ADBMS Assignment

Name: Kapil Kumar Chhipa
Roll No: M1B0265CA

19.10

Q 1 Consider a relation that is fragmented horizontally by plant_number.

employee (name, address, salary, plant_number)

Assume that each fragment has two replicas: one stored at the New York site and one stored locally at the plant site. Describe a good Processing strategy for the following queries entered at the San Joe site.

(a) Find all employees at the Boca plant.

Sol (i) Send the query Π_{name}(employee) to the Boca plant.
(ii) Have the Boca location send back the answer.

(b) Find the average salary of all employees.

Sol (i) Compute average at New York.
(ii) Send Answer to San Jose.

(c) Find the highest paid employee at each of the following sites: Toronto, Edmonton, Vancouver, Montreal

Sol (i) Send the query to find the highest salaried employee to Toronto, Edmonton, Vancouver, and Montreal.
(ii) Compute the queries at those site
(iii) Return Answer to San Jose.

Scanned with CamScanner

(d) Find the lowest-paid employee in the Company.

Sol (i) Send the query to find the lowest salaried employee to New York.

(ii) Compute the query at New York.

(iii) Send answer to San Jose.

@ 19.23

Consider the relation:

Employee (name, address, salary, plant_number)

Machine (machine_number, type, plant_number)

Assume that the Employee relation is fragmented horizontally by plant_number and each fragment is stored locally at its corresponding plant site.

Assume that the machine relation is stored in its entirely at the Armonk site. Describe a good strategy for processing each of the following queries.

a) Find all employees at the plant that contains machine number 1130.

Sol (i) Perform $\Pi_{plant\_number}(\sigma_{machine\_number = 1130}(machine))$ at Armonk.

(ii) Send the query $\Pi_{name}(Employee)$ to all site which are in the result of the previous query.

(iii) Those site Compute the answers.

(iv) Union the answers at the destinations site.

Scanned with CamScanner

b) Find all employees at plants that contain machines whose type is "milling machine".

Sol The strategy is the same as (a) except the first step should be to perform $\Pi_{plant\text{-}number}$

$(\sigma_{type = 'milling\ machine'}(machine))$ at Armonk.

c) Find all machines at the Almaden plant.

Sol (i) Perform $\sigma_{plant\text{-}number = x}(machine)$ at Armonk, where x is the plant number for almaden.

(ii) Send the answers to the destination site.

d) Find Employee M machine

Sol Strategy 1

(i) Group machine at Armonk by plant number.

(ii) Send the groups to the sites with the corresponding plant number.

(iii) Perform a local join between the local data and the remaining data.

(iv) Union the results at the destination site.

Strategy 2

Send the machine relation at Armonk, and all the fragments of the employee relation to the destination site. Then perform the join at the destination site.

There is parallelism in the join computation according to the final strategy but not in the second. Nevertheless, in a WAN the amount of data to be shipped is the main cost factor.

Scanned with CamScanner

we Expect that each plant will have more than one machine, hence the result of the local join at each site will be a Cross Product of the employee tuples and machines at that plant.

This Cross-Product's size is greater than the size of the employee fragment at that site. As a result the second strategy will result in less data shipping and will be more efficient.

Scanned with CamScanner