$\therefore V(Y, D) = 25.$

Question 2.3 Estimate Result Size (5 Points)

Finally consider the full expression $\mathbf{U} = \pi_{ACD}[(\sigma_{A=3 \wedge B=5}\mathbf{R}) \bowtie \mathbf{S}]$. Compute the following values.

1. T(U) = Projection doesn't impact the tuple size.

   ∴ T(U) = 25

2. S(U) = The attributes of Π_{ACD} are A, C, D

   ∴ Size = 10 + 10 + 10 = 30 bytes.

3. V(U, A) = The distinct value of A = 1

   Because we have mentioned in Condition A=3.

4. V(U, C) = The distinct value of C = 25

   Since we dont know the unique value of C, we assume that all the rows have unique value of C.

5. V(U, D)=

   The distinct value of D = 25

   Since we dont know the unique value of C, we assume that all the rows have the unique D values.

Question 2.4 I/O Cost Estimation (15 Points)

Assume a database system that holds two important relations, R and S, that are frequently joined over a common attribute A. The relations are currently stored as rows.

- Relation R has three attributes, A, B, and C where each 10 bytes long.
- Relation S has three attributes, A, D, and E, where A is 10 bytes long and D and E each 15 bytes long.
- Each of R and S contain 64000 tuples.
- In addition to its header, each disk block can hold 6400 bytes.
- To perform R ⋈ S we use a simple hash join algorithm as described in our class notes. Assume there is enough main memory.
- The expected number of resulting tuples in R ⋈ S is 10.
- The tuples of relation R are stored contiguously in blocks. They are not sorted.
- The S tuples are also contiguous, spanned, and unsorted.

What is the number of IOs needed for the hash join? (Do not include the cost of writing the final result to disk.)

In Relation R

Each tuple is 30 bytes ∴ 30 * [illegible]

In Relation S

Each tuple is 40 byte (10+15+15) ∴ 40 [illegible]

∴ Total size of R = 30 * 64000

= 1920000

Total size of S = 40 * 64000

= 2560000

Each block holds 6400 byte

∴ R = $\frac{1920000}{6400}$ = 300 blocks

S = $\frac{2560000}{6400}$ = 400 blocks

Join operation = 3(B(R)) + 3 B(S)

∴ 3(300) + 3(400)

= 900 + 1200 = 2100 disk IOs.

Part 3 Index Structures (Total: 30 Points)

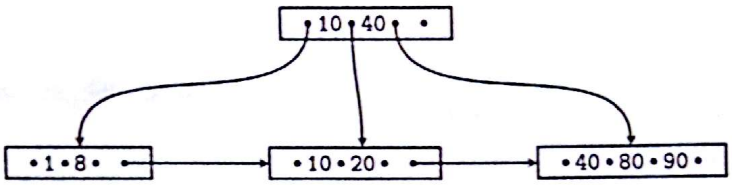
Question 3.1 B+-tree Operations (20 Points)

Given is the B+-tree shown below ($n = 3$). Execute the following operations and write down the resulting B+-tree after each step:

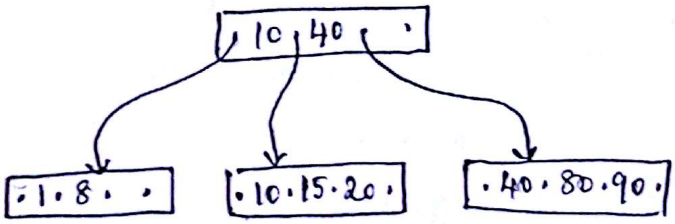
insert(15), insert(30), insert(110), delete(30), delete(10), delete(80)

When splitting or merging nodes follow these conventions:

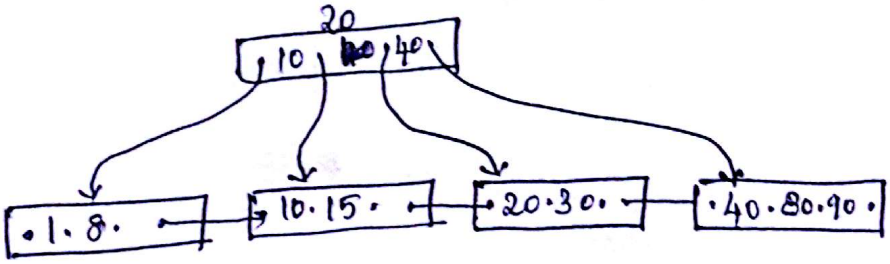
- **Leaf Split**: In case a leaf node needs to be split, the left node should get the extra key if the keys cannot be split evenly.
- **Non-Leaf Split**: In case a non-leaf node is split evenly, the "middle" value should be taken from the right node.
- **Node Underflow**: In case of a node underflow you should first try to redistribute and only if this fails merge. Both approaches should prefer the left sibling.



(i) insert 15 :

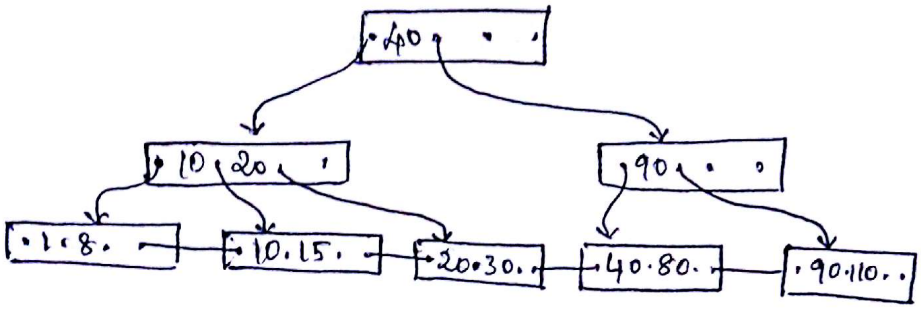


(ii) insert 30 :

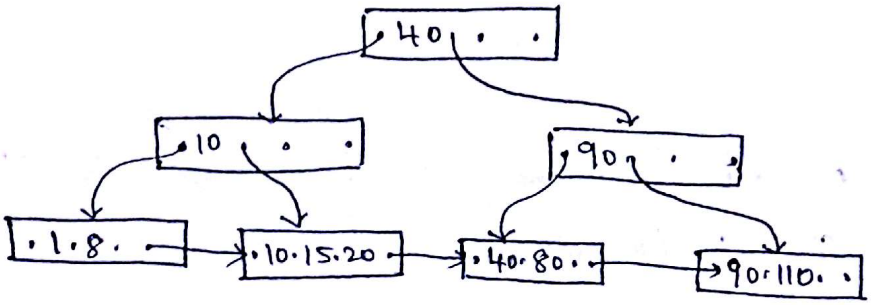


This page left blank intentionally. There are no more questions.

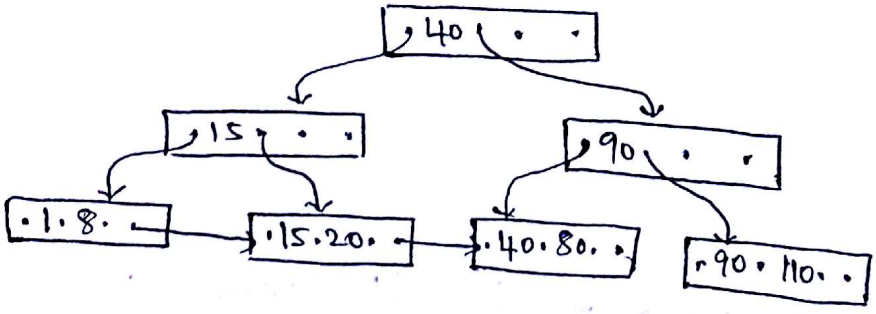
Insert 110 :



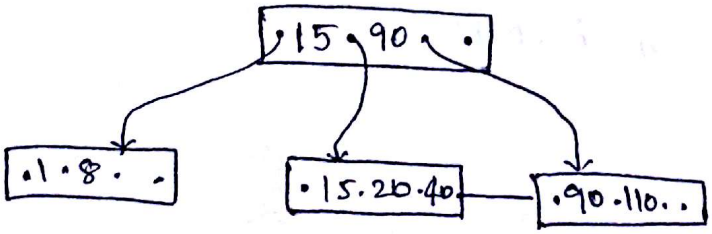
delete 30 :



delete 10 :



delete 80 :



Question 3.2 Extensible Hash Operations (10 Points)

Consider an **extensible hash** structure with the following characteristics:

- Buckets can hold up to two records.
- No overflow blocks are allowed.
- The hash function we use generates `b = 4` bits total.
- Initially the extensible hash table is empty.

Say we insert `X` records, where the search key of each record generates a distinct 4-bit hash value (no collisions). No records are deleted during this process. We are told that after the `X` insertions, 4 buckets have been allocated. (Note that the previous sentence does not refer to the size of the directory.)

4.1. What is the minimum possible value of `X`?

X = 3 is the minimum possible value if there are 4 buckets.

4.2. In the same scenario, what is the maximum possible value of `X`?

X = 8 is the maximum possible value if there are 4 buckets.