

CSE 4331/5331 Sample Test 3 Questions

Sharma Chakravarthy

Hoping to improve the performance, I am posting this sample test 2 questions. Hope it will improve your preparations for the test. [n] indicates points assigned.

1) Map/Reduce, cloud computing, Big Data Sample questions

[5 + 5 + 5]

We have a relation with the following attributes: stud id, department, gender, country, and cumulative GPA (cgpa). We want to compute the average GPA of all students in each department and by gender. This is equivalent to a GROUP BY on department and gender and average computation in SQL. We want to project department, gender, and AVG(cgpa) as output. Each tuple (record) in the input contains values for the above attributes separated by a “;”.

Write mapper, reducer, and combiner (if needed) **pseudo code** to

- i) Show what the mapper should output when it processes each record. Show the key, value pairs output by the mapper clearly and explain why
- ii) Indicate what the reducer will compute, how, and output the result.
- iii) Is there a need for a combiner for this problem? If so, please show what the combiner needs to do to the output generated by the mapper and why it is beneficial. If a combiner is not needed, explain/justify why

(6 + [4 + 5 + 5])

2) Query Processing sample questions

You are given the following schema (italicized attributes are keys)

EMP(*eid*: integer, did: integer, sal: integer, hobby: char(20))

DEPT(*did*: integer, dname: char(20), floor: integer, phone: char(20))

FINANCE(*did*: integer, budget: real, sales: real, expenses: real)

And the query along with the following information:

```
SELECT      D.dname, F.budget
FROM        EMP E, DEPT D, FINANCE F
WHERE       E.did=D.did and D.did = F.did
```

And D.floor=1 and E.sal >= 59000 and E.hobby = 'yodelling'

Unclustered B+ tree indexes exist on EMP.did, EMP.sal, DEPT.floor, DEPT.did, and FINANCE.did. The system's statistics indicate that employee salaries range from 10000 to 60000, employees enjoy 200 different hobbies, and the company owns 2 floors in the building. There are a total of 50000 employees and 5000 departments (each with corresponding financial information) in the database. The DBMS used by the company supports nest loop join and index nested loops

- i) For each of the query's base relations (EMP, DEPT and FINANCE) estimate the number of tuples that would be initially selected from the relation if all of the **non-join predicates** on that relation were applied to it before any join processing begins. In other words, compute the reduction factor for EMP, DEPT, and FINANCE relations
- ii) Using the cardinality (size) of the resulting relations from above: (a) determine the join order and for that (b and c) show **two different physical plans** for computing the above query including join algorithms for each join (e.g., scan relation or retrieve using B+ tree index on which attribute, etc.) using the index and other information given.