Name

CWID

Quiz
2

Due Nov 21th, 2018

CS525 - Advanced Database Organization

Solutions

Please leave this empty! 1 2 3 4 5 6 7

Sum

Instructions

- Multiple choice questions are graded in the following way: You get points for correct answers and points subtracted for wrong answers. The minimum points for each questions is **0**. For example, assume there is a multiple choice question with 6 answers - each may be correct or incorrect - and each answer gives 1 point. If you answer 3 questions correct and 3 incorrect you get 0 points. If you answer 4 questions correct and 2 incorrect you get 2 points. ...
- For your convenience the number of points for each part and questions are shown in parenthesis.
- There are 3 parts in this quiz
  1. Disk Organization and Buffering
  2. Index Structures
  3. Result Size Estimations
  4. I/O Cost Estimation
  5. Schedules
  6. ARIES (Optional)
  7. Physical Optimization (Optional)

Part 1 Disk Organization and Buffering (Total: 15 Points)

Question 1.1 Page Replacement Clock (15 Points)

Consider a buffer pool with 3 pages using the **Clock** page replacement strategy. Initially the buffer pool is in the state shown below. We use the following notation ${}^{flag}[page]^{dirty}_{fix}$ to denote the state of each buffer frame. *page* is the number of the page in the frame, fix is its fix count, *dirty* is indicating with an Asterix that the page is dirty, and $flag$ is the reference bit used by the Clock algorithm. E.g., ${}^{1}[5]^{*}_{2}$ denotes that the frame stores page 5 with a fix count 2, that the page is dirty, and that the reference bit is set to 1. Recall that Clock uses a pointer S that points to the current page frame (the one to be checked for replacement next). The page frame S is pointing to is indicated by by $\downarrow$. In your solution draw an arrow to the page frame that S is pointing to.

Current Buffer State

$${}^{1}[3]^{*}_{0} \quad {}^{1}\overset{\downarrow}{[10]}_{0} \quad {}^{0}[6]_{0} \quad {}^{1}[4]_{1}$$

Execute the following requests and write down state of the buffer pool after each request.

- p stands for pin
- u for unpin
- d for marking a page as dirty

```
p(3),p(1),p(4),u(3),u(4),u(1),p(7)
```

Solution

p(3)

$${}^{1}[3]^{*}_{1} \quad {}^{1}\overset{\downarrow}{[10]}_{0} \quad {}^{0}[6]_{0} \quad {}^{1}[4]_{1}$$

p(1)

$${}^{1}[3]^{*}_{1} \quad {}^{0}[10]_{0} \quad {}^{1}[1]_{1} \quad {}^{1}\overset{\downarrow}{[4]}_{1}$$

p(4)

$${}^{1}[3]^{*}_{1} \quad {}^{0}[10]_{0} \quad {}^{1}[1]_{1} \quad {}^{1}\overset{\downarrow}{[4]}_{2}$$

u(3)

$${}^{1}[3]^{*}_{0} \quad {}^{0}[10]_{0} \quad {}^{1}[1]_{1} \quad {}^{1}\overset{\downarrow}{[4]}_{2}$$

u(4)

$${}^{1}[3]^{*}_{0} \quad {}^{0}[10]_{0} \quad {}^{1}[1]_{1} \quad {}^{1}\overset{\downarrow}{[4]}_{1}$$

u(1)

$${}^{1}[3]^{*}_{0} \quad {}^{0}[10]_{0} \quad {}^{1}[1]_{0} \quad {}^{1}\overset{\downarrow}{[4]}_{1}$$

p(7)

$${}^{0}[3]^{*}_{0} \quad {}^{1}[7]_{1} \quad {}^{1}\overset{\downarrow}{[1]}_{0} \quad {}^{1}[4]_{1}$$

Part 2 Index Structures (Total: 25 Points)

Assume that you have the following table:

Student

name	gpa	credits
Will Wonton	4.0	40
Joe Joeton	3.5	34
Jill Johnson	2.4	15
John Johnson	3.1	3
Bo Zhao	3.1	6
Yu Wei	3.0	9
Heinz Bert	2.2	30

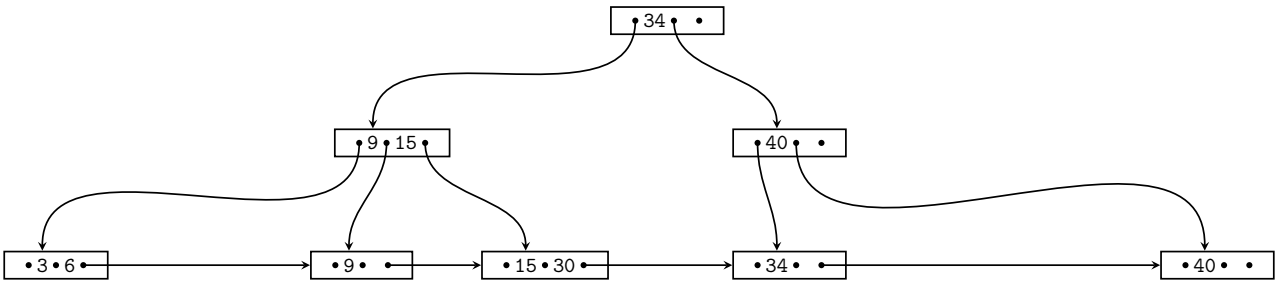
Question 2.1 Construction (9 Points)

Create a B+-tree for table *Student* over attribute *credits* with $n = 2$ (up to two keys per node). You should start with an empty B+-tree and insert the keys in the order shown in the table above. Write down the resulting B+-tree after each step.

When splitting or merging nodes follow these conventions:

- **Leaf Split**: In case a leaf node needs to be split during insertion and n is even, the left node should get the extra key. E.g, if $n = 2$ and we insert a key 4 into a node [1,5], then the resulting nodes should be [1,4] and [5]. For odd values of n we can always evenly split the keys between the two nodes. In both cases the value inserted into the parent is the smallest value of the right node.
- **Non-Leaf Split**: In case a non-leaf node needs to be split and n is odd, we cannot split the node evenly (one of the new nodes will have one more key). In this case the "middle" value inserted into the parent should be taken from the right node. E.g., if $n = 3$ and we have to split a non-leaf node [1,3,4,5], the resulting nodes would be [1,3] and [5]. The value inserted into the parent would be 4.
- **Node Underflow**: In case of a node underflow you should first try to redistribute values from a sibling and only if this fails merge the node with one of its siblings. Both approaches should prefer the left sibling. E.g., if we can borrow values from both the left and right sibling, you should borrow from the left one.

Solution

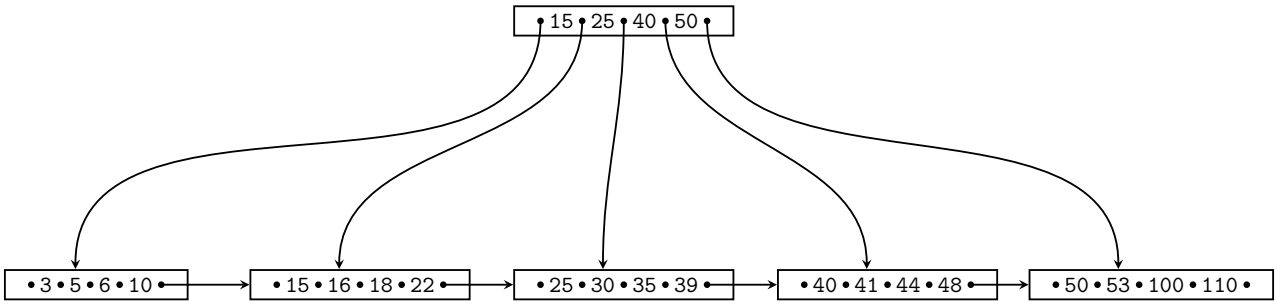


Question 2.2 Operations (9 Points)

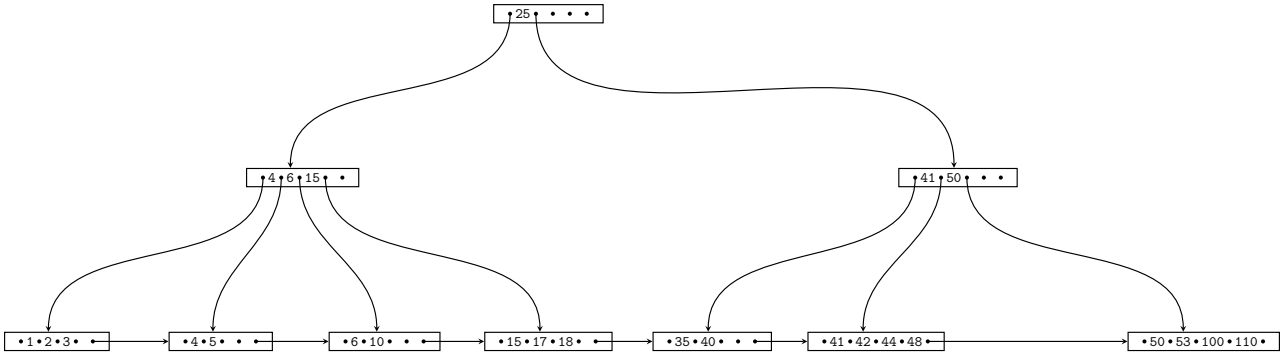
Given is the B+-tree shown below ($n = 4$). Execute the following operations and write down the resulting B+-tree after each operation:

delete(30), delete(16), delete(22), delete(25), delete(39), insert(42), insert(17), insert(1), insert(2), insert(4)

Use the conventions for splitting and merging introduced in the previous question.



Solution



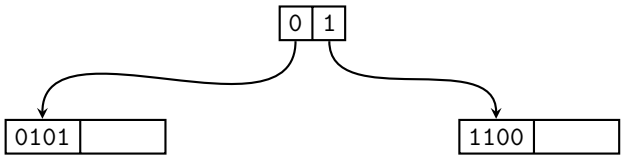
Question 2.3 Extensible Hashing (7 Points)

Consider the extensible Hash index shown below that is the result of inserting values 6 and 4. Each page holds two keys. Execute the following operations

`insert(1),insert(2),insert(3),insert(5),insert(6)`

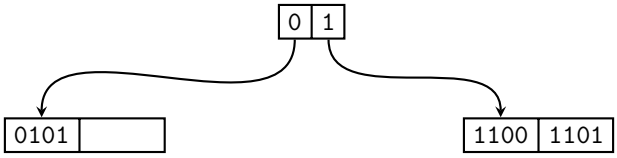
and write down the resulting index after each operation. Assume the hash function is defined as:

x	h(x)
0	1000
1	1101
2	0111
3	0000
4	1100
5	0100
6	0101
7	1001
8	1110

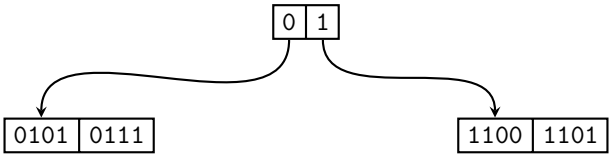


Solution

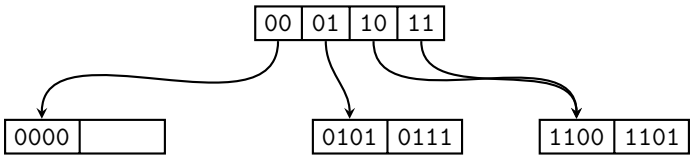
insert(1)



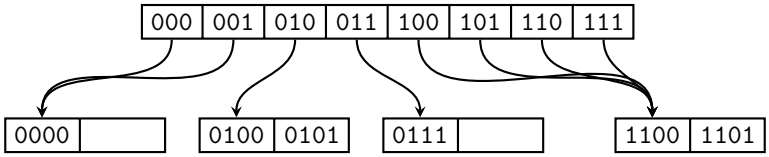
insert(2)



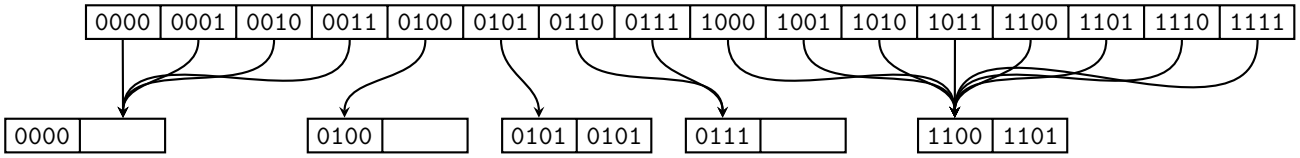
insert(3)



insert(5)



insert(6)



Part 3 Result Size Estimations (Total: 20 Points)

Consider a table book with attributes ISBN, title, author, edition (primary key is ISBN), a table library with loc, budget, public (primary key is loc), and a table catalog with attributes library and book. catalog.library is a foreign key to relation library. Attribute book of relation catalog is a foreign key to relation book. Given are the following statistics:

$$T(book) = 100,000 \qquad T(library) = 100 \qquad T(catalog) = 200,000$$
$$V(book, ISBN) = 100,000 \qquad V(library, loc) = 100 \qquad V(catalog, library) = 100$$
$$V(book, title) = 50,000 \qquad V(library, budget) = 40 \qquad V(catalog, book) = 90,000$$
$$V(book, author) = 30,000 \qquad V(library, public) = 2$$
$$V(book, edition) = 15$$

Question 3.1 Estimate Result Size (4 Points)

Estimate the number of result tuples for the query $q = \sigma_{author=Goethe}(book)$ using the first assumption presented in class (values used in queries are uniformly distributed within the active domain).

Solution

$$T(q) = \frac{T(book)}{V(book, author)} = \frac{100,000}{30,000} = \frac{10}{3} \approx 3.3$$

Question 3.2 Estimate Result Size (5 Points)

Estimate the number of result tuples for the query $q = \sigma_{title=Inferno \vee author=Dante}(book)$ using the first assumption presented in class.

Solution

$$T(q) = (1 - [(1 - \frac{1}{V(book, title)}) \cdot (1 - \frac{1}{V(book, author)})]) \cdot T(book)$$
$$= 1 - (1 - \frac{1}{50,000}) \cdot (1 - \frac{1}{30,000}) \cdot 100,000 = (1 - 0.9025) \cdot 30,000 \approx 5.33$$

Question 3.3 Estimate Result Size (5 Points)

Estimate the number of result tuples for the query $q = \sigma_{(edition=1 \vee edition=3) \wedge author=Jenifer\,Widom}(book)$ using the first assumption presented in class. Assume that the minimal and maximal values in the `edition` attribute are 1 and 15,

Solution

$$T(q) = (1 - [(1 - \frac{1}{V(book, edition)}) \cdot (1 - \frac{1}{V(book, edition)})]) \cdot \frac{1}{V(book, author)} \cdot T(book) \approx 0.43$$

Question 3.4 Estimate Result Size (6 Points)

Estimate the number of result tuples for the query
$q = \sigma_{title=Poetics \wedge author=Aristotle}(book) \bowtie_{title=book} catalog \bowtie_{library=loc} \sigma_{public=false}(library)$
using the first assumption presented in class.

Solution

Let $q_1 = \sigma_{title=Poetics \wedge author=Aristotle}(book)$ and $q_2 = \sigma_{public=false}(library)$.
To estimate the selection result size q_1:

$$T(q_1) = \frac{T(book)}{V(book, title) \cdot V(book, author)} = \frac{100,000}{50,000 \cdot 30,000} = \frac{1}{15,000} \approx 0.0006$$

To estimate the selection result size q_2:

$$T(q_2) = \frac{T(library)}{V(library, public)} = \frac{100}{2} = 50$$

Now for the full query we get

$$\begin{aligned} T(q) &= \frac{T(q_1) \cdot T(catalog) \cdot T(q_2)}{max(V(q_1, title), V(catalog, book)) \cdot max(V(catalog, library), V(q_2, loc))} \\ &= \frac{0.0006 \cdot 200,000 \cdot 50}{max(1, 90,000) \cdot max(100, 50)} = \frac{20,400}{9,000,000} \approx 0.0006 \end{aligned}$$

Part 4 I/O Cost Estimation (Total: 20 Points)

Question 4.1 External Sorting (4 Points)

You have $M = 401$ memory pages available and should sort a relation R with $B(R) = 400,000$ blocks. Compute the number of I/Os necessary to sort R using the external merge sort algorithm introduced in class.

Solution

$$\begin{aligned} IO &= 2 \cdot B(R) \cdot (1 + \lceil log_{M-1}(\frac{B(R)}{M}) \rceil \\ &= 2 \cdot 400,000 \cdot (1 + 2) \\ &= 2,400,000 \end{aligned}$$

Question 4.2 External Sorting (4 Points)

You have $M = 5$ memory pages available and should sort a relation R with $B(R) = 5,000,000,000$ blocks. Compute the number of I/Os necessary to sort R using the external merge sort algorithm introduced in class.

Solution

$$\begin{aligned} IO &= 2 \cdot B(R) \cdot (1 + \lceil log_{M-1}(\frac{B(R)}{M}) \rceil \\ &= 2 \cdot 5,000,000,000 \cdot (1 + 15) \\ &= 160,000,000,000 \end{aligned}$$

Question 4.3 I/O Cost Estimation (6 = 2+2+2 Points)

Consider two relations R and S with $B(R) = 1,000$ and $B(S) = 200,000$. You have $M = 801$ memory pages available. Estimate the minimum number of I/O operations needed to join these two relations using **block-nested-loop join**, **merge-join** (the inputs are not sorted), and **hash-join**. You can assume that the hash function evenly distributes keys across buckets. Justify you result by showing the I/O cost estimation for each join method.

Solution

- **BNL**: R is smaller, thus, keep chunks of R in memory
  $\lceil \frac{B(R)}{M-1} \rceil \cdot [B(S) + min(B(R), (M-1))] = 2 \cdot [200,000 + 800] = 401,600$ I/Os
- **MJ**: We can generate sorted runs of size 801. We need 1 merge pass for the sort for R and 1 merge passes for S. The last merge of R requires 2 pages and the last merge of S requires 250 pages, i.e., the number of sorted runs from R and S small enough to keep one page from each run of both R and S in memory. $3 \cdot B(R) + 3 \cdot B(S) = 3 \cdot 1,000 + 3 \cdot 200,000 = 603,000$ I/Os.
- **HJ**: We need 1 partitioning pass, because we can create 800 buckets and the bucket sizes of R will be 2 after one pass. The cost is $(2+1) \cdot (B(R) + B(S)) = 3 \cdot (1,000 + 200,000) = 603,000$ I/Os.

Question 4.4 I/O Cost Estimation (6 = 2+2+2 Points)

Consider two relations R and S with $B(R) = 6,000,000$ and $B(S) = 20$. You have $M = 21$ memory pages available. Compute the minimum number of I/O operations needed to join these two relations using **block-nested-loop join**, **merge-join** (the inputs are not sorted), and **hash-join**. You can assume that the hash function evenly distributes keys across buckets. Justify you result by showing the I/O cost estimation for each join method.

Solution

- **BNL**: S is smaller, thus, keep chunks of S in memory
  $\lceil \frac{B(S)}{M-1} \rceil \cdot [B(R) + min(B(S), (M-1))] = 1 \cdot [6,000,000 + 20] = 6,000,020$ I/Os
- **MJ**: We can Generate sorted runs of size 21 that means the number of sorted runs from R and S is not low enough after 5 merge passes for R and 0 merge passes for S to keep one page from each run of both R and S in memory. $13 \cdot B(R) + 3 \cdot B(S) = 78,000,000 + 60 = 78,000,060$ I/Os.
- **HJ**: Relation S fits into memory and no partitioning phases are needed. The cost is $B(R) + B(S) = 6,000,020$ I/Os.

Part 5 Schedules (Total: 20 Points)

Question 5.1 Schedule Classes (20 Points)

Indicate which of the following schedules belong to which class. Recall transaction operations are modelled as follows:

$w_1(A)$	transaction 1 wrote item A
$r_1(A)$	transaction 1 read item A
c_1	transaction 1 commits
a_1	transaction 1 aborts

$$S_1 = w_2(A), r_4(B), r_3(E), w_2(C), c_2, r_1(E), w_4(A), w_4(B), c_4, r_3(D), w_3(C), w_3(D), c_3, r_1(D), c_1$$
$$S_2 = r_2(C), r_2(D), w_1(A), r_1(D), r_2(A), c_2, w_3(C), w_3(A), c_3, r_1(B), c_1$$
$$S_3 = r_3(B), w_3(B), r_3(A), w_4(A), r_2(B), w_2(B), c_2, r_4(B), c_4, r_3(A), c_3$$
$$S_4 = r_2(D), r_1(C), r_2(A), r_1(E), r_1(A), r_2(C), w_2(A), r_2(E), c_2, r_1(A), w_1(D), w_1(C), c_1$$

- ■ S_1 is recoverable
- ■ S_1 is cascade-less
- ■ S_1 is strict
- ■ S_1 is conflict-serializable
- ■ S_1 is 2PL

- ❑ S_2 is recoverable
- ❑ S_2 is cascade-less
- ❑ S_2 is strict
- ■ S_2 is conflict-serializable
- ■ S_2 is 2PL

- ❑ S_3 is recoverable
- ❑ S_3 is cascade-less
- ❑ S_3 is strict
- ❑ S_3 is conflict-serializable
- ❑ S_3 is 2PL

- ■ S_4 is recoverable
- ■ S_4 is cascade-less
- ■ S_4 is strict
- ❑ S_4 is conflict-serializable
- ❑ S_4 is 2PL

Part 6 Optional: ARIES (Total: 10 Optional Points)

Question 6.1 Recovery (10 Points)

Consider the state of the log and pages on disk shown below. For simplicity we do not show the actual undo/redo actions for updates, but instead show only the affected page. Assume a crash occurred after the last log entry. Answer the following questions:

1. **Analysis**: Write down the result of the analysis phase (`RedoLSN`, Transaction Table, Dirty Page Table)
2. **Redo**: Which pages will be loaded from disk during redo? Which pages will be modified during redo?
3. **Undo**: Write down the additional log entries that will be written during undo.

Log

LSN	Type	TID	PrevLSN	UndoNxtLSN	Data
1	begin	1	-	-	-
2	update	1	1	-	Page 1
3	begin	2	-	-	-
4	begin	3	-	-	-
5	update	2	4	-	Page 1
6	begin_cp	-	-	-	-
7	update	3	5	-	Page 1
8	update	3	7	-	Page 3
9	update	2	5	-	Page 4
10	update	1	2	-	Page 4
11	update	2	9	-	Page 4
12	commit	2	12	-	-

Disk

PageID	PageLSN
1	5
2	0
3	8
4	9
5	0

Solution

(1):
RedoLSN: 2
Transaction Table: $< T_1, u, 10, - >, < T_3, u, 8, - >$
Dirty Page Table: $< 1, 2 >, < 3, 7 >, < 4, 8 >$

(2):
Pages 1,3, and 4 have to be loaded from disk.
Only page 1 and 4 will be modified based on redo info from log entries 7, 10, and 11.

(3):
Transactions T_1 and T_3 will be rolled back. The CLRs written during undoing these transactions' updates is shown below.

LSN	**Type**	**TID**	**PrevLSN**	**UndoNxtLSN**	**Data**
14	CLR	1	-	2	Page 4
15	CLR	3	-	7	Page 3
16	CLR	3	-	4	Page 1
17	CLR	1	-	1	Page 1

Part 7 Bonus: Physical Optimization (Total: 10 Bonus Points)

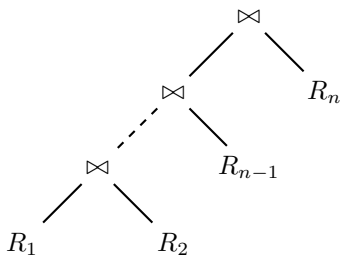
Consider the following relations $R(A, B)$, $S(C, D)$, $T(F, G)$ with $S = \frac{1}{10}$ (10 tuples fit on each page). The sizes and value distributions are:

$N(R) = 1,000$	$V(R, A) = 1,000$	$V(R, B) = 500$
$N(S) = 10$	$V(S, C) = 10$	$V(S, D) = 10$
$N(T) = 1,000$	$V(T, F) = 10$	$V(T, G) = 500$

Question 7.1 Greedy Join Enumeration (10 Points)

Use the greedy join enumeration algorithm to find the cheapest plan for the join $R \bowtie_{B=C} S \bowtie_{D=F \wedge G=A} T$. Assume that **nested-loop** (not the block based version) is the only available join implementation with the left input being the "outer" (for each tuple from the outer we have to scan the whole inner relation). Furthermore, there are no indicies defined on any of the relations (that is you have to use **sequential scan** for each of the relations). As a cost model consider the **total number of I/O operations**. For example, if you join two relations with $5,000$ and $10,000$ tuples with $S = \frac{1}{10}$, where the $5,000$ tuple relation is the outer, then the cost would be $5,000,000$ (scan the inner 5000 times) + 500 to scan the other once. The total cost is then $5,000,500$ I/Os. Assume that the system supports pipelining for the outer input of a join. That is if you join the result of a join with a relation where the join result is the outer, then there is no I/O cost for scaning the outer. Also under these assumptions you never have to store join results to disk. Hint: You will have to estimate the size of intermediate results. Use the estimation based on the number of values and not the one based on the size of the domain. Use the assumption that the number of values in a join attribute of a join result is the minimum of the number of values in the join attribute of each input.

Write down the state after each iteration of the algorithm using the following notation. Write $(((R_1, R_2), \ldots, R_{n-1}), R_n)^{C,S}$ to denote a plan as shown below with I/O cost C and result size S. Alternatively you are also allowed to draw join trees as shown below.



Solution

Calculate Result Sizes:

Using the formula from class the estimated result sizes are:

$$T(R \bowtie S) = \frac{T(R) \cdot T(S)}{max(V(R,B), V(S,C))} = \frac{1,000 \cdot 10}{max(500,10)} = 20$$
$$T(S \bowtie T) = \frac{T(S) \cdot T(T)}{max(V(S,D), V(T,F))} = \frac{10 \cdot 500}{max(10,10)} = 500$$
$$T(R \bowtie T) = \frac{T(R) \cdot T(T)}{max(V(R,A), V(T,G))} = \frac{1,000 \cdot 1,000}{max(1,000,500)} = 1,000$$
$$R(R \bowtie S \bowtie T) = \frac{T(R) \cdot T(S) \cdot T(T)}{max(V(R,B), V(S,C)) \cdot max(V(S,D), V(T,F)) \cdot max(V(T,G), V(R,A))} = \frac{1,000 \cdot 10 \cdot 500}{500 \cdot 10 \cdot 1,000} = 1$$

Initialization:

$$(R)^{100;1000}, (S)^{1;10}, (T)^{100;1,000}$$

---

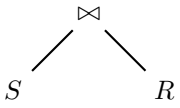
n = 1:

Here we have 6 different options how to join two of the plans from the initialization:

$$(R \bowtie S)^{1,100;20}, (R \bowtie T)^{100,100;1,000}, (S \bowtie R)^{1,001;20},$$
$$(S \bowtie T)^{1,001;500}, (T \bowtie R)^{100,100;1,000}, (T \bowtie S)^{1,100;500}$$

As an example take the join $(R \bowtie S)$. Here R is the outer and S is the inner. The cost is computed as: For each tuple from R ($T(R)$) we have to scan S once (1 I/O). Thus, the cost is $B(R)+T(R)\cdot B(S) = 100+1,000\cdot 1 = 1,100$ I/Os. Greedy join enumeration chooses the plan with the lowest cost (either $S \bowtie R$ or $S \bowtie T$). Let us use $S \bowtie R$ which has a smaller result.

$1,001;20$



$, (T)^{100;1,000}$

---

n = 2:

Now we need to consider two join options.
For $((S \bowtie R) \bowtie T)$ we pipeline the result of $(S \bowtie R)$ so the cost is:

$$Cost(S \bowtie R) + T(S \bowtie R) \cdot B(T) = 1,001 + 20 \cdot 100 = 3,001$$

Recall that the assumption is that only the outer input of the join can be pipelined. For $(T \bowtie (S \bowtie R))$, the result of the join $(S \bowtie R)$ is the "inner", so we have to store the result of $S \bowtie R$ on disk resulting in $B(S \bowtie R)$ additional I/O. Since $(S \bowtie R)$ has 20 result tuples and $S(S \bowtie R) = S(S)+S(R) = 1/10+1/10 = 1/5$ it follows that $B(S \bowtie R) = 4$. Thus, the total cost is

$$Cost(S \bowtie R) + B(S \bowtie R) + B(T) + T(T) \cdot B(S \bowtie R) = 1,001 + 4 + 100 + 1,000 \cdot 4 = 5,401$$

$$(S \bowtie R) \bowtie T)^{3,001;1}, (T \bowtie (S \bowtie R)^{5,401;1}$$

$3,001;1$