Principles of Distributed and Parallel Database Systems

Introduction to Fragmentation

Ira A. Fulton Schools of Engineering
Arizona State University

Objectives

Objective

Understand data fragmentation
& replication models

Fragmentation Alternatives: Horizontal

PROJ

PNO	PNAME	BUDGET	LOC
P1	Instrumentation	150000	Montreal
P2	Database Develop.	135000	New York
P3	CAD/CAM	250000	New York
P4	Maintenance	310000	Paris
P5	CAD/CAM	500000	Boston

PROJ 1: projects with budgets less than $200,000

PNO	PNAME	BUDGET	LOC
P1	Instrumentation	150000	Montreal
P2	Database Develop.	135000	New York

PROJ 2: projects with budgets greater than or equal to $200,000

PNO	PNAME	BUDGET	LOC
P3	CAD/CAM	250000	New York
P4	Maintenance	310000	Paris
P5	CAD/CAM	500000	Boston

Fragmentation Alternatives: Vertical

PROJ

PNO	PNAME	BUDGET	LOC
P1	Instrumentation	150000	Montreal
P2	Database Develop.	135000	New York
P3	CAD/CAM	250000	New York
P4	Maintenance	310000	Paris
P5	CAD/CAM	500000	Boston

PROJ 1: information about project budgets

PNO	BUDGET
P1	150000
P2	135000
P3	250000
P4	310000
P5	500000

PROJ 2: information about project names and locations

PNO	PNAME	LOC
P1	Instrumentation	Montreal
P2	Database Develop.	New York
P3	CAD/CAM	New York
P4	Maintenance	Paris
P5	CAD/CAM	Boston

Which are good fragmentations?

$\mathbf{F} = \{ F_1, F_2 \}$

$F_1 = \sigma_{sal<10} E$ $\quad\quad$ $F_2 = \sigma_{sal>20} E$

Some tuples lost

Which are good fragmentations?

F = { F3, F4 }

F3 = σ sal<10 E F4 = σ sal>5 E

Tuples with 5 < sal < 10 are duplicated

Better Replication

$$\mathbf{F} = \{ F_5, F_6, F_7 \}$$

$$F_5 = \sigma_{sal \leq 5} E \qquad F_6 = \sigma_{5 < sal < 10} E \qquad F_7 = \sigma_{sal \geq 10} E$$

Then replicate F6 if convenient (part of allocation problem)

Desired Properties of Fragmentation

Completeness

- Decomposition of relation R into fragments R_1, R_2, ..., R_n is complete if and only if each data item in R can also be found in some R_i

Reconstruction

- If relation R is decomposed into fragments R_1, R_2, ..., R_n, then there should exist some relational operator ∇ such that
  - $R = \nabla_{1 \le i \le n} R_i$

Disjointness

- If relation R is decomposed into fragments R_1, R_2, ..., R_n, and data item d_i is in R_j, then d_i should not be in any other fragment R_k ($k \neq j$).