

Tutorial 1: Distributed Database Design

Semester 1, 2004

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 |

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

Solutions:

- Horizontal fragmentation based on the given predicates

Fragment 1

| ID | NAME | SAL |
|------|--------|-------|
| 8907 | Sally | 67050 |
| 7047 | Elvin | 51980 |
| 0986 | Robert | 54000 |

Fragment 2

| ID | NAME | SAL |
|------|--------|-------|
| 1289 | John | 12000 |
| 6543 | Emily | 28760 |
| 2345 | Thomas | 29999 |

- This is **not** a correct fragmentation.
- When we fragment the relation using the given predicates we end up excluding one tuple, which violates the *completeness* rule for correct fragmentation. Correct predicates: p1: SAL ≥ 30000, p2: SAL < 30000 or P1: SAL > 30000, p2: SAL ≤ 30000

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 |

Solutions: F1 is a correct vertical fragment. F2 is not a correct vertical fragment, as it does not contain the primary key attribute, ST-NO, of the original relation. F3 is a not correct vertical fragment at all (it is a horizontal fragment).

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 |

Computer the value of the following express:

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

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

Solutions:

- (a) Below is one (of many) vertical data fragmentation schema to distribute for two sites.

$$r_{11} = ABC, r_{12} = ABD$$

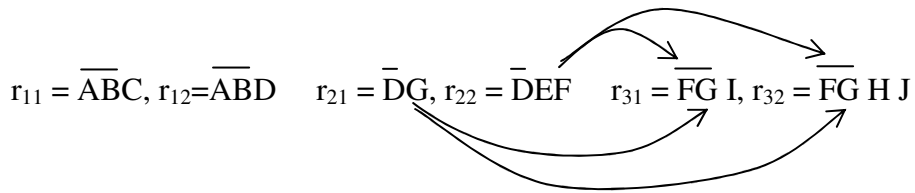
$$r_{21} = DG, r_{22} = DEF$$

$$r_{31} = FGI, r_{32} = FGHIJ$$

- (b) Referential integrity constraints are:

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

These referential integrity constraints should be maintained in the distributed database too:



(c) To perform the following query

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

when the database is not fragmented, the result is

| $r_2(DEF G)$ | $r_3(FGH I J)$ | $\rightarrow DEFGHIJ$ | $\rightarrow EFH$ |
|--------------|----------------|-----------------------|-------------------|
| 1 2 3 1 | 3 1 0 2 0 | 1 2 3 1 0 2 0 | 2 3 0 |
| 2 4 3 2 | 3 2 6 1 0 | 2 4 3 2 6 1 0 | 4 3 6 |
| 3 1 3 2 | 2 2 1 1 1 | 3 1 3 2 6 1 0 | 1 3 6 |

If r_{11} is on site 1 and r_{12} is on site 2, site 2 has all the attributes required by this query ($r_{22}(DEF)$ and $r_{32}(FGHIJ)$).

| $r_{22}(DEF)$ | $r_{32}(FGHIJ)$ | $\rightarrow DEFGHIJ$ | $\rightarrow EFH$ |
|---------------|-----------------|-----------------------|-------------------|
| 1 2 3 | 3 1 0 0 | 1 2 3 1 0 0 | 2 3 0 |
| 2 4 3 | 3 2 6 0 | 1 2 3 2 6 0 | 2 3 6 |
| 3 1 3 | 2 2 1 1 | 2 4 3 1 0 0 | 4 3 0 |
| | | 2 4 3 2 6 0 | 4 3 6 |
| | | 3 1 3 1 0 0 | 1 3 0 |
| | | 3 1 3 2 6 0 | 1 3 6 |

The result here is not correct (it's a superset of the correct result). The problem is caused by violation of distributed referential integrity constraints between r_2 and r_3 .

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

(a) Check its correctness

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

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

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

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

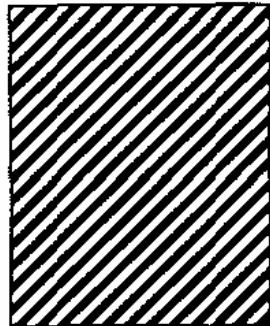
$$R_5 = \pi_{AC} \sigma_{C < 5}(r)$$

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

Solutions:

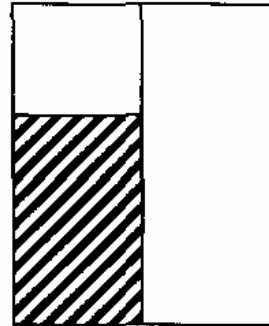
(a) We will illustrate each of the fragments R_1, \dots, R_5 first:

r (A B C D)



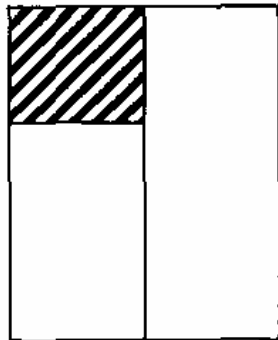
r₁ (A B C D)

$A \geq 2$



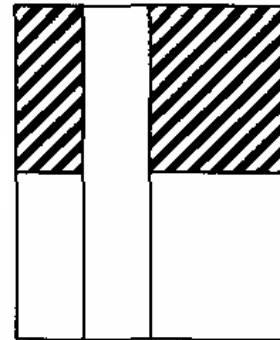
r₂ (A B C D)

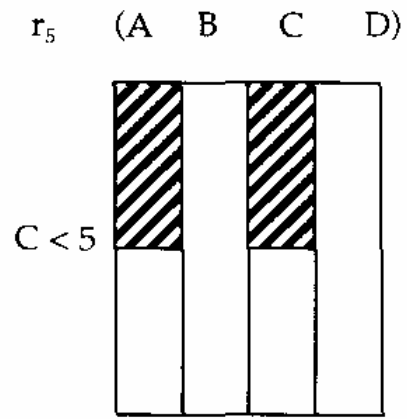
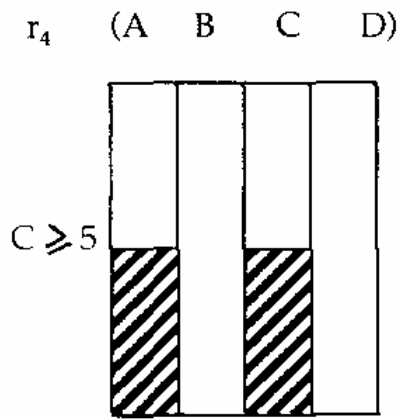
$A < 2$



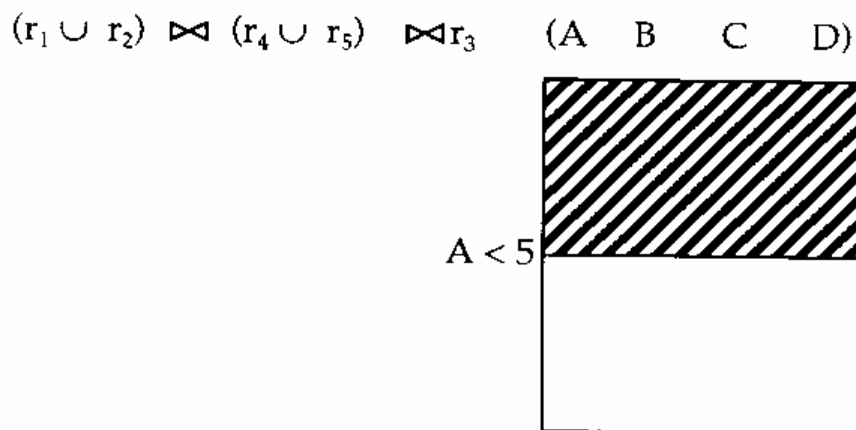
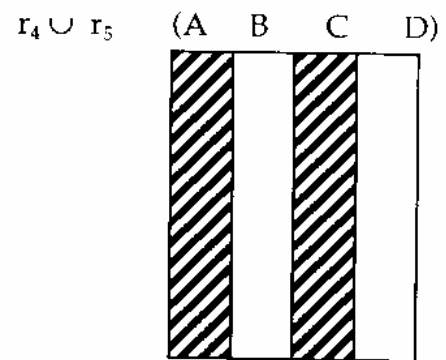
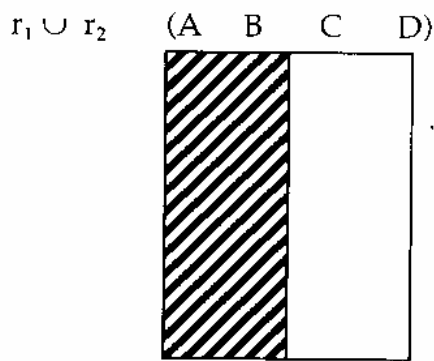
r₃ (A B C D)

$C < 5$





Clearly a tuple $t = (\bullet, \bullet, \bullet, 7)$ is not represented after reconstruction.



Finally, the proposed fragmentation is incomplete, thus incorrect.

There is substantial data replication.