CS411 Database Systems

Fall 2004, Prof. Chang

Department of Computer Science
University of Illinois at Urbana-Champaign

Final Examination

December 17, 2004

Solution

Problem 1 (*12 points*) Misc. Concepts

(1) True

(2) True

(3) False

(4) True

(5) False

(6) False

(7) False

(8) True

(9) True

(10) False

(11) False

(12) True

Problem 2 (*18 points*) Short Answer Questions

(1) 100

(2) *AB CD CA*

(3) $\pi_{a,d}(\sigma_{a>10}(R \bowtie_{b=c} S))$

(4) (SELECT *a, b* FROM R WHERE $a > 5$)
EXCEPT (SELECT *a, b* FROM S)

(5) 3 nodes

(6) Index, repeating fields, and so on

(7) 3NF

(8) 200

(9) $\sigma_\theta(R \times S)$

Problem 3 (*8 points*) Schema Decomposition

(a) No. Unless $A \rightarrow B$ or $A \rightarrow C$ is satisfied in R, $R1 \bowtie R2$ may generate entries that do not belong to R.

(b) See figure 1.

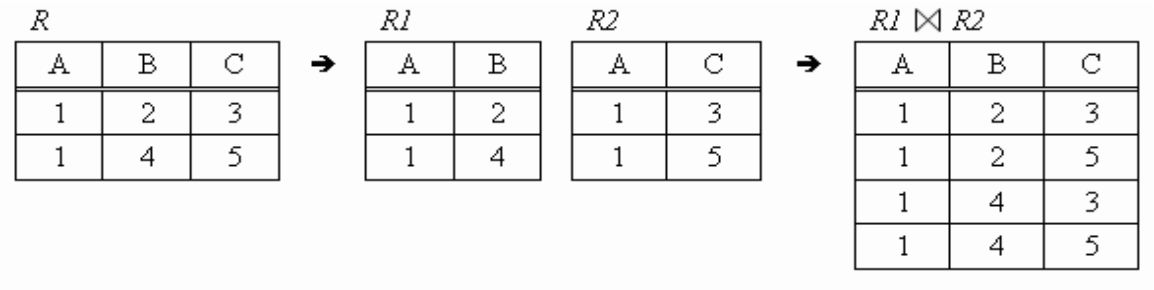

Figure 1: Problem 3.(b)

Problem 4 (*9 points*) Query Languages

(a) $\pi_{score}(\sigma_{name\ =\ "Alex"\ and\ exam="final"}(Scores))$

(b) $\pi_{name}(\sigma_{M.score\ <\ F.score}((\rho_M(Scores)) \bowtie_{M.name\ =\ F.name} (\rho_F(Scores))))$

(c) SELECT *score*, COUNT(*) FROM *Scores* WHERE *exam* = "midterm"
GROUP BY *score* ORDER BY *score*

Problem 5 (*10 points*) Indexing: B+ tree

(a) See figure 2.

(b) Yes. We can change ordering so that more leaf nodes are full.
For example, 10 20 40 50 70 80 30 60 90 100
See figure 3.

Problem 6 (*10 points*) Query Processing

(a) M=3
Need one block to read one block of relation R into memory buffer.
Need one block to read one block of relation S into memory buffer.
Need one block to hold intermediate result of $R \bowtie S$ to write to disk.

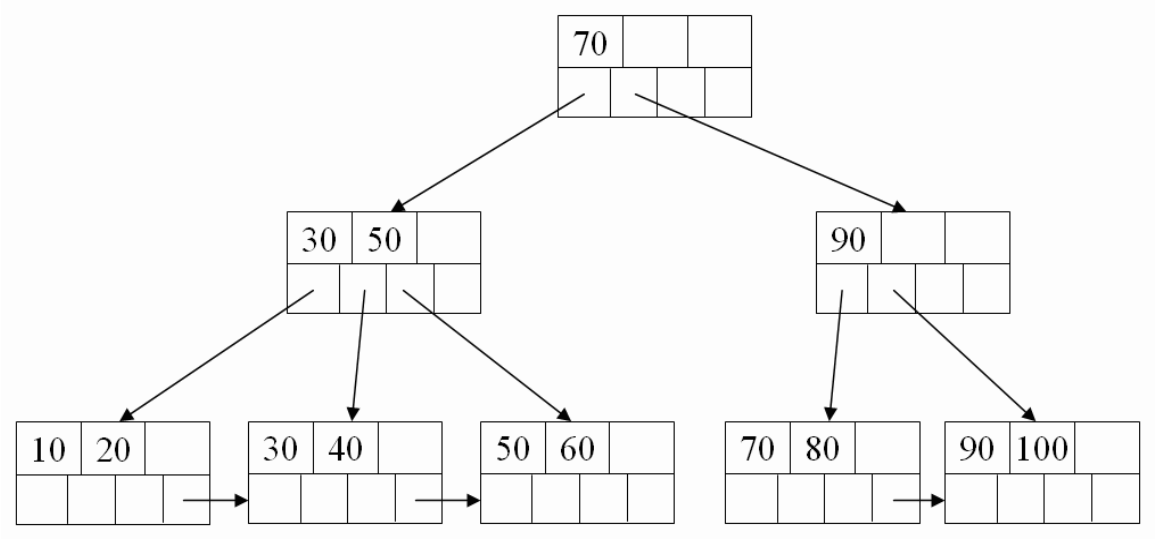

Figure 2: Problem 5.(a)

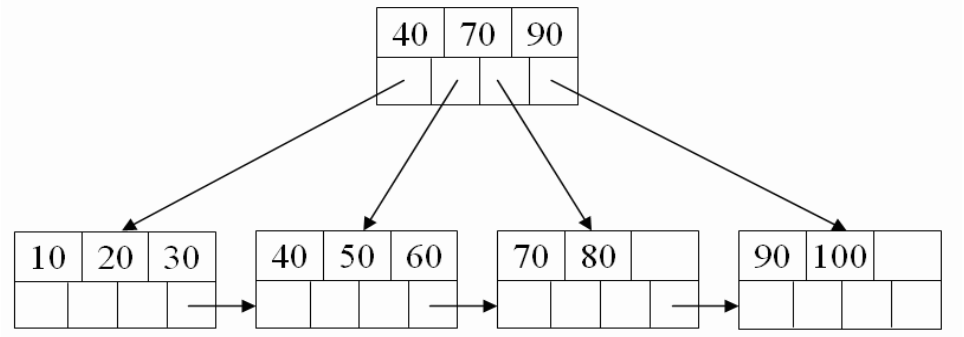

Figure 3: Problem 5.(b)

(b) 1. $B_s < M^2$
There are B_s/M runs after the first phase. In the second phase, we are going to use one block for each sorted sublist and another block for output.
2. B_t
This is the number of blocks of T.
3. $1 + B_s/M + B_t \leq M$
We need only one block for reading R since we have already sorted $R.b$, B_s/M blocks for reading the sorted sublists of S, B_t blocks for T.

Problem 7 (*19 points*) Query Optimization

(a) 15
The total number of tree shapes for 4 relations is 5 as figure 4. But, because we assume that join orders are symmetric, the tree shapes (b), (d), and (e) in figure 4 are same as the tree shapes (a). So, we only consider shape (a) and (c).
Shape (a) - left deep tree: ${}_4C_2 \times 2 = 12$
Shape (c) - busy tree: ${}_4C_2/2 = 3$

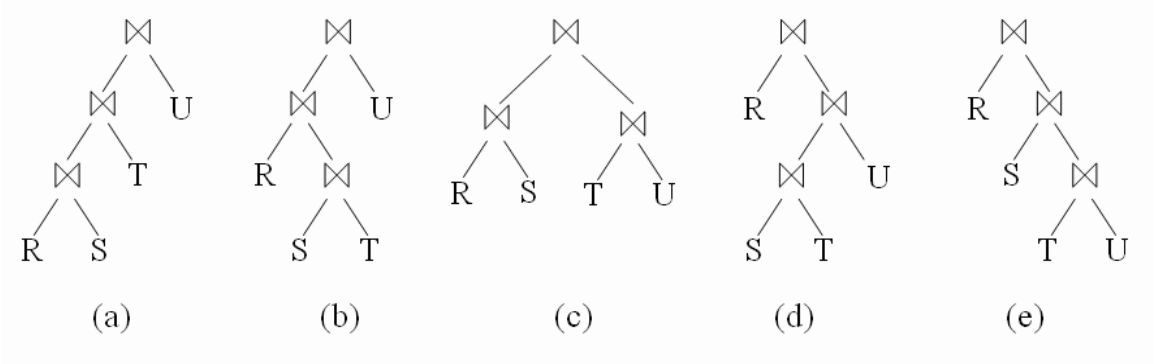

Figure 4: Problem 7

(b) 12

(c) 1. $E.cid$ = "cs411" AND $C.cid$="cs411"
To reduce the size of the intermediate results of the two joins.
2. $\pi_{sname,\ ctitle,\ iname}((\sigma_{E.cid="cs411"}(S \bowtie E)) \bowtie (\sigma_{C.cid="cs411"}(C \bowtie I)))$

Problem 8 (*14 points*) Failure Recovery

(a) Output A, Output B before Action 5.(for *T1*)
Output B before Action 7.(for *T2*)
Output A, Output B before Action 12.(for *T3*)
Output C before Action 15.(for *T4*)
Output D before Action 17.(for *T5*)

(b) 1. Between Action 4 and 5, <START Checkpoint(*T1, T2*)>
2. Between Action 7 and 8, <END Checkpoint>

(c) We need to backtrack to <START Checkpoint (*T1, T2*)>

(d) Output A, Output B, Output C, Output D after Action 17.