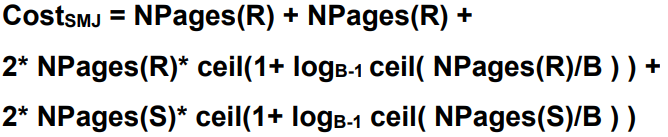
Question1

A = 10,000 Disk pages. B = 1,000 pages. R = 502 buffer pages. No index, simple heap

For each strategy, provide the formulae you use to calculate your cost estimates.

1. Page-oriented Nested Loops Join. Consider A as the outer relation.  
   Formula: Since A is outer relation, R = 10,000, S = 1000  
   So
2. Block-oriented Nested Loops Join. Consider A as the outer relation  
   Formula =

Since A is outer relation, R = 10,000, S = 1000, Block size B = 502  
So

1. Sort-Merge Join  
   Formula:  
   

R = 10,000; S = 1,000; B = 52;  
So, = 11000+2\*10,000\*2+ 2\*1,000\*2 = 55000

1. Hash Join

Formula:   
R = 10,000; S = 1,000;

So,

1. What would the lowest possible I/O cost be for joining A and B using any join algorithm and how much buffer space would be needed to achieve this cost. Explain briefly.

The optimal case would be when buffer size = 1000. So this is the case that S can be fits in memory, therefore only hashing will happen.

So,

Question 2

Consider a relation with the following schema:

Executives (id: integer, name:string, title:string, level: integer)

The Executives relation consists of 100,000 tuples stored in disk pages. The relation is stored as simple heap file and each page stores 100 tuples. There are 10 distinct titles in the Executives hierarchy and 20 distinct levels ranging from 0-20.

Suppose that the following SQL query is executed frequently using the given relation:

SELECT E.ename

FROM Executives

WHERE E.title = “CEO” and E.level > 15;

Your job is to analyze the query plans given below and estimate the cost of the best plan utilizing the information given about different indexes in each part.

Since there are 100,000 tuples and 100 tuples/page, we have 1,000 pages.

1. Compute the estimated result size and the reduction factor (selectivity) of this query  
   For E.title:  
    Formula:   
    So,

For E.level > 15:

Formula:

So,

Therefore,

1. Compute the estimated cost of the best plan assuming that a clustered B+ tree index on (title, level) is (the only index) available. Suppose there are 200 index pages, and the index uses Alternative 2. Discuss and calculate alternative plans

Formula:

So, in this case, we have

The Cost is

1. Compute the estimated cost of the best plan assuming that an unclustered B+ tree index on (level) is (the only index) available. Suppose there are 200 index pages, and the index uses Alternative 2. Discuss and calculate alternative plans.’

Formula:

So, in this case we have

The Cost is

Question2 Continue:

1. Compute the estimated cost of the best plan assuming that an unclustered Hash index on (title) is (the only index) available. The index uses Alternative 2. Discuss and calculate alternative plans

Formula:

So, in this case, we have

The cost is

1. Compute the estimated cost of the best plan assuming that an unclustered Hash index on (level) is (the only index) available. The index uses Alternative 2. Discuss and calculate alternative plans.

Formula

Since hashing cannot deal with range value, we have to calculate cost for finding each specific level, then add them up.

In this case,

So, for cost of search for one level, we have

So there are 4 levels that is >15, we need to sum up the cost:  
So

Therefore, the total cost is 55,000

Question 3

Consider the following relational schema and SQL query. The schema captures information about employees, departments, and company finances (organized on a per department basis)

Emp(eid: integer, did: integer, sal: integer, hobby: char(20))

Dept(did: integer, dname: char(20), floor: integer, phone: char(10))

Finance(did: integer, budget: real, sales: real, expenses: real)

Consider the following query:

SELECT D.dname, F.budget

FROM Emp E, Dept D, Finance F

WHERE E.did=D.did AND D.did=F.did

AND E.sal ≥ 59000 AND E.hobby = ‘yodeling’;

The system’s statistics indicate that employee salaries range from 10,000 to 60,000, and employees enjoy 200 different hobbies. There are a total of 50,000 employees and 5,000 departments (each with corresponding financial record in the Finance relation) in the database. Each relation fits 100 tuples in a page. Suppose there exists a clustered B+ tree index on (Emp.did) of size 50 pages.

1. Compute the estimated result size and the reduction factors (selectivity) of this query  
   For E.hobby:  
    Formula:   
    So,

For E.sal > 59000:

Formula:

So,

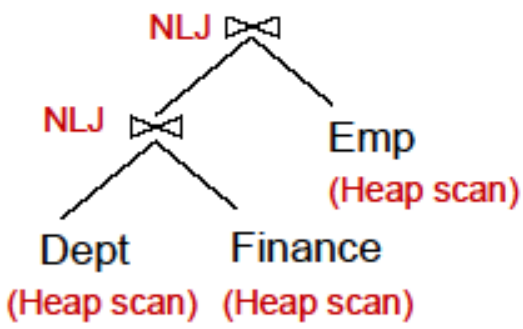
Therefore,

Since there are 50,000 employees, and 100 tuples in a pages, we have 500 pages record employees, So,

1. Compute the cost of the plans shown below. Assume that sorting of any relation (if required) can be done in 2 passes: 1 st pass to produce sorted runs and 2 nd pass to merge runs. Similarly hash join can be done in 2 passes: 1st pass to produce partitions, 2 nd pass to join corresponding partitions. NLJ is a Page-oriented Nested Loops Join. Assume that did is the candidate key, and that 100 tuples of a resulting join between Emp and Dept fit in a page. Similarly, 100 tuples of a resulting join between Finance and Dept fit in a page.

Question3 b) continue

We have

1. 

Cost of Scanning Dept = as it is Page Oriented

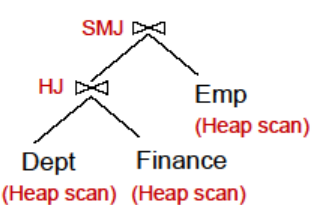
Cost to Join Dept with Finance using NLJ:

Result Size

Cost to Join (Dept Join Finance) with Emp:

As we already have result from Dept and Finance in memory.

So, total cost is

1. 

Cost of Hash Join Dept and Finance:

Formula:

In this case

So, assume memory cannot fit any relation entirely

Cost of Join this table with Emp:

Formula:

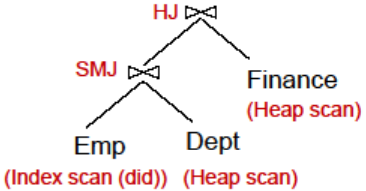
In this case,

Cost of Sorting the result Hashing table: 2\*50\*2 = 200

Cost of Sorting the Emp: 2\*300\*2 = 1200

Cost of Joining EMP to existing Hashing table = 500, as we already read existing table

So, Total Cost is

1. 

In this case, we have:

Cost of Joining Emp with Dept:

Cost of Sorting Emp = 500 + 50 = 550, as Emp is indexed

Cost of Sorting Dept = 50 \* 2 \* 2 = 200, as #pass = 2

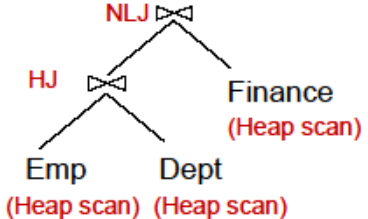
Cost of Joining this two is 550 + 50 = 600

Total cost of SMJ between Emp and Dept = 550 + 200 +600 =1350

Result pages size of Emp Joing Dept: = 500, Npages(Finance) = 50

Cost to Hash Join with Finance: 2\*500+3\*50 = 1150

So, the total cost is: 1150 + 1350 = 2500

1. 

This case, we have   
Cost of Hash Join Emp and Dept:

3\*(500 + 50) = 1650.

Result Page Size = 500.

Cost of NLJ Between result page and Finance:

500\*50 = 25000, as Result table has already read

Total Cost of operation is:

1650 + 25000 = 26650.