

Tutorial 10: Concurrency Control

CS3402 Database Systems

Question 1

- Consider the following table scheme, and assume that R_1 has 1000 tuples, R_2 has 1500 tuples and R_3 has 750 tuples.
- $R_1 (\underline{A}, B, C)$
 - $R_2 (\underline{C}, D, E)$
 - $R_3 (\underline{E}, F)$
- a) Estimate the size (max and min numbers of tuples) of $R_1 * R_2 * R_3$ (where $*$ denotes Natural Join).
- b) There are two ways to perform the Natural Join in (a), which one is more efficient in terms of number of comparison?
- $(R_1 * R_2) * R_3$
 - $R_1 * (R_2 * R_3)$

Question 2 (1/2)

- A canonical query tree is a tree structure that corresponds to a relational algebra expression or an SQL query directly, without doing any optimization. As such, it is usually not the most efficient way of executing the query.
- Consider the relations:

EMPLOYEE(ENAME, SSN, BDATE, ADDRESS, DNUM)

PROJECT(PNAME, PNUMBER, PLOCATION, DNUM)

WORKS_ON(ESSN, PNO, HOURS)

Question 2 (2/2)

- And the following SQL query:

```
SELECT      ENAME
FROM        EMPLOYEE, WORKS_ON, PROJECT
WHERE       PNAME="HeavenRay" AND PNUMBER=PNO
           AND ESSN=SSN AND BDATE > 'OCT-11-1966';
```

- a) Draw a canonical query tree for the above SQL query.
- b) Apply the optimization rules to the above query tree and come up with the most optimized query tree.