

# Tutorial 1: Distributed Database Design

*Semester 1, 2004*

## Correctness rules of fragmentation:

- (a) *Completeness.* If a relation instance  $R$  is decomposed into fragments  $R_1, R_2, \dots, R_n$ , each data item that can be found in  $R$  can also be found in one or more  $R_i$ 's. This property, which is identical to the lossless decomposition property of normalisation, is also important in fragmentation since it ensures that the data in a global relation is mapped into fragments without any loss. Note that in the case of horizontal fragmentation, the "item" typically refers to a tuple, while in the case of vertical fragmentation, it refers to an attribute.
- (b) *Reconstruction.* If a relation  $R$  is decomposed into fragments  $R_1, R_2, \dots, R_n$ , it should be possible to define a relational operator  $\nabla$  such that

$$R = \nabla R_i, \forall R_i$$

The operator  $\nabla$  will be different for the different forms of fragmentation; it is important, however, that it can be identified. The reconstructability of the relation from its fragments ensures that constraints defined on the data in the form of dependencies are preserved.

- (c) *Disjointness.* If a relation  $R$  is horizontally decomposed into fragments  $R_1, R_2, \dots, R_n$ , and data item  $d_i$  is in  $R_j$ , it is not in any other fragment  $R_k$  ( $k \neq j$ ). This criterion ensures that the horizontal fragments are disjoint. If relation  $R$  is vertically decomposed, its primary key attributes are typically repeated in all its fragments. Therefore, in case of vertical partitioning, disjointness is defined only on the nonprimary attributes of a relation.

**Question 1:**

Given the following relation and the predicates p1: SAL > 30000, p2: SAL < 30000

ID	NAME	SAL
1289	John	12000
8907	Sally	67050
7643	Elvin	51980
0988	Kelly	30000
6543	Emily	28760
0986	Robert	54000
2345	Thomas	29999

- a) Perform a horizontal fragmentation of the table based on the given predicates.
- b) Is this a correct fragmentation?
- c) If the answer to (b) is no, explain why, and give the predicates that would correctly fragment the table

**Question 2:** A Student table is given as follows

ST-NO	NAME	AGE	DEGREE
89-087	Nadine	23	B.Sc
45-867	Sarah	24	M.Sc
12-098	Zhou	21	B.Sc
34-765	Kamil	30	M.Sc
09-008	Jenny	29	PhD
87-124	Sonya	19	B.Sc
09-546	Stephan	45	PhD

Are any of the following vertical fragments of the student table incorrectly constructed? If so, what is the problem?

### Fragment 1

ST-NO	NAME	AGE
89-087	Nadine	23
45-867	Sarah	24
12-098	Zhou	21
34-765	Kamil	30
09-008	Jenny	29
87-124	Sonya	19
09-546	Stephan	45

### Fragment 2

NAME	AGE	DEGREE
Nadine	23	B.Sc
Sarah	24	M.Sc
Zhou	21	B.Sc
Kamil	30	M.Sc
Jenny	29	PhD
Sonya	19	B.Sc
Stephan	45	PhD

### Fragment 3

ST-NO	NAME	AGE	DEGREE
89-087	Nadine	23	B.Sc
45-867	Sarah	24	M.Sc
12-098	Zhou	21	B.Sc
09-546	Stephan	45	PhD

**Question 3:** The following global schema is given:

$$R_1 = \overline{ABCD}, R_2 = \overline{DEFG}, R_3 = \overline{FGHIJ}$$

- (a) Design a correct vertical data fragmentation of the above schema suitable for distributed databases with two sites.
- (b) Identify all referential integrity constraints that must be maintained on each site of the distributed system and between sites of the system.
- (c) Assume the following is the database population:

$$D = \{r_1, r_2, r_3\}$$

$r_1(A B C D)$	$r_2(D E F G)$	$r_3(F G H I J)$
1 2 1 1	1 2 3 1	3 1 0 2 0
3 4 2 2	2 4 3 2	3 2 6 1 0
3 2 3 2	3 1 3 2	2 2 1 1 1

Compute the value of the following expression:

$$\Pi_{EFH} (r_2 \bowtie r_3)$$

in the centralized database and then in your designed distributed version.  
(Show complete algebraic expression required for distributed computation).

**Question 4:** The following table  $r(\overline{A B C D})$  is given. Let  $R_1, R_2, R_3, R_4, R_5$  be a fragmentation.

- (a) Check its correctness

$$R_1 = \pi_{AB} \sigma_{A \geq 2} (r)$$

$$R_2 = \sigma_{A < 2} \pi_{AB} (r)$$

$$R_3 = \pi_{ACD} \sigma_{C < 5} (r)$$

$$R_4 = \pi_{AC} \sigma_{C \geq 5} (r)$$

$$R_5 = \pi_{AC} \sigma_{C < 5} (r)$$

- (b) Is there any data replication in such design?