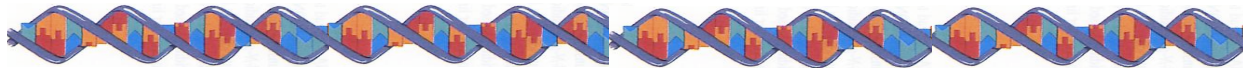


Problem Set 7

CS347,
11/24/09
V1.0



Assigned: 11/23/09

Due: 12/1/09 by midnight, email to the TA. stanv@cs.utexas.edu, place "CS347 PS7" in the subject line.

Objectives:

This homework walks you through each of the formula's concerning computing I/O and the size of a result of a relational operator, one case per question, largely in order of the presentation in class. i.e. There is a principle in each question that you are should know, but the actual method/calculation should not be that troublesome. Do observe that these questions, given their succinctness and lack of ambiguity make for good test questions.

Use the following database schema and associated catalog data for all the problems in this homework. If it is not otherwise specified you should assume the query comprises a natural join.

schema	R(a,b,c)	S(a)	U(b)	V(c)	W(b,c)
blocks	B(R)= 1000	B(S)=5	B(U)= 20	B(V)= 10	B(W) = 100
rows	T(R)= 1,000,000	T(S)=1,000	T(U)=10,000	T(V)= 100	T(W) = 10,000
values	V(R,a) = 100	V(S,a)= 100			
	V(R,b) = 200		V(U,b)=400		V(W,b)=20
	V(R,c) = 50			V(c)= 100	V(W,c)= 10

I/O Cost. which is a function of the number of blocks read.

- 1) In the linear cost model, what is the cost of reading the entire contents of relation R using a table scan?
- 2) Assume R is 100% clustered. (i.e. all the blocks are contiguous on disk) In the affine cost model, what is the cost of reading the entire contents of relation R using a table scan?
- 3) Assume c is the primary key of relation W.
 - a. Using the linear cost model, what is the cost of reading the data pages of W, (i.e. ignore the B+tree nodes), for the following query?
b. How many records do you estimate will be in the output?
Select *
From W
Where W.c = "Texas"
- 4) Assume (c,b,a) is a compound primary key for relation R,
 - a. Using the linear cost model, what is the cost of reading the data pages of R, (i.e. ignore the I/O for the B+tree nodes), for the following query?

b. How many records do you estimate will be in the output?

Select *

From R

Where R.a = 'Austin' AND R.b = 'Travis' And R.c = 'Texas'

5) Assume (c,b,a) is a compound primary key for relation R,

a. Using the linear cost model, what is the cost of reading the data pages of R, (i.e. ignore the I/O for the B+tree nodes, but do assume you have the benefit of the primary index), for the following query?

b. How many records do you estimate will be in the output?

Select *

From R

Where R.b = 'Travis' And R.c = 'Texas'

6) Assume that W has no useful index for the following question. How many rows do you estimate to be in the output of the following queries

a. Select *

From W

Where W.b = 'Travis' AND W.c = 'Texas'

b. Select *

From W

Where W.b = 'Travis' OR W.c = 'Texas'

Join Size Result: What is the size in, in rows, of the following join queries.

7) R |X| S

8) R |X| U

9) S |X| U

10) R |X| S |X| U |X| V

11) V |X| W

12) R |X| W

Think about:

Determine the optimal left deep join order for the query in question 10.