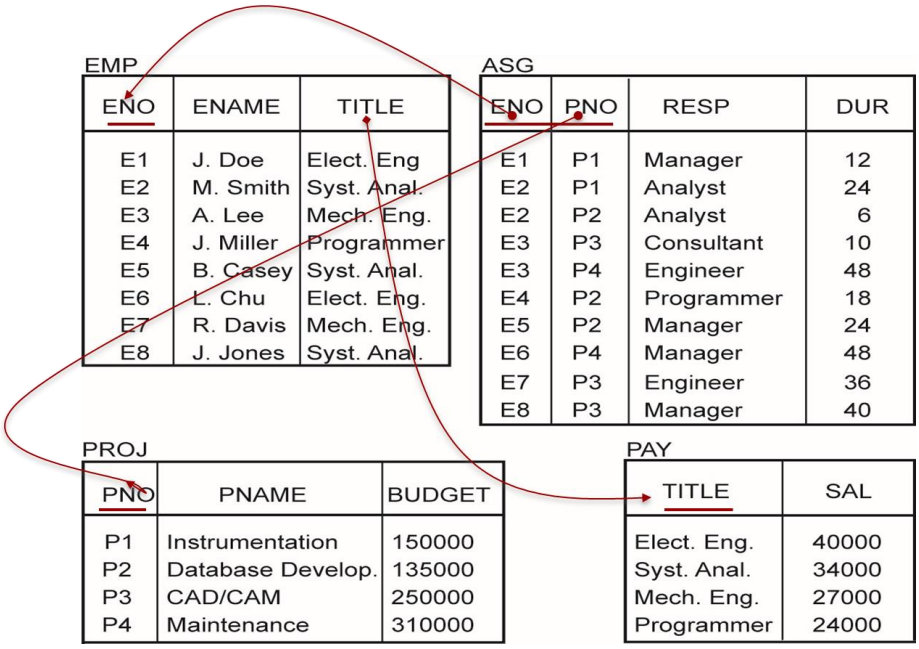
INFS3200 Advanced Database Systems

Tutorial 2: Distributed Query Processing

Semester 1, 2021

Question 1: At the global level, distributed query processing consists of three main steps: query decomposition, data localization and global optimization. Discuss these three steps, focusing on the input, output, and objectives for each step.

Question 2: Consider the following database (arrows are foreign-key linkages):



EMP

ENO	ENAME	TITLE
E1	J. Doe	Elect. Eng
E2	M. Smith	Syst. Anal.
E3	A. Lee	Mech. Eng.
E4	J. Miller	Programmer
E5	B. Casey	Syst. Anal.
E6	L. Chu	Elect. Eng.
E7	R. Davis	Mech. Eng.
E8	J. Jones	Syst. Anal.

ASG

ENO	PNO	RESP	DUR
E1	P1	Manager	12
E2	P1	Analyst	24
E2	P2	Analyst	6
E3	P3	Consultant	10
E3	P4	Engineer	48
E4	P2	Programmer	18
E5	P2	Manager	24
E6	P4	Manager	48
E7	P3	Engineer	36
E8	P3	Manager	40

PROJ

PNO	PNAME	BUDGET
P1	Instrumentation	150000
P2	Database Develop.	135000
P3	CAD/CAM	250000
P4	Maintenance	310000

PAY

TITLE	SAL
Elect. Eng.	40000
Syst. Anal.	34000
Mech. Eng.	27000
Programmer	24000

(a) Assume that relation PROJ is horizontally fragmented as follows:

$PROJ_1 = \sigma_{PNO \leq "P2"}\ PROJ$

$PROJ_2 = \sigma_{PNO > "P2"}\ PROJ$

Transform the following query into a reduced query on fragments:

```
SELECT ENO, PNAME
FROM   PROJ, ASG
WHERE  PROJ.PNO = ASG.PNO AND PNO = "P4"
```

(b) Assume PROJ is fragmented as above, and ASG is fragmented as below:

$ASG_1 = \sigma_{PNO \leq "P2"}\ ASG$

$ASG_2 = \sigma_{"P2" < PNO \leq "P3"}\ ASG$

$ASG_3 = \sigma_{PNO > "P3"}\ ASG$

Transform the following query into a reduced query on fragments, and determine whether it is better than the localized query:

```
SELECT RESP, BUDGET
FROM ASG, PROJ
WHERE ASG.PNO = PROJ.PNO AND PNAME = "CAD/CAM"
```

Question 3: Let R(A, B) and S(B, C, D) be two relations as shown below:

R(A, B)

A	B
1	4
1	5
2	4
2	6
3	7

S(B, C, D)

B	C	D
4	5	0
4	7	8
5	0	1
5	1	1

(a) Compute $R \ltimes S$
(b) Compute $S \ltimes R$
(c) Assume R is at site 1 and S is at site 2, and a query $R \bowtie S$ is issued at site 2. List the steps for a query processing strategy using semi-join, and check if the semi-join is a beneficial option in this case (ignore local processing cost).