Sheet-5

(a) Selectivity,

$$f_{i,j} = \frac{|R_i \bowtie_{p_{i,j}} R_j|}{|R_i \times R_j|}$$

$\therefore$ Cardinality of result/ join result

$|R_i \bowtie_{p_{i,j}} R_j| = f_{i,j} * R_i * R_j$

Given, relations A, B, C with cardinalities

$|A| = 30$
$|B| = 50$
$|C| = 80$ and

Selectivities,

$f_{A,B} = 0.2$

$f_{B,C} = 0.4$

Cout is cost function and

Cross products are allowed

We know, cost function

$$Cost(T) = \begin{cases} 0 & \text{if } T \text{ is a leaf } R_i \\ |T| + Cost(T_1) + Cost(T_2), & \text{if } T = T_1 \bowtie T_2 \end{cases}$$

Here,

$$Cost(A) = Cost(B) = Cost(C) = 0$$

$\Rightarrow Cost(A \bowtie B) = |A \bowtie B| + Cost(A) + Cost(B)$

$$= f_{A,B} \times |A| \times |B| + 0 + 0$$

$$= 0.2 \times 30 \times 50 = 300$$

$\Rightarrow Cost(B \bowtie C) = |B \bowtie C| + Cost(B) + Cost(C)$

$$= f_{B,C} \times |B| \times |C|$$

$$= 0.4 \times 50 \times 80 = 1600$$

$\Rightarrow Cost(A \times C) = |A \times C| + Cost(A) + Cost(C)$

$$= f_{A,C} \times |A| \times |C|$$

$$= 1 \times 30 \times 80 = 2400$$

$\Rightarrow Cost((A \bowtie B) \bowtie C) =$

$|(A \bowtie B) \bowtie C| + Cost(A \bowtie B) + Cost(C)$

$= f_{A,B} \times f_{B,C} \times |A| \times |B| \times |C| + 300 + 0$

$= 0.2 \times 0.4 \times 30 \times 50 \times 80 + 300$

$= 9900$

$\Rightarrow Cost((B \bowtie C) \bowtie A) =$

$|(B \bowtie C) \bowtie A| + Cost(B \bowtie C) + Cost(A)$

$= f_{B,C} \times f_{A,B} \times |B| \times |C| \times |A| + 1600 + 0$

$= 0.2 \times 0.4 \times 50 \times 80 \times 30 + 1600$

$= 9600 + 1600 = 11200$

$\Rightarrow Cost((A \times C) \bowtie B) =$

$|(A \times C) \bowtie B| + Cost(A \times C) + Cost(B)$

$= f_{A,B} \times f_{B,C} \times A \times C \times B + 2400 + 0$

$= 9600 + 2400 = 12000$

So the DP table would be,

Relations	T	Cost
A, B	$A \bowtie B$	300
B, C	$B \bowtie C$	1600
A, C	$A \times C$	2400
A, B, C	$(A \bowtie B) \bowtie C$	9900
A, B, C	$(B \bowtie C) \bowtie A$	11200
A, B, C	$(A \times C) \bowtie B$	12000

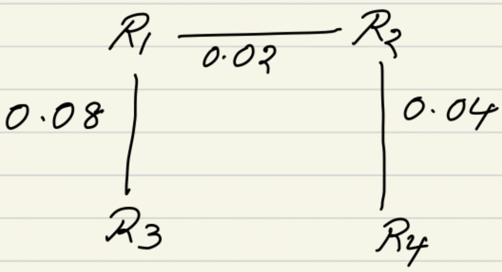
(b) Given, $|R_1| = 20$

$|R_2| = 10$

$|R_3| = 30$

$|R_4| = 10$

and the following query graph:-



that means,

$f_{R_1, R_2} = 0.02$

$f_{R_1, R_3} = 0.08$

$f_{R_2, R_4} = 0.04$

We've to find optimal Bushy tree for

$\{R_1, R_2, R_3, R_4\}$

Calculating Cout for possible combinations:

$\Rightarrow \{R_1\} \cup \{R_2, R_3, R_4\}$

$= Cout(R_1 \bowtie ((R_2 \bowtie R_4) \bowtie R_3))$

$= |R_1 \bowtie ((R_2 \bowtie R_4) \bowtie R_3))| + Cout(R_1) + Cout((R_2 \bowtie R_4) \bowtie R_3)$

$= f_{R_1, R_2} \times |R_1| \times |(R_2 \bowtie R_4) \bowtie R_3)| + 0 + 124$

$= 0.02 \times 20 \times 120 + 124$

$= 172$

$\Rightarrow \{R_2\} \cup \{R_1, R_3, R_4\}$

$= Cout\ (R_2 \bowtie ((R_1 \bowtie R_3) \bowtie R_4))$

$= |R_2 \bowtie ((R_1 \bowtie R_3) \bowtie R_4))| + Cout(R_2) + Cout(R_1 \bowtie (R_3 \bowtie R_4))$

$= f_{R_2,R_1} \times f_{R_2,R_4} \times |R_2| \times |((R_1 \bowtie R_3) \bowtie R_4))| + 528$

$= 0.02 \times 0.04 \times 10 \times 480 + 528$

$= 531.84$

$\Rightarrow \{R_3\} \cup \{R_1, R_2, R_4\}$

$= Cout\ (R_3 \bowtie ((R_1 \bowtie R_2) \bowtie R_4))$

$= |R_3 \bowtie ((R_1 \bowtie R_2) \bowtie R_4))| + Cout\ (R_3) + Cout\ ((R_1 \bowtie R_2) \bowtie R_4))$

$= f_{R_3,R_1} \times |R_3| \times |((R_1 \bowtie R_2) \bowtie R_4)| + 0 + 5.6$

$= 0.08 \times 30 \times 1.6 + 5.6$

$= 9.44$ ✓ optimal

$\Rightarrow \{R_4\} \cup \{R_1, R_2, R_3\}$

$= Cout(R_4 \bowtie ((R_1 \bowtie R_2) \bowtie R_3))$

$= |R_4 \bowtie ((R_1 \bowtie R_2) \bowtie R_3))| + Cout(R_4) + Cout((R_1 \bowtie R_2) \bowtie R_3))$

$= f_{4,2} \times |R_4| \times |((R_1 \bowtie R_2) \bowtie R_3))| + 0 + 13.6$

$= 0.04 \times 10 \times 9.6 + 13.6$

$= 17.44$

$\Rightarrow \{R_1, R_2\} \cup \{R_3, R_4\}$

$= Cout((R_1 \bowtie R_2) \bowtie (R_3 \bowtie R_4))$

$= |(R_1 \bowtie R_2) \bowtie (R_3 \bowtie R_4))| + Cout(R_1 \bowtie R_2) + Cout(R_3 \bowtie R_4)$

$= f_{1,3} \times f_{2,4} \times |R_1 \bowtie R_2| \times |R_3 \bowtie R_4| + 4 + 300$

$= 0.08 \times 0.04 \times 4 \times 300 + 304$

$= 307.84$

$\Rightarrow \{R_1, R_4\} \cup \{R_2, R_3\}$

$= Cout((R_1 \bowtie R_4) \bowtie (R_2 \bowtie R_3))$

$= |((R_1 \bowtie R_4) \bowtie (R_2 \bowtie R_3))| + Cout(R_1 \bowtie R_4)$
$+ Cout(R_2 \bowtie R_3)$

$= f_{1,2} \times f_{1,3} \times f_{4,2} \times |(R_1 \bowtie R_4)| \times |R_2 \bowtie R_3| +$
$200 + 300$

$= 0.02 \times 0.08 \times 0.04 \times 200 \times 300 + 500$

$= 503.84$

From above calculations we can see that $(R_3 \bowtie ((R_1 \bowtie R_2) \bowtie R_4))$ is the optimal Bushi tree with minimum Cout.

1.(a)

Given $V(R,b) = V(S,b) = 20$

Let, $Res(R(a,b) \bowtie S(b,c))$ = Result of join.

Also given,

$T(R) = 60$; $T(S) = 80$

Histogram for R.b		Histogram for S.b	
0	5	0	10
1	4	1	8
2	10	2	5
3	5	4	7
others	36	others	50

Case 1: when $b=0$

$R(a,0) \bowtie S(0,c) = 50$ tuples

Case 2: when $b=1$

$4 \times 8 = 32$ tuples

case 3: when b = 2

$10 \times 5 = 50$ tuples

case 4: when b = 3

There are 5 tuples in R with R.b = 3. There are 50 tuples in S with 20 − 4 = 16 other values, so average is $50/16 \approx 4$ tuples (ceiling)

So, $R(a,4) \bowtie S(4,c) = 5 \times 4 = 20$

case 4: when b = 4

7 tuples in B with S.b = 4, There are 36 other tuples in R with 20 − 4 = 16 other values. So average is,

$36/16 \approx 3$ (ceiling)

$\therefore R(a,4) \bowtie S(4,c) = 3 \times 7 = 21$

Now,

R has average 2 and S has 3 tuples per value. So for 15 other values,

$R(a, \text{other}) \bowtie S(\text{other}, c)$

$= 15 \times 2 \times 3 = 90$

$\therefore Res(R(a,b) \bowtie S(b,c)) = 50 + 32 + 50 + 20 + 21 + 90 = \boxed{263}$

For, $T(R) = 60$ and $T(S) = 80$

we find no. of tuples

$(60 \times 80)/20 = 240$

So the difference is $(263 - 240)$

$= 23$ tuples

(b) Size of natural join

$R(a, b) \bowtie S(b, c)$ for following histogram info:-

	$b < 0$	$b = 0$	$b > 0$
R	400	100	200
S	400	300	800

At first we calculate total no. of tuples that can be there after join operation:-

For $b < 0$: $400 \times 400 = 160000$

For $b = 0$: $100 \times 300 = 30000$

For $b > 0$: $200 \times 800 = 160000$

So, total no. of tuples =

$160000 + 30000 + 160000 = 35000$

Now,

$T(R) = 400 + 100 + 200 = 700$

$T(S) = 400 + 300 + 800 = 1500$

Now we calculate $T(R \bowtie S)$ below for all three cases:-

For $b < 0$: $(700 \times 1500)/400 = 2625$

For $b = 0$: $(700 \times 1500)/300 = 3500$

For $b > 0$: $(700 \times 1500)/800 \approx 1313$

In total $= 7438$

So upper bound will be 35000 and lower bound will be 7438

Question 4:

Nested Query	Solution
SELECT DISTINCT P. playerId FROM Player P WHERE (SELECT COUNT (G.id) FROM Game G WHERE G. playerId = P. playerId) >10	SELECT DISTINCT P.name, COUNT(G.playerId) FROM Player P INNER JOIN Game G ON P.playerId = G.playerId GROUP BY G.playerId HAVING COUNT(G.playerId) > 10
SELECT DISTINCT P.name , (SELECT count (*) FROM Game G WHERE P. playerId = G. playerId) FROM Player P	SELECT DISTINCT P.name, COUNT(G.playerId) FROM Player P LEFT OUTER JOIN Game G ON P.playerId = G.playerId GROUP BY G.playerId

In Query 1, for each tuple in relation Player P, the nested count query will execute. This is an example of Type JA (dependent and aggregation) nested query.

Also, in both cases, sub queries are dependent on outer query.

To un-nest first query, we used INNER JOIN, because we want to list all the distinct players and their #played games, who has played more than 10 games. Here LEFT JOIN would not be necessary because players who have not played any game would be truncated by the HAVING clause.

To un-nest second query, we used LEFT OUTER JOIN, because we want to list all the distinct players (all from left table) and their #played games (only matching right table, null values will be replaced as 0 by aggregate function), even those who have not played any game.