

Tutorial 1: Distributed Database Design

INFS3200/7907
Advanced Database Systems

1

Announcement

- The assignment is available on course website.
- It is optional for INFS3200.
- However, you are encouraged to attempt this as an exercise.
- You can also undertake it formally, in which case your mark distribution will be same as for INFS7907 students,
 - i.e. 10%Assignment+30%Mid Semester Exam + 60%Final Exam

2

Fragmentation

- What is fragmentation?
 - A fragment is subset of a relation.
- Why fragmentation?
 - Application views are subset of relations, e.g. 80-20 rule
 - If there is no fragmentation
 - An entire table is at one site – high volume of remote data access
 - An entire table is replicated – problems in executing updates
 - Parallel execution of query, since sub-queries operate on fragments – higher concurrency (better performance)

3

Fragmentation

Types of Fragmentation

- Horizontal Fragmentation
- Vertical Fragmentation
- Hybrid Fragmentation

4

Types of Fragmentation

- Horizontal Fragmentation
 - Characteristics
 - A relation (table) is divided along tuples.
 - The fragments are mutually exclusive/disjointed.
 - How?
 - SELECT Operation in Relational Algebra

5

Types of Fragmentation

- Vertical Fragmentation
 - Characteristics
 - A relation is partitioned with attributes.
 - Each fragment must contain a primary key that is used as a foreign key for the reconstruction using JOIN operation.
 - How?
 - Project Operation in Relational Algebra

6

Types of Fragmentation

- Hybrid fragmentation
 - The combination of Horizontal and Vertical Fragmentation

7

Fragmentation

- What is **good/correct** fragmentation?
 - The fragmentation process must comply with all of the **Correctness rules**.

8

Fragmentation

- Correctness rules For Fragmentation
 - Completeness
 - Reconstruction
 - Disjointness

9

Fragmentation

- Correctness rules For Fragmentation
 - Completeness
 - Each data item found in a relation will be found in one or more of the relation's fragments.

10

Fragmentation

- Correctness rules For Fragmentation
 - Reconstruction
 - The dependency constraints on the data can be preserved by ensuring that every relation can be **reconstructed** from its fragments using a relational (SQL) operation.
 - Horizontal fragmentation - **Union**
 - Vertical fragmentation - **Join** (by primary keys)

11

Fragmentation

- Correctness rules For Fragmentation
 - Disjointness
 - No redundant data is allowed among the fragments of a relation.
 - Horizontal fragmentation - a row cannot appear in more than a fragment.
 - Vertical fragmentation
 - non-key attributes cannot appear in more than a fragment.
 - Disjointness does not apply to the primary key attributes.

12

Referential Integrity

- Following these rules can get the fragmentation right.
- However, it doesn't guarantee the correctness of query results based on fragments if the referential integrity is broken inadvertently during fragmentation(Q3).

13

Question 1 - Horizontal Fragmentation

Question: Given the following relation and the predicates
p1: Pos < 4, p2: Pos > 4

EventID	CompID	Pos
SwM100Free	Thorpe	3
SwM100Free	Hoogenband	1
SwM100Free	Schoman	2
SwM100Free	Iles	7
ShotputW	Cumba	1
ShotputW	Ostapchuk	4
ShotputW	Li	9

14

Q1a

Perform a horizontal fragmentation of the table based on the given predicates

Solution for performing horizontal fragments with the given predicates:

EventID	CompID	Pos	EventID	CompID	Pos
SwM100Free	Thorpe	3	SwM100Free	Iles	7
SwM100Free	Hoogenband	1	ShotputW	Li	9
SwM100Free	Schoman	2			
ShotputW	Cumba	1			

Fragment 1: Pos < 4

Fragment 2: Pos > 4

15

Q1b

Is this a correct fragmentation?

They are **INCORRECT** fragments

- Completeness, one of the correctness rules of Fragmentation, is violated.
 - Tuple (ShotputW, Ostapchuk, 4) disappears.

16

Q1c

If the answer to (Q1b) is no, explain why, and give the predicates that would correctly fragment the table

The **CORRECT** predicates are either:-

- p1: Pos ≥ 4, p2 Pos < 4 OR
- p1: Pos > 4, p2 Pos ≤ 4

17

Question 2 - Vertical Fragmentation

A Results table is given as follows

EventID	CompID	Pos	Record?
SwM100Free	Thorpe	3	NA
SwM100Free	Hoogenband	1	No
SwM100Free	Schoeman	2	NA
SwM100Free	Iles	7	NA
ShotputW	Cumba	1	No
ShotputW	Ostapchuk	4	NA
ShotputW	Li	9	NA

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

18

Question 2 (Continue)

EventID	CompID	Pos	Record?	EventID	CompID	Pos
SwM100Free	Thorpe	3	NA	SwM100Free	Thorpe	3
SwM100Free	Hoogenband	1	No	SwM100Free	Hoogenband	1
SwM100Free	Schoeman	2	NA	SwM100Free	Schoeman	2
SwM100Free	Iles	7	NA	SwM100Free	Iles	7
ShotputW	Cumba	1	No	ShotputW	Cumba	1
ShotputW	Ostapchuk	4	NA	ShotputW	Ostapchuk	4
ShotputW	Li	9	NA	ShotputW	Li	9

The given table

Fragment 1

Explanation:

- Fragment 1 is a **CORRECT** vertical fragment.
- Fragment 1 contains the primary keys.

19

Question 2 (Continue)

EventID	CompID	Pos	Record?	CompID	Pos	Record?
SwM100Free	Thorpe	3	NA	Thorpe	3	NA
SwM100Free	Hoogenband	1	No	Hoogenband	1	No
SwM100Free	Schoeman	2	NA	Schoeman	2	NA
SwM100Free	Iles	7	NA	Iles	7	NA
ShotputW	Cumba	1	No	Cumba	1	No
ShotputW	Ostapchuk	4	NA	Ostapchuk	4	NA
ShotputW	Li	9	NA	Li	9	NA

The given table

Fragment 2

Explanation:

- Fragment 2 is an **INCORRECT** vertical fragment.
- Fragment 2 does not contain all **PRIMARY KEYS**.

20

Question 2 (Continue)

EventID	CompID	Pos	Record?	EventID	CompID	Pos	Record?
SwM100Free	Thorpe	3	NA	SwM100Free	Schoeman	2	NA
SwM100Free	Hoogenband	1	No	SwM100Free	Iles	7	NA
SwM100Free	Schoeman	2	NA	ShotputW	Cumba	1	No
SwM100Free	Iles	7	NA	ShotputW	Ostapchuk	4	NA
ShotputW	Cumba	1	No				
ShotputW	Ostapchuk	4	NA				
ShotputW	Li	9	NA				

The given table

Fragment 3

Explanation:

- Fragment 3 is an **INCORRECT** vertical fragment.
- Fragment 3 is a **horizontal fragment**.

21

Question 3

Global DB schema

- Results (EventID, CompID, Pos, Time)
 - CompID = Competitor's ID
 - Pos = Position
- Competitors (CompID, *Country*, *CompName*, NMedals, NEventsEntered)
 - CompName = Competitor's name
 - NMedals = Number of Medals
 - NEventsEntered = Number of Events participated
- Records (EventID, *Country*, *CompName*)

Italic – foreign key/alternate key

22

Question 3 (Continues)

Π Ctry, EventID, Time (Records \bowtie Competitors \bowtie Results)

Competitors				
CompID	Country	CompName	NMedals	NEventsEntered
ThorpeI	AUS	Thorpe	2	5
KlimtI	AUS	Klimt	1	4
Hoogen	NED	Hoogen	2	3

Results				Records		
EventID	CompID	Pos	Time	EventID	Country	CompName
SW1	ThorpeI	1	2:56	SW1	AUS	Thorpe
SW1	KlimtI	4	3:10			
SW2	Hoogen	1	1:32			

23

Question 3 (Continues)

Query:

Π Ctry, EventID, Time (Records \bowtie Competitors \bowtie Results)

Solution

The join in the **centralized database** has the result
Country = AUS, EventID=SW1, Time = 2:56

24

Question 3 (Continues)

- Design a vertical fragmentation
- Results (EventID, CompID, Pos, Time)
 - Site 1: Results1(EventID, CompID, Pos)
 - Site 2: Results2(EventID, CompID, Time)
- Competitors (CompID, Country, CompName, NMedals, NEventsEntered)
 - Site 1: Competitors1(CompID, CompName, NMed)
 - Site 2: Competitors2(CompID, Country, NEvent)
- Records (EventID, Country, CompName)
 - Site 1: Records1(EventID, Country)
 - Site 2: Records2(EventID, CompName)

25

Question 3 (Continues)

Data Population for the vertical fragmentation

Π Ctry, EventID, Time (Records \bowtie Competitors \bowtie Results)

Results1 (Site 1)			Results2 (Site 2)		
EventID	CompID	Pos	EventID	CompID	Time
SW1	ThorpeI	1	SW1	ThorpeI	2:56
SW1	KlimtI	4	SW1	KlimtI	3:10
SW2	Hoogen	1	SW2	Hoogen	1:32

Competitors1 (Site 1)			Competitors2 (Site 2)		
CompID	CompName	NMed	CompID	Country	NEvent
ThorpeI	Thorpe	2	ThorpeI	AUS	5
KlimtI	Klimt	1	KlimtI	AUS	4
Hoogen	Hoogen	2	Hoogen	NED	3

Records1 (Site 1)		Records2 (Site 2)	
EventID	Country	EventID	CompName
SW1	AUS	SW1	Thorpe

26

Question 3 (Continues)

Solution

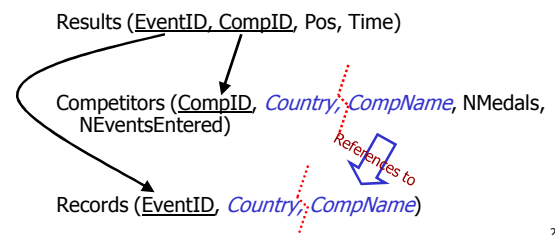
EventID	Country	Time
SW1	AUS	2:56
SW1	AUS	3:10

- The join is in the distributed database system
 - Only the first row is valid
 - Why ?

27

Question 3 (Continues)

There are violations on *referential integrity* constraints



28

Question 3 (Continues)

- Conclusion
 - Even though we have correct horizontal, vertical or hybrid fragmentation in the distributed database systems, the query results could be **INCORRECT** if the fragmentation has inadvertently **broken** the referential integrity.

29

Question 4 - Hybrid Fragmentation

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

Q (A) Check its correctness

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

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

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

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

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

Π = Projection
 σ = Selection

30

Question 4 (Continue)

Visualization of the Fragmentation

R_1

A	B	C	D
1			
2			
3			
4			

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

R_2

A	B	C	D
1			
2			
3			
4			

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

R_3

A	B	C	D
		3	
		4	
		5	
		6	

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

R_4

A	B	C	D
		3	
		4	
		5	
		6	

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

R_5

A	B	C	D
		3	
		4	
		5	
		6	

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

31

Question 4 (Continue)

R_1

A	B	C	D

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

R_2

A	B	C	D

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

$R_1 \cup R_2$

A	B	C	D

Union

32

Question 4 (Continue)

R_4

A	B	C	D

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

R_5

A	B	C	D

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

$R_4 \cup R_5$

A	B	C	D

Union

33

Question 4 (Continue)

$R_1 \cup R_2$

A	B	C	D

$R_4 \cup R_5$

A	B	C	D

$(R_1 \cup R_2) \bowtie (R_4 \cup R_5)$

A	B	C	D

Join

34

Question 4 (Continue)

$(R_1 \cup R_2) \bowtie (R_4 \cup R_5)$

A	B	C	D

R_3

A	B	C	D

$(R_1 \cup R_2) \bowtie (R_4 \cup R_5) \bowtie R_3$

A	B	C	D

Join

Outcome:

- The proposed fragmentation is incomplete, thus incorrect.
- There is substantial DATA redundancy for area .

35

Question 4 (Continue)

$(R_1 \cup R_2) \bowtie (R_4 \cup R_5) \bowtie R_3$

A	B	C	D

The Expected Result

A	B	C	D

36



Question 4 (Continue)

- Obviously, the proposed fragmentation has violated rules:
- **Completeness:** The fragmentation has lost all values equal to or great than **five** under all attributes.
- **Reconstruction:** the original table is no longer achieved by combining all fragments from R_1 to R_5 .
- **Disjointness:** There are redundant/overlapped values less than 5 under attribute C.

37



Thank You

38