

CSE 4125: Distributed Database Systems

Chapter – 4 (Part – B)

Distributed Database Design

The Design of Horizontal Fragmentation

Horizontal Fragments

A horizontal fragment R_i of a relation R consists of all the tuples of R that satisfy a minterm predicate m_i .

Which minterm predicate should we use?

- We have to decide on the *set of simple predicates that are the basis for the minterm predicates*.

Desirable properties of the set of simple predicates

The set should be complete and minimal.

Informally, the set should include only predicates with attributes and conditions that are used in the applications

Example

Consider the following global relation, fragmentation schema and applications.

J	<u>JNO</u>	JNAME	BUDGET	LOC
	J1	Instrumental	150,000	Montreal
	J2	GUI	135,000	New York
	J3	CAD/CAM	250,000	New York
	J4	Database Dev	310,000	Orlando

Fragmentation schema: Three horizontal fragments JP1, JP2, JP3 based on locations respectively Montreal, New York, Orlando.

Applications: APP1 = *SELECT * FROM J*;

APP2 = *SELECT * FROM J WHERE BUDGET < 200,000*;

Determine the set of simple predicates Pr to obtain J from its fragments.

If APP1 and APP2 are issued, do you think Pr will satisfy completeness property? If not, then make necessary changes to make the set complete. Justify your answers.

Example of Completeness

Pr= { LOC="Montreal",
LOC="New York",
LOC="Orlando"

$$JP_1 = SL \text{ } LOC = \text{MONTREAL} \text{ } J$$

JP1

JNO	JNAME	BUDGET	LOC
J1	Instrumental	150,000	Montreal

$$JP_2 = SL \text{ } LOC = \text{NEW YORK} \text{ } J$$

JP2

JNO	JNAME	BUDGET	LOC
J2	GUI	135,000	New York
J3	CAD/CAM	250,000	New York

$$JP_3 = SL \text{ } LOC = \text{ORLANDO} \text{ } J$$

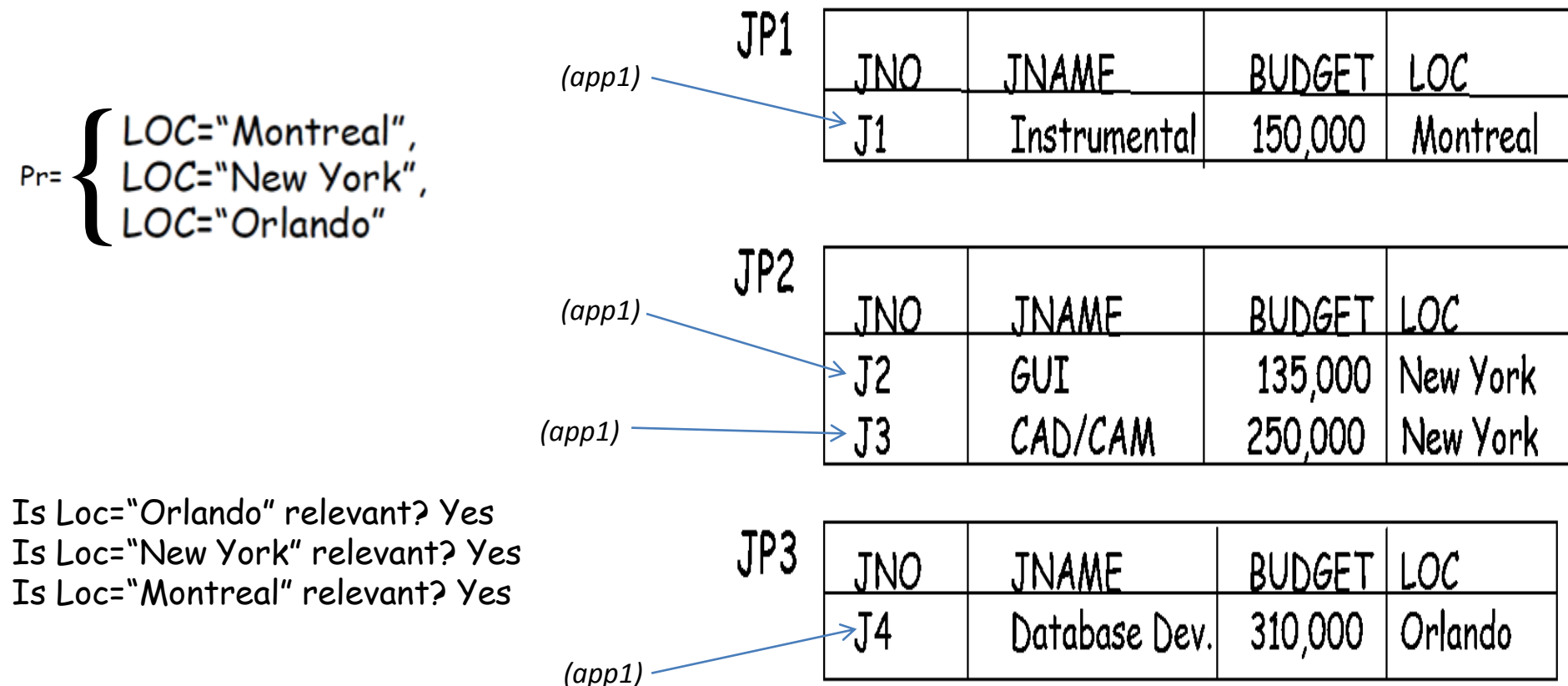
JP3

JNO	JNAME	BUDGET	LOC
J4	Database Dev.	310,000	Orlando

Case 1: An application (*app1*) is issued in three sites. It wants to access the tuples according to the location (any location).

APP1 = *SELECT * FROM J*;

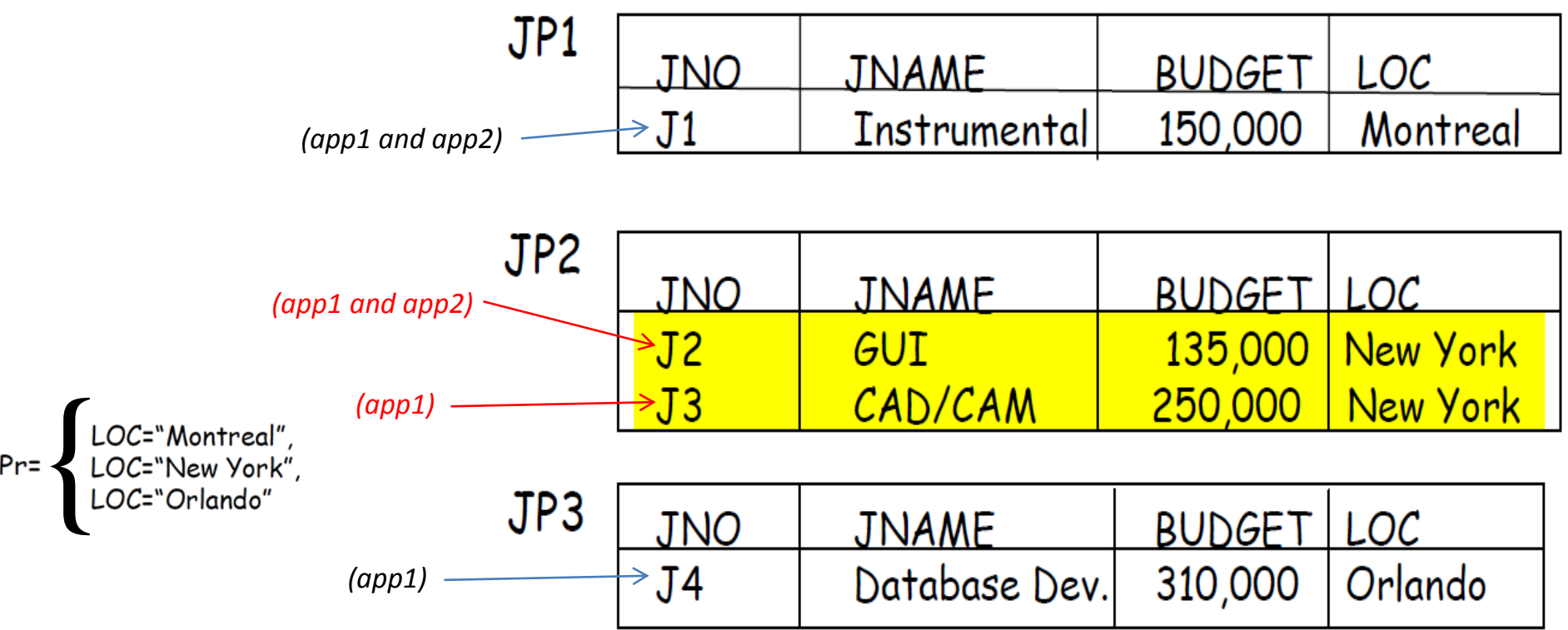
In this case, P_r is complete because each tuple of each fragment has the same probability of being accessed.



Case 2: There is a second application (*app2*) which is also issued in three sites. It accesses only those tuples where *budget is less than \$200,000*.

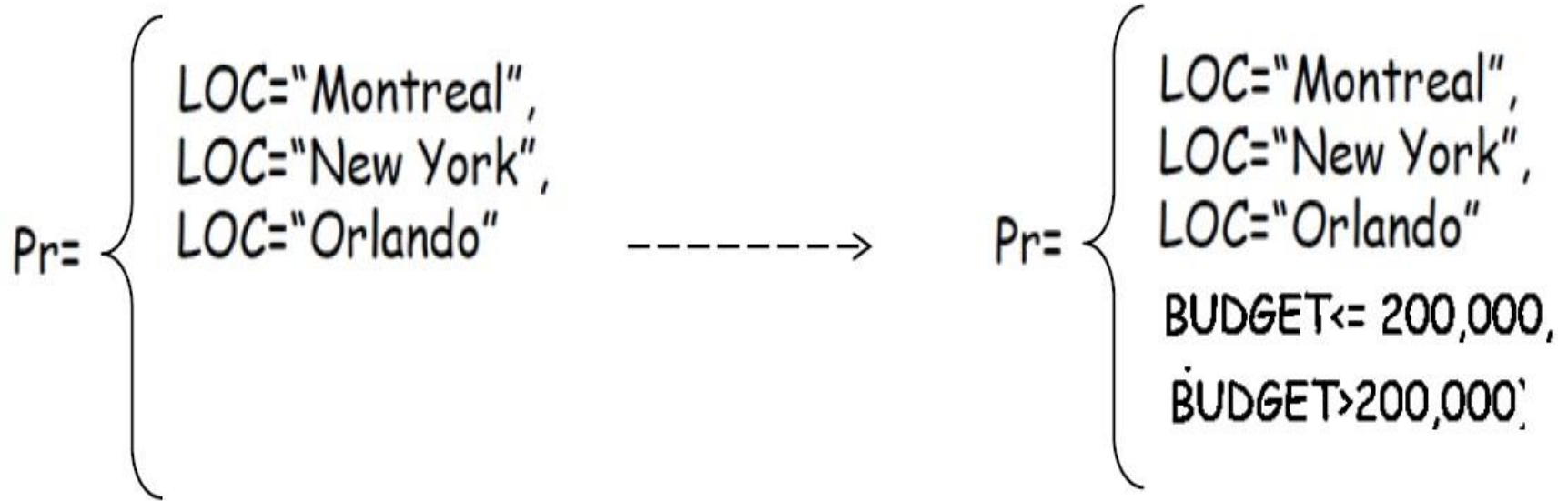
APP2 = *SELECT * FROM J WHERE BUDGET < 200,000;*

Tuple J_2 has higher access probability than tuple J_3 in JP_2 . In this case, P_r is not complete since some tuples (J_i) in JP_i has higher access probability.



To make the set complete, we need to add –

$(BUDGET \leq 200,000, BUDGET > 200,000)$ to P_r .



Pr= {
 LOC="Montreal",
 LOC="New York",
 LOC="Orlando"
 BUDGET <= 200,000,
 BUDGET > 200,000

SL LOC = 'Montreal' J

JP1

JNO	JNAME	BUDGET	LOC
J1	Instrumental	150,000	Montreal

SL LOC = 'New York' and budget <= 200,000 J

JP21

JNO	JNAME	BUDGET	LOC
J2	GUI	135,000	New York

SL LOC = 'New York' and budget > 200,000 J

JP22

JNO	JNAME	BUDGET	LOC
J3	CAD/CAM	250,000	New York

SL LOC = 'Orlando' J

JP3

JNO	JNAME	BUDGET	LOC
J4	Database Dev.	310,000	Orlando

$(app_1 \text{ and } app_2)$ JP1

JNO	JNAME	BUDGET	LOC
J1	Instrumental	150,000	Montreal

$(app_1 \text{ and } app_2)$ JP21

JNO	JNAME	BUDGET	LOC
J2	GUI	135,000	New York

(app_1) JP22

JNO	JNAME	BUDGET	LOC
J3	CAD/CAM	250,000	New York

(app_1) JP3

JNO	JNAME	BUDGET	LOC
J4	Database Dev.	310,000	Orlando

Is Loc="Orlando" relevant? Yes
 Is Loc="New York" relevant? Yes
 Is Loc="Montreal" relevant? Yes
 Is Budget \leq 200,000 relevant? Yes
 Is Budget $>$ 200,000 relevant? Yes

Considering the previous *app1* and *app2* , and the set P_r .

➤ If we add the predicate $JNAME = \text{"Instrument"}$ to P_r .

➤ Is resulting P_r complete?

$P_r = \left\{ \begin{array}{l} LOC = \text{"Montreal"}, \\ LOC = \text{"New York"}, \\ LOC = \text{"Orlando"}, \\ BUDGET \leq 200,000, \\ BUDGET > 200,000, \\ JNAME = \text{"Instrument"} \end{array} \right. ??$

Is Loc="Orlando" relevant? Yes
Is Loc="New York" relevant? Yes
Is Loc="Montreal" relevant? Yes
Is Budget <= 200,000 relevant? Yes
Is Budget > 200,000 relevant? Yes
Is Jname = "Instrumental" relevant? No

Completeness

A set of simple predicate P_r is said to be **complete** if and only if –

There is an *equal probability* of *access* by *every application (i.e. query)* to *any tuple* belonging to *any minterm fragment* that is defined by P_r

Minimality

The set of predicates P_r is minimal if and only if there is at least one application (i.e. query) that accesses the fragment.

If all the predicates of a set P_r is relevant, then P_r is Minimal.

$P_r = \left\{ \begin{array}{l} \text{LOC} = \text{"Montreal"}, \\ \text{LOC} = \text{"New York"}, \\ \text{LOC} = \text{"Orlando"} \\ \text{BUDGET} \leq 200,000, \\ \text{BUDGET} > 200,000. \end{array} \right.$

$P_r = \left\{ \begin{array}{l} \text{LOC} = \text{"Montreal"}, \\ \text{LOC} = \text{"New York"}, \\ \text{LOC} = \text{"Orlando"} \\ \text{BUDGET} \leq 200,000, \\ \text{BUDGET} > 200,000, \\ \text{JNAME} = \text{"Instrument"} \end{array} \right. ??$

Let's See

Say, only APP3 application accesses J wants to access project located in New York.

APP3 = SELECT * FROM J WHERE LOC = "NEW YORK";

Pr= { LOC="Montreal",
LOC="New York",
LOC="Orlando"

Is Loc="Orlando" relevant? No
Is Loc="New York" relevant? Yes
Is Loc="Montreal" relevant? No

Is the set Complete? Minimal?

JP1

JNO	JNAME	BUDGET	LOC
J1	Instrumental	150,000	Montreal

JP2

JNO	JNAME	BUDGET	LOC
J2	GUI	135,000	New York
J3	CAD/CAM	250,000	New York

JP3

JNO	JNAME	BUDGET	LOC
J4	Database Dev.	310,000	Orlando

Exercise

Consider the following global relation, fragmentation schema and applications.

STUDENT

SNUM	SNAME	DEPT	SEM
1	A	CSE	1
2	B	EEE	2
3	C	CSE	3
4	D	EEE	4
5	E	CSE	6
6	F	CSE	7

Fragmentation Schema:

$STUDENT_1 = SL_{DEPT = "CSE"} STUDENT$

$STUDENT_2 = SL_{DEPT = "EEE"} STUDENT$

Applications:

$APP_1 = \text{SELECT } * \text{ FROM } STUDENT \text{ WHERE DEPT} = "CSE" \text{ OR DEPT} = "EEE";$

$APP_2 = \text{SELECT } * \text{ FROM } STUDENT \text{ WHERE SEM} = 6;$

Now, answer the following questions:

- Define *simple predicate* and *minterm predicate*. [2]
- Determine the set of simple predicates P_r to obtain **STUDENT** from its fragments. If **APP₁** and **APP₂** are issued, do you think P_r will be completed? If not, then make necessary changes to make the set complete. Justify your answer. [5]