
King Fahd University of Petroleum and Minerals

Information and Computer Science Department

ICS 541: Database Design & Implementation

DO NOT OPEN UNTIL INSTRUCTED TO DO SO!!!!

Write clearly, precisely, and briefly!!

ID:							
Name:							

Grades		
Section	Max	Scored
A	26	
B	42	
C	12	
D	8	
TOTAL	88	

A. Questions in query execution**[26 pts]**

1. If $B(S) = B(R) = 10,000$ and $M = 1000$, what is the number of disk I/O's required for hybrid hash join? **[8]**

2. What is the number of disk I/O's if we join S and R using simple hash join? **[4]**

3. Suppose $B(R) = 1000$ and $T(R) = 40,000$. Let \mathbf{b} be one of the attributes of R and there is an index on \mathbf{b} . If we ignore the cost of accessing index blocks, what will be the worst-case number of disk I/O's required to

execute the operation $\sigma_{b=0}(R)$ if: [10]

a) R is clustered, but we don't use the index

b) R is clustered, but we don't use the index

c) $V(R, b) = 200$ and the index is clustering

d) $V(R, b) = 20$ and index is nonclustering

e) If $V(R, b) = 40,000$

4. Explain why the clock buffer-replacement strategy is an approximation of LRU. [4]

B. Questions in query compilation.**[42 pts]**

1. Assume the following schema and query in solving the subsequent questions.

Schema:

Student(Id,Name,Major,Year)

Department(Code,DeptName,Location)

Query:

```
SELECT DeptName FROM Department WHERE Code IN  
(SELECT Major FROM Student WHERE Year=4)
```

- a) Draw the parse tree of the above query.

[6]

b) Convert the parse tree of question (a) to logical query plan. [4]

- c) Using heuristics optimize the logical query plan of question (b).
[6]

2. Below are the vital statistics for relations W, X, Y, and Z. [16]

W(a, b)	X(b, c)	Y(c, d)	Z(d, e)
T(W) = 100	T(X) = 200	T(Y) = 300	T(Z) = 400
V(W, a) = 20	V(X, b) = 50	V(Y, c) = 50	V(Z, d) = 40
V(W, b) = 60	V(X, c) = 100	V(Y, d) = 50	V(Z, e) = 100

Estimate the sizes of relations that are the result of the following expressions:

a) $W \bowtie X \bowtie Z$

b) $\sigma_{a=10}(W)$

c) $\sigma_{c=20}(Y)$

d) $\sigma_{c=20}(Y) \bowtie Z$

e) $W \bowtie X \bowtie Y$

f) $\sigma_{c > 10}(Z)$

g) $\sigma_{a=10 \text{ AND } b=2}(W)$

h) $\sigma_{a=10 \text{ AND } b > 2}(W)$

3. Consider the join of four relations R, S, T, and U. Assume each have 2000 tuples. Their attributes and the estimated sizes of value sets for each relation is:

R(a, b)	S(b, c)	T(C, d)	U(d, a)
V(R, a) = 200			V(U, a) = 100
V(R, b) = 400	V(S, b) = 200		
	V(S, c) = 1000	V(T, c) = 40	
		V(T, d) = 100	
			V(U, d) = 2000

- a) Show the join groupings that a greedy algorithm will choose. [6]

- b) What is the total cost (the size of the intermediate relations) of the join grouping chosen by the greedy algorithm? [4]

C. Questions in recovery.**[12 pts]**

1. Explain in detail how the Checkpointing is used during recovery in each of the following logging strategies. (You need to explain in detail how far back we go in the log files during recovery).

a) Undo logging

b) Redo logging

c) Undo/redo logging

D. Questions in data integration**[8 pts]**

1. Explain the difference between federated databases, mediation, and data warehousing.
2. Assume the association rules and following transactions in answering the subsequent questions:

Transaction ID	Time	Item bought
1	6:30	Milk, bread, cookies, juice
2	7:30	Milk, juice
3	8:30	Milk, eggs
4	9:30	Milk, bread
5	10:30	Coffe, bread, juice
6	11:30	Bread cookies, coffee

- a) What is the support of milk \rightarrow juice
- b) What is the support of bread \rightarrow juice
- c) What is the confidence of milk \rightarrow juice
- d) What is the confidence of bread \rightarrow juice