# Database Management Systems
Practice Exam Questions

Question 1 Query Optimization Consider the following relations where the key of the relation is underlined, all attributes are declared to be non-null, and the statistics known to the optimizer are given below. There is a unclustered B$^+$-tree index on Employee.Sal, a clustered Hash index on Emp.Name, and a clustered B$^+$-tree on Project.Name. Both B$^+$-trees have a height of two.

Emp[<u>Name</u>, Proj, Sal, Dept], Dept[<u>Name</u>, HQ, Head], Project[<u>Name</u>, Budget]

Pages(Emp)=1,000 Tuples(Emp)=10,000 Number Distinct Values of Emp.Proj (i.e., $|Emp.Proj|$) is 100

Number of Distinct Values of Emp.Sal=5,000, lowest value is 10,001, highest is 110,000

Number of distinct values of Emp.Dept=1,000

Pages(Dept)=100 Tuples(Dept)=1,000 $|Dept.HQ| = 1,000 |Dept.Head| = 50$

Pages(Project)=100 Tuples(Project)=1,000 $|Project.Budget| = 1,000$

Assume there are ten buffer pages available.

SELECT DISINTCT D.Name, P.Name
FROM Emp E, Dept D, Project P
WHERE E.Dept = D.Name and E.Proj = P.Name and E.Sal = 100,000

1. What is the estimated cost of the following plan? To receive partial credit, clearly show any estimates for intermediate results. (Note: SortScan(P,Name) just refers to a file scan of P. But because P is sorted on Name, this scan produces tuples in Name order. For the Nested Loop, please use a Block Nested-Loop join.)
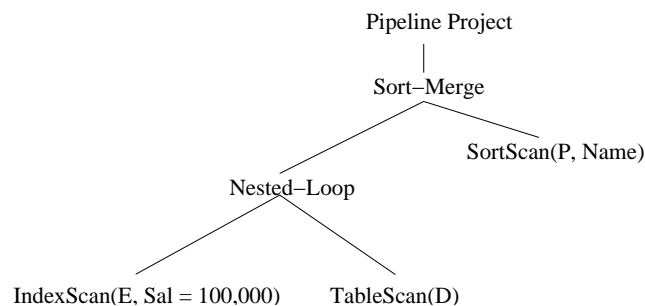
2. What is the estimated size of the result of this query? How would this change if the predicate **E.Sal = 100,000** were changed to **E.Sal > 100,000**?

3. How far off could these size estimates be? Give the maximum and minimum number of tuples that could be returned by this query (with the original **E.Sal = 100,000** predicate).

Question 2 Transaction Management Indicate whether each of the following schedules is *serial, seri-alizable, 2PL, strict 2PL*. Recall that a 2PL schedule is one that could be produced by a two-phase locking scheduler. A strict 2PL schedule is one that could be prodcued by a two phase locking scheduler which holds write locks until end of transaction.

1. $R_1(A), W_2(A), R_1(B), Commit_1, W_3(B), R_3(B), W_3(A), Commit_3, R_2(C), Commit_2$.
2. $R_1(A), W_2(B), R_1(B), Commit_1, Commit_2$
3. $R_1(A), W_2(B), R_1(B), Commit_2, Commit_1$
4. $R_1(A), W_2(B), Commit_2, R_1(B), W_1(B)_1, Commit_1$

Question 3 Consider the following update.

UPDATE Part set price = price * 1.1
WHERE pid in (SELECT U.pid
                FROM Supply U, Supplier S
                WHERE S.sid = U.sid AND S.city = 'Toronto')

Consider what indices might be useful for the update in the previous question. Indicate yes (Y) or no (N).

_____ 1. Would an optimizer consider the use of a clustered hash index on Supply with key sid?

_____ 2. Would an optimizer consider the use of a clustered B+tree index on Part keyed on price?

_____ 3. Assuming no indices are available, then this update could cause a phantom.