Principles of Distributed and Parallel Database Systems Introduction to Fragmentation



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
Р3	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
Р3	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
Р3	CAD/CAM	250000	New York
P4	Maintenance	310000	Paris
P5	CAD/CAM	500000	Boston

PROJ 1: information about project budgets

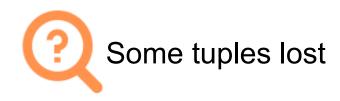
PNO	BUDGET
P1	150000
P2	135000
Р3	250000
P4	310000
P5	500000

PROJ 2: information about project names and locations

PNO	PNAME	LOC
P1	Instrumentation	Montreal
P2	Database Develop.	New York
Р3	CAD/CAM	New York
P4	Maintenance	Paris
P5	CAD/CAM	Boston

Which are good fragmentations?

$$F = \{ F_1, F_2 \}$$
 $F_1 = \mathbf{O}_{sal<10} E$
 $F_2 = \mathbf{O}_{sal>20} E$



Which are good fragmentations?

$$F = \{ F3, F4 \}$$
 $F3 = \mathbf{O} \text{ sal} < 10 \text{ E}$
 $F4 = \mathbf{O} \text{ sal} > 5 \text{ E}$

Better Replication

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

$$F_5 = \mathbf{O}_{sal \le 5} E$$

$$F_5 = \mathbf{O}_{sal \le 5} E$$
 $F_6 = \mathbf{O}_{5 \le sal \le 10} E$ $F_7 = \mathbf{O}_{sal \ge 10} E$



Then replicate F6 if convenient (part of allocation problem)

Desired Properties of Fragmentation

Completeness

Decomposition of relation R into fragments R₁, R₂, ..., 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 R₁, R₂, ...,
 Rₙ, 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 R₁, R₂, ..., Rₙ, and data item dᵢ is in Rᵢ, then dᵢ should not be in any other fragment Rᵢ (k ≠ j).