CS3402 Tutorial 7:

1. Consider the following table scheme, and assume that R1 has 1000 tuples, R2 has 1500 tuples and R3 has 750 tuples.

**R1** (*A, B, C*)

**R2** (*C*, D, E)

**R3** (*E, F*)

1. Estimate the size (max and min of tuples) of R1 \* R2 \* R3 (where \* denotes Natural Join).
2. There are two ways to perform the Natural Join in (a), which one is more efficient in terms of number of comparison?
   1. (R1 \* R2) \* R3
   2. R1 \* (R2 \* R3)
3. 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)

as well as 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';

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