

**Final Exam in Distributed Database**

Name: .....

**Exercise 01: (10 pts)**

EMPLOYEE (ENR, ENAME, JOB, SALARY)

PROJECT (PNR, PNAME, BUDGET)

ASSIGNMENT (ENR, PNR, DURATION)

The following relation schema is given:

Assume that the relation EMPLOYEE is to be fragmented horizontally according to the attribute ENR to as follow: Site 1) ENR < 150, Site 2) 150 <= ENR < 300, Site 3) ENR >= 300

Q1) How to fragment the relation EMPLOYEE?

Q2) Given that the relation ASSIGNMENT is to be fragment to three fragments so that the fragments should be also stored at Site1, Site 2, and Site 3. How to do that?

Q3) Based on EMPLOYEE and ASSIGNMENT fragmentations, give the optimized execution plan of the query:

SELECT ENAME, DURATION FROM EMPLOYEE E, ASSIGNMENT A WHERE E.ENR = A.ENR

**THE ANSWER**

A1) (3 pts)

$$EMPLOYEE1 = \sigma_{ENR < 150} (EMPLOYEE)$$

$$EMPLOYEE2 = \sigma_{ENR \geq 150 \wedge ENR < 300} (EMPLOYEE)$$

$$EMPLOYEE3 = \sigma_{ENR \geq 300} (EMPLOYEE)$$

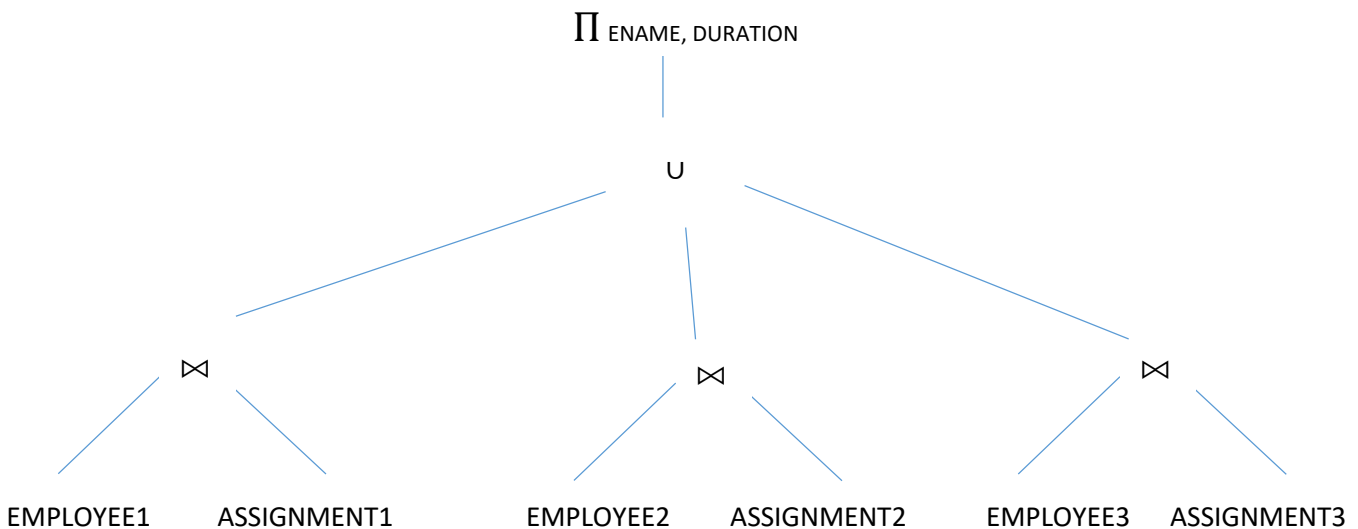
A2) (3 pts)

$$ASSIGNMENT1 = ASSIGNMENT \bowtie EMPLOYEE1$$

$$ASSIGNMENT2 = ASSIGNMENT \bowtie EMPLOYEE2$$

$$ASSIGNMENT3 = ASSIGNMENT \bowtie EMPLOYEE3$$

A3) (4 pts)



**Exercise 02 (10 pts):**

EMPLOYEE (ENR, ENAME, JOB, SALARY)

PROJECT (PNR, PNAME, BUDGET)

ASSIGNMENT (ENR, PNR, DURATION)

Given the following schema:

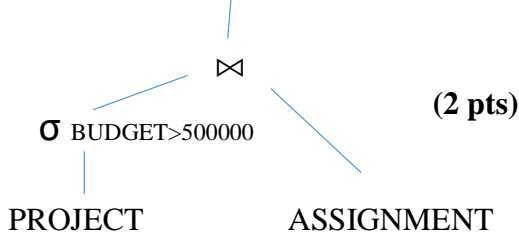
Each of the three relations is stored on a different node. Furthermore, the following statistics are known:  
card(ASSIGNMENT)=1.500, card(PROJECT)=200.

SELECT \* FROM PROJECT P, ASSIGNMENT A  
WHERE P.PNR=A.PNR AND P.BUDGET>500000

Q1) Estimate the best query communication cost for the optimized plan, given that:

- the query is initiated at node PROJECT and the result must be returned there.
- 20% of the projects have BUDGET>500000
- 25% of the projects are not assigned to any employee (do not have entries in the relation ASSIGNMENT)

**THE ANSWER**



**Analysis (2 pts)**

Card ( $\sigma$  BUDGET>500000 (PROJECT)) = 200 \* 0.20 = 40

Assigned projects are 75% of total projects = 200 \* 0.75 = 150

⇒ on average 1500 assignments for 150 projects

⇒ 10 assignments/project

So, card( $\sigma$  BUDGET>500000 (PROJECT) ⋈ ASSIGNMENT) = 40 \* 10 = 400

**The best Cost (6 pts)**

1) Join in PROJECT Node: N ASSIGNMENT  $\xrightarrow{1500}$  N PROJECT (Cost = 1500)

2) Join in PROJECT Node: N PROJECT  $\xrightarrow{40}$  N ASSIGNMENT  $\xrightarrow{400}$  N PROJECT (Cost = 440)

**So the the best plan is (2)**