Assignment 3 – INFO20003 Semester 2 2018

Assignment 3: Query Processing and Query Optimization

Due: 6pm Friday 5th of October 2018

Submission: Via LMS https://lms.unimelb.edu.au

Weighting: 10% of your total assessment. The assignment will be graded out of 20 marks.

Question 1 (5 marks)

Consider two relations called Invoice and Customers. Imagine that the Customers relation has 1,000 pages and the Invoice relation has 5,000 pages. Consider the following SQL statement:

```
SELECT *
FROM Customers INNER JOIN Invoice
  ON Customers.Cust_ID = Invoice.Cust_ID;
```

Assignment Project Exam Help

We wish to evaluate an equijoin between Customers and Invoice, with an equality condition Customers.Cust_ID = Invoice.Cust_ID. There are 502 buffer pages available in memory for this operation. Both relations are stored as (unsorted) heap files. Neither relation has any indexes built on it.

https://powcoder.com

Add WeChat powcoder

Consider the alternative join strategies described below and calculate the cost of each alternative. Evaluate the algorithms using the number of disk I/O's (i.e. pages) as the cost. For each strategy, provide the formulae you use to calculate your cost estimates.

a) Page-oriented nested loops join. Consider Customers as the outer relation. **(1 mark)**
b) Block-oriented nested loops join. Consider Customers as the outer relation. **(1 mark)**
c) Sort-Merge join. Assume that Sort-Merge Join can be done in 2 passes. **(1 mark)**
d) Hash Join **(1 mark)**
e) What would be the lowest possible cost to perform this query, assuming that no indexes are built on any of the two relations, and assuming that sufficient buffer space is available? What would be the minimum buffer size required to achieve this cost? Explain briefly. **(1 mark)**

Question 2 (5 marks)

Consider a relation called MobilePhones that stores information about purchased phones. Imagine that the relation MobilePhones consists of 1,000 data pages and each page stores 100 tuples. Imagine that the *amount* attribute can take any value between 0 and 1000 ([0,1000]), and imagine that *discount* can take any value between 1 and 100 ([1,100]). Suppose that the following SQL query is executed frequently using the given relation:

```
SELECT *
FROM MobilePhones
WHERE amount > 500 AND discount = 20;
```

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

a) Compute the estimated result size for the query and the reduction factor of each filter. **(1 mark)**

b) Compute the estimated cost of the *best plan* assuming that a *clustered B+ tree* index on *(discount, amount)* is (the only index) available. Suppose there are 400 index pages. Discuss and calculate alternative plans. **(1 mark)**

c) Compute the estimated cost of the *best plan* assuming that an *unclustered B+ tree* index on *(amount)* is (the only index) available. Suppose there are 200 index pages. Discuss and calculate alternative plans. **(1 mark)**

d) Compute the estimated cost of the *best plan* assuming that an *unclustered Hash* index on *(discount)* is (the only index) available. Discuss and calculate alternative plans. **(1 mark)**

e) Compute the estimated cost of the *best plan* assuming that an *unclustered Hash* index on *(amount)* is (the only index) available. Discuss and calculate alternative plans. **(1 mark)**

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Question 3 (10 marks)

Consider the following relational schema and SQL query. The schema captures information about employees, departments, and company finances. The finance sector is organized on a per-department basis, i.e. each department has its own financial budget (and thus a corresponding record in the Finance relation).

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: double, sales: double, expenses: double)

Consider the following query:

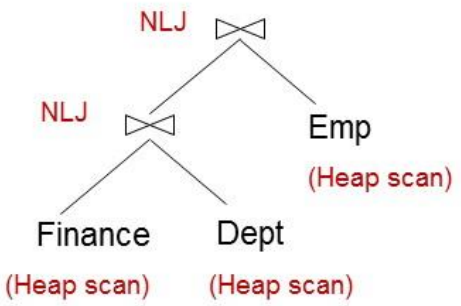
```
SELECT D.dname, F.budget
FROM Emp E, Dept D, Finance F
WHERE E.did = D.did AND D.did = F.did
  AND E.sal >= 90000 AND E.hobby = 'skydiving';
```

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

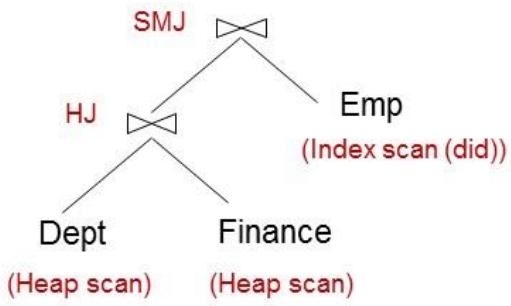
The system's statistics indicate that employee salaries range from 50,000 to 100,000, and employees enjoy 100 different hobbies. There are a total of 100,000 employees and 6,000 departments in the database. Each department has a corresponding financial record in the Finance relation. Each relation fits 100 tuples in a page. Suppose there exists a *clustered B+ tree* index on *(Emp.did)* of size 100 pages. Since selection over filtering predicates is not marked in the plans, assume it will happen on-the-fly *after* all joins are performed (as the last operation in the plan).

a) Compute the estimated result size and the reduction factors (selectivity) of this query. **(2 marks)**
b) Compute the cost of the plans shown below. Assume that sorting of any relation (if required) can be done in *2 passes.* NLJ is a *Page-oriented* Nested Loops Join. Assume that *did* is the candidate key of the Dept relation, 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. **(8 marks, 2 marks per plan)**

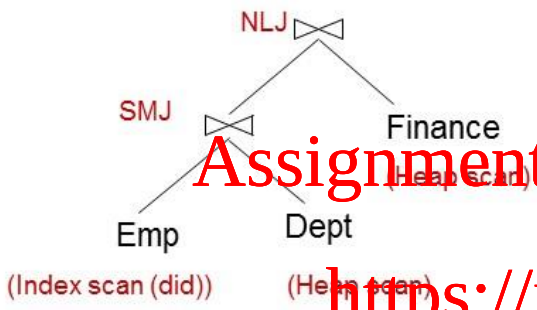
1)



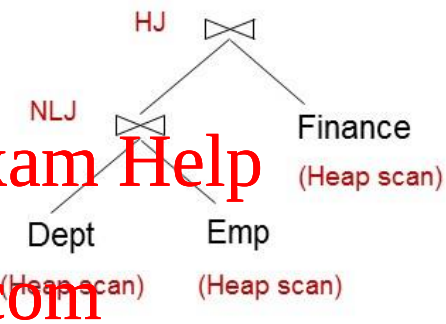
2)



3)



4)



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder