CSE 4125: Distributed Database Systems
Chapter – 6
(Part – B)

Optimization of Access Strategies.

Pre-requisites

- Knowledge of Chapter 5

Topics to be discussed -

- Optimization Graph

Optimization Graph

Optimization Graph

- A model to describe query optimization.
- Convenient than operator tree.
- Include only *critical* operations (critical for data transmission).

- Unary operations are *not critical.*
  - –Effect only by reducing operands and **do not need data transmission.**
  - –These operations are collected by a program called ***fragment reducer.***

- Binary operations are *critical.*
  - –When operands are not in the same site, they need data transmission.
  - –CP, DF and SJ are not considered as they are rare. **JN** and **UN** are kept which gives us a graph called **optimization graph.**

Example

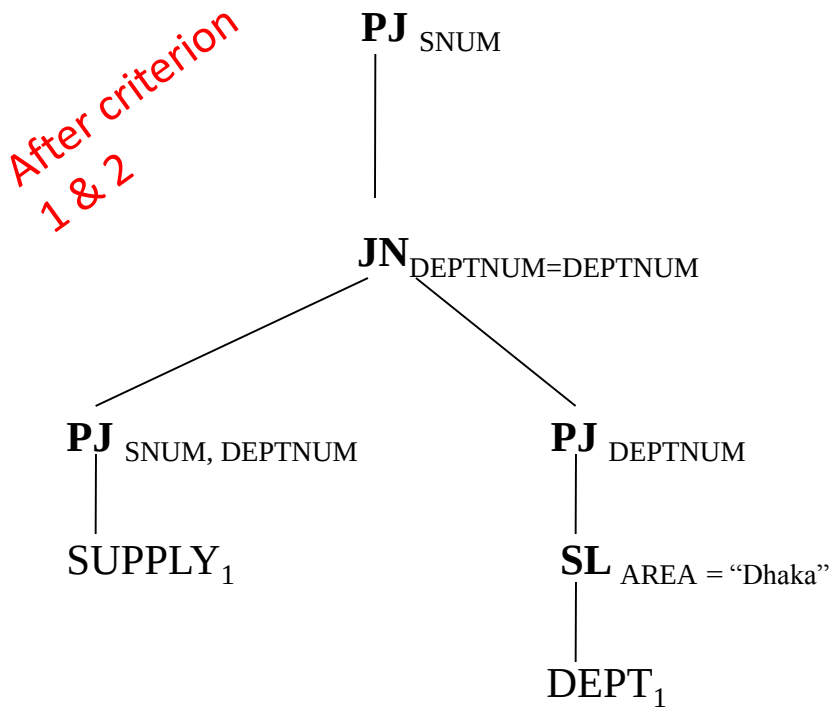
Consider the following Global Relational Schema, query & corresponding Database Profile.

$SUPPLY_1$ (snum, pnum, deptnum, quan)

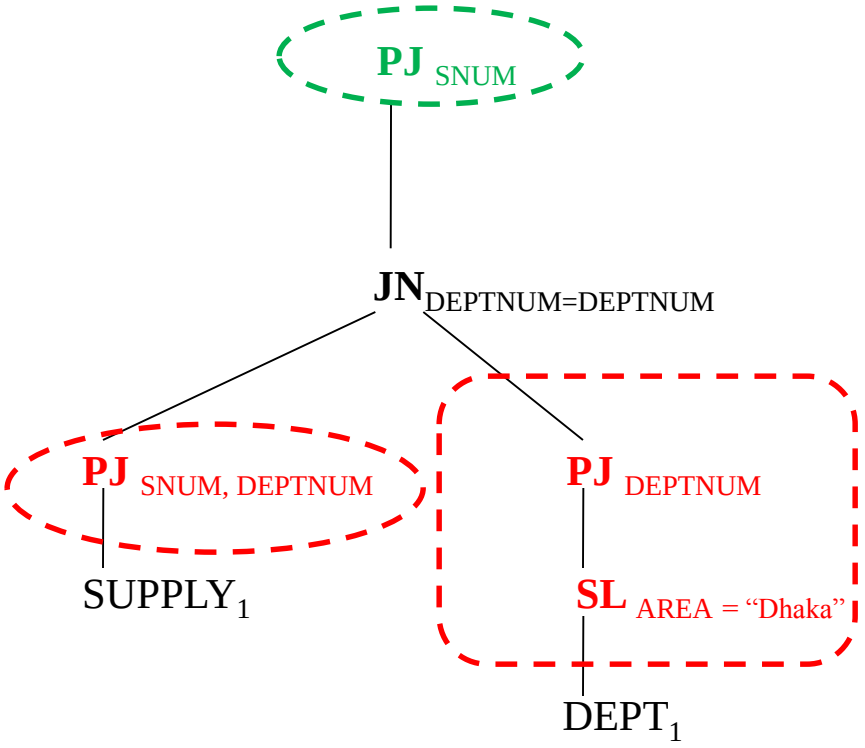
$DEPT_1$ (deptnum, name, area, mgrnum)

Q: $\mathbf{PJ}_{\text{SNUM}}\ (\text{SUPPLY}_1\ \mathbf{JN}_{\text{DEPTNUM=DEPTNUM}}\ (\mathbf{SL}_{\text{AREA = “Dhaka”}}\ \text{DEPT}_1\ ))$

Optimization Graph (example)



Optimization Graph (example)



Fragment Reducer Program:

Before binary operation:

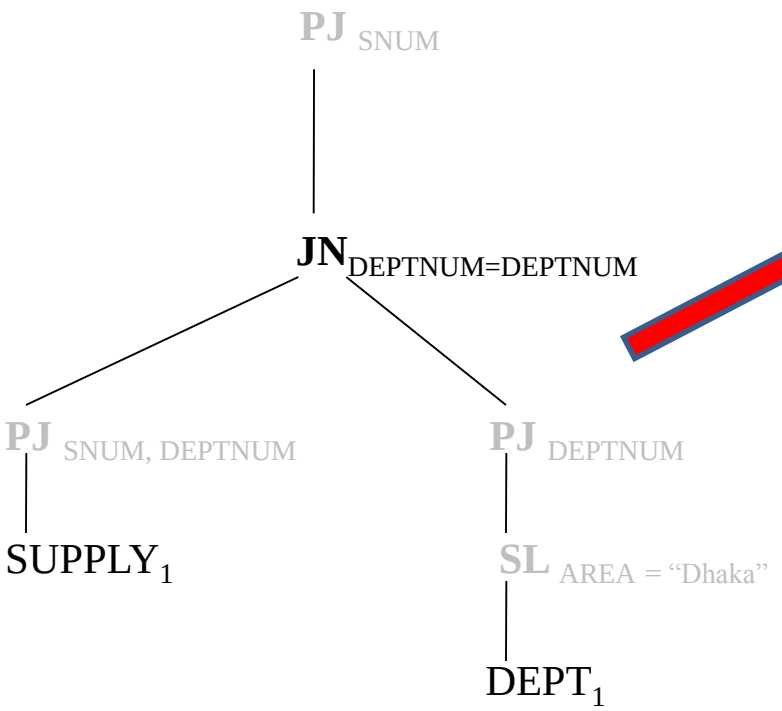
Reducer for $SUPPLY_1$: **PJ** SNUM, DEPTNUM

Reducer for $DEPT_1$: **PJ** DEPTNUM **SL** AREA="Dhaka"

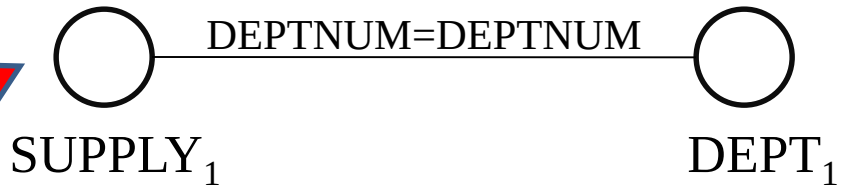
After binary operation:

Reducer for Result: **PJ** SNUM

Optimization Graph (example)



Optimization Graph



** In **Optimization Graph,** nodes represent reduced fragments, joins are represented by edges between nodes which are labeled with the join specification.

*** Unions are represented by hypernodes enclosing their operands. [book-p.138]

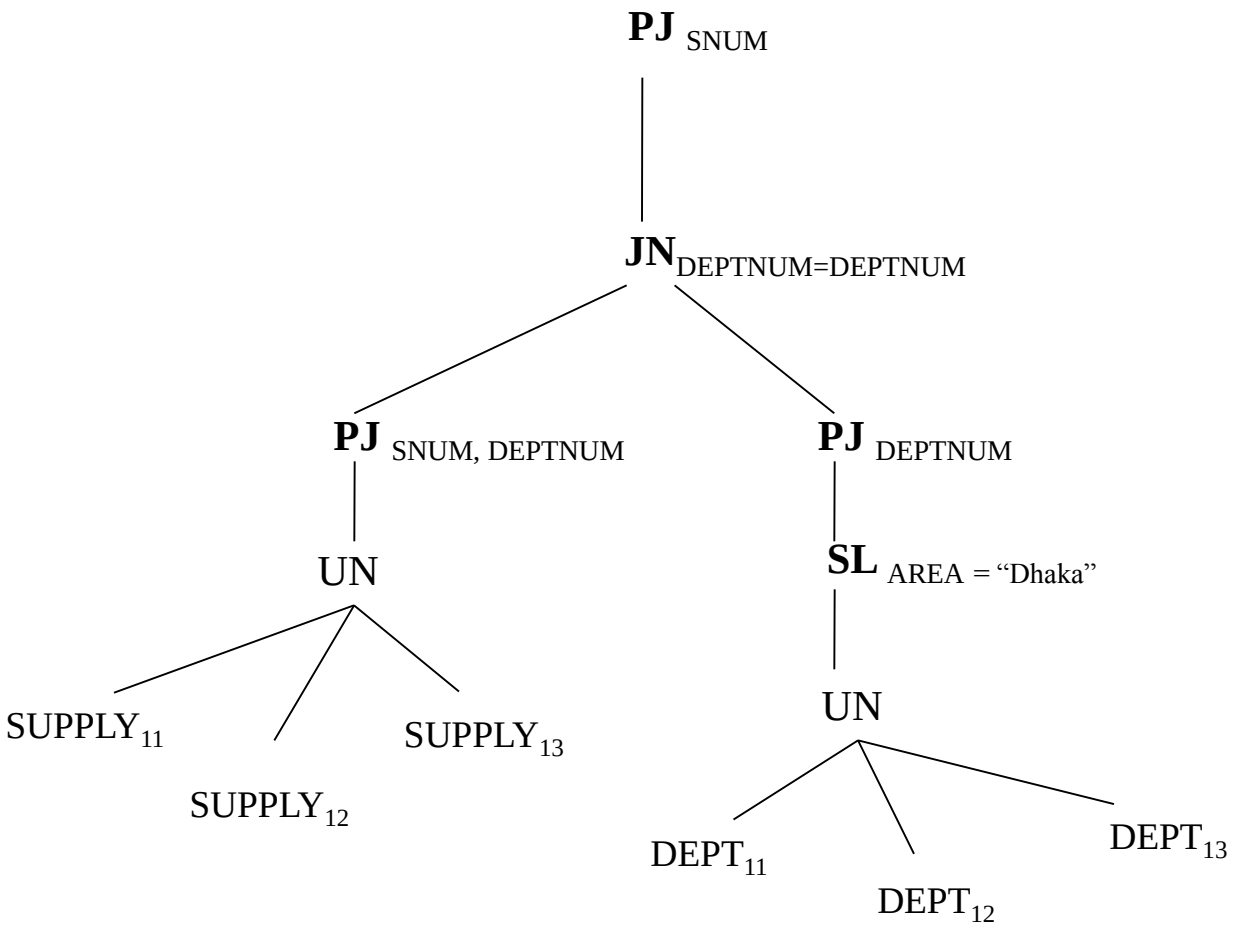
Optimization Graph (Distributed)

What if $SUPPLY_1$ and $DEPT_1$ both have three horizontal fragments each?

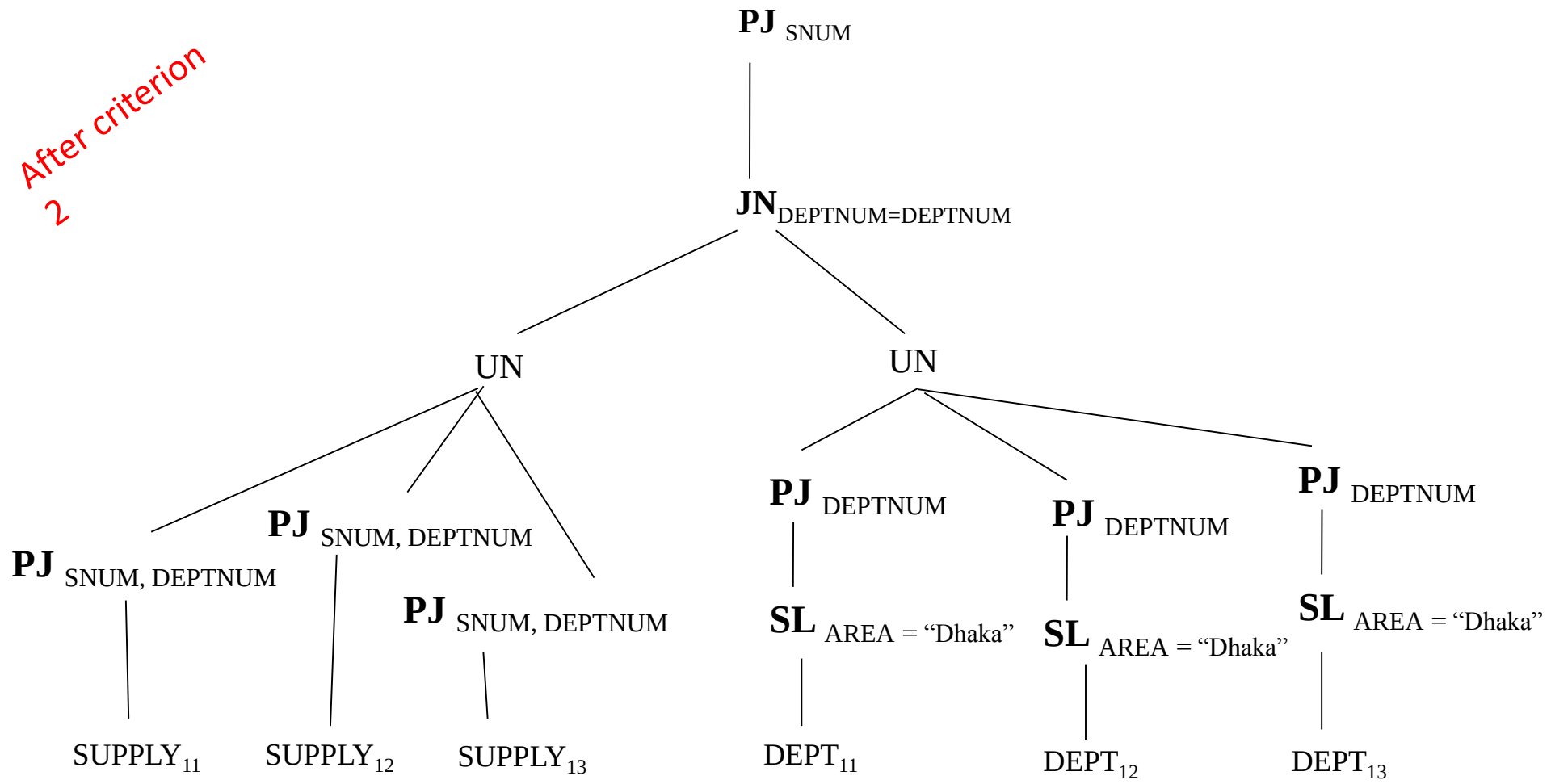
$SUPPLY_1$ has three fragments -> $SUPPLY_{11,}$ $SUPPLY_{12,}$ and $SUPPLY_{13}$

$DEPT_1$ has three fragments -> $DEPT_{11,}$ $DEPT_{12,}$ and $DEPT_{13}$

Optimization Graph (Distributed)



Optimization Graph (Distributed)



Fragment Reducer Program:

Before binary operation:

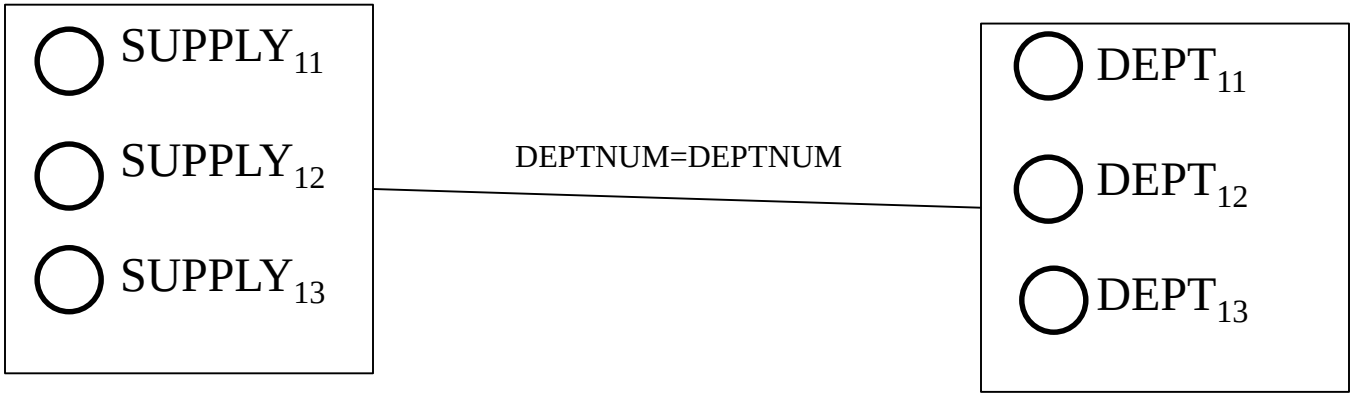
Reducer for $SUPPLY_{11}$: **PJ** $_{SNUM, DEPTNUM}$
Reducer for $SUPPLY_{12}$: **PJ** $_{SNUM, DEPTNUM}$
Reducer for $SUPPLY_{13}$: **PJ** $_{SNUM, DEPTNUM}$

Reducer for $DEPT_{11}$: **PJ** $_{DEPTNUM}$ **SL** $_{AREA=\text{“Dhaka”}}$
Reducer for $DEPT_{12}$: **PJ** $_{DEPTNUM}$ **SL** $_{AREA=\text{“Dhaka”}}$
Reducer for $DEPT_{13}$: **PJ** $_{DEPTNUM}$ **SL** $_{AREA=\text{“Dhaka”}}$

After binary operation:

Reducer for Result: **PJ** $_{SNUM}$

Optimization Graph



Exercise

Consider the following global relational schemata.

EMP (ID, NAME, SAL, AGE, MGRNUM, DEPTNUM)
DEPT (ID, AREA, DEPTNUM, MGRNUM)

Corresponding fragmentation schemata:

$$EMP_1 = SL_{SAL < 25K}\, EMP$$
$$EMP_2 = SL_{SAL > 25K}\, EMP$$
$$DEPT_1 = SL_{AREA = "North"}\, DEPT$$
$$DEPT_2 = SL_{AREA = "South"}\, DEPT$$

Consider the following query **Q** with the *global relational and fragmentation schemata of question 4(b)*.

$$\boldsymbol{Q}: PJ_{NAME}\,((EMP\; JN_{DEPTNUM=DEPTNUM}\; SL_{MGRNUM=375}\; DEPT)\;\; DF$$
$$(SL_{SAL > 25000}\; EMP\; JN_{DEPTNUM=DEPTNUM}\;\; SL_{MGRNUM=375}\; DEPT))$$

Write a *fragment reducer program* for the query **Q** to optimize the corresponding operator tree. Draw the obtained *optimization graph*. [3+1]