COSC265: Tutorial 4 – Solutions

1.
   - $\mathcal{F}\ \mathcal{SUM}\ \mathcal{QUANTITY}(PRODUCT)$ The result is a single-row, single-column table, containing just one value: 21.
   - $Number\ \mathcal{F}\ \mathcal{SUM}\ \mathcal{QUANTITY}(PRODUCT)$ This time, we get the following table:

Number	Sum Quantity
1	2
2	3
3	4
4	2
5	4
6	1
7	5

   - $Type\ \mathcal{F}\ \mathcal{SUM}\ \mathcal{QUANTITY}(PRODUCT)$

Type	Sum Quantity
A	7
B	7
C	5
D	2

   - $\mathcal{F}\ MAXIMUM\ Quantity(PRODUCT)$
     Answer: 5
   - $\mathcal{F}\ MINIMUM\ Quantity(PRODUCT)$
     Answer: 1
   - $\mathcal{F}\ AVERAGE\ Quantity(PRODUCT)$
     Answer: 3
   - $\mathcal{F}\ COUNT\ Number(PRODUCT)$
     Answer: 7
   - $\mathcal{F}\ SUM\ Quantity(\sigma_{Type='A'}\ (PRODUCT))$
     Answer: 7

2. (a) $R \times S$

A	B	C	D	E
a1	b1	c1	a1	b1
a1	b1	c1	a1	b2
a1	b1	c1	a2	b1
a1	b2	c3	a1	b1
a1	b2	c3	a1	b2
a1	b2	c3	a2	b1
a2	b1	c2	a1	b1
a2	b1	c2	a1	b2
a2	b1	c2	a2	b1
a1	b1	c2	a1	b1
a1	b1	c2	a1	b2
a1	b1	c2	a2	b1
a1	b1	c3	a1	b1
a1	b1	c3	a1	b2
a1	b1	c3	a2	b1
a2	b2	c3	a1	b1
a2	b2	c3	a1	b2
a2	b2	c3	a2	b1
a2	b1	c3	a1	b1
a2	b1	c3	a1	b2
a2	b1	c3	a2	b1

(b) $R - S$: not possible, as R and S are not union compatible relations.

(c) $R \div S$ cannot directly be computed, because this operation requires the attributes to have the same names. Therefore, it is necessary first to rename the attributes D and E:
$S1(A, B) \leftarrow S$

$R \div S1$

C
c3

3. (a) $\pi_{Name}(HOTEL)$

(b) $\sigma_{Type='single' \ AND \ Price<40}(ROOM)$

(c) $\pi_{GUEST.Name}((GUEST * BOOKING) * (\sigma_{Hotel='Park'}(HOTEL)))$

(d) $PHR \leftarrow \sigma_{Name='Park'}(HOTEL * ROOM)$
$GR \leftarrow BOOKING * GUEST$
$TR \leftarrow (PHR \bowtie_{PHR.Hotel\_No=GR.Hotel\_No \ AND \ PHR.Room\_No=GR.Room\_No} GR)$
(Note — the Join in the previous line is the left outer join!)
$RESULT \leftarrow \pi_{Room\_No,Type,Price,GUEST.Name}(TR)$

4. (a) R=30+9+40+9+8+1+4+4+4+3+1=113B

(b) $bfr = \lfloor \frac{B}{R} \rfloor = \lfloor \frac{512}{113} \rfloor = 4$ records per block

$b = \lceil \frac{r}{bfr} \rceil = \lceil \frac{20000}{4} \rceil = 5,000$ blocks

(c) For linear search, on average we search half the blocks (2,500 blocks)

(d) For binary search, $\lceil (log_2 b) \rceil = \lceil (log_2 5000) \rceil = 13$

5. (a) R = 115 B

(b) $bfr = \lfloor \frac{B}{R} \rfloor = \lfloor \frac{512}{115} \rfloor = 4$ records per block

$b = \lceil \frac{30000}{4} \rceil = 7,500$

(c) Index entry size: $R_i = 9 + 6 = 15B$
$bfr_i = \lfloor \frac{B}{R_i} \rfloor = \lfloor \frac{512}{15} \rfloor = 34$ index entries per block

$r_i = 7,500$ entries (the number of blocks in the file)
$b_1 = \lceil \frac{7500}{34} \rceil = 221$ blocks

Multi-level index:
$r_2 = 221$ entries
$b_2 = \lceil \frac{221}{34} \rceil = 7$ blocks
$r_3 = 7$
$b_3 = 1$ block
Therefore, the number of levels is 3. The total number of blocks for the index is 221 + 7 + 1 = 229 blocks.
The number of block accesses to search for a record = 3 + 1 = 4

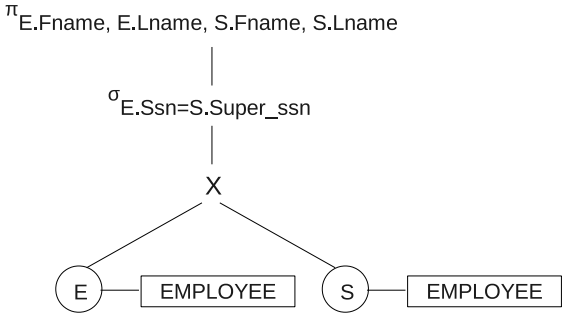
(d) Index entry size is still 15B, and the blocking factor is 34.
$r_1 = 30000$ (dense index)
$b_1 = 883$ blocks
The number of levels:
$r_2 = 883$
$b_2 = \lceil \frac{883}{34} \rceil = 26$
$r_3 = 26$
$b_3 = 1$

The total number of blocks for the index is 883 + 26 + 1 = 910
The number of block accesses is 3 + 1 = 4

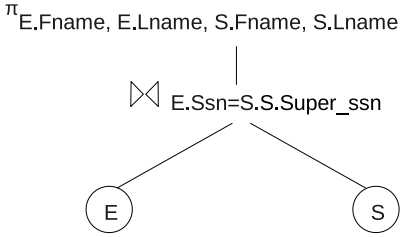
(e) For internal nodes:
$p * P + (p-1) * V_{SSN} <= B$
$6p + 9(p-1) <= 512$
$p = 34$
For leaf nodes: $p_{leaf} * (9+7) + 6 <= 512$
$16p_{leaf} <= 506$
$p_{leaf} = 31$

6. (a) The relational algrebra expression for the first SQL query:
$E \leftarrow EMPLOYEE$
$S \leftarrow EMPLOYEE$
$\pi_{E.Fname,E.Lname,S.Fname,S.Lname}(E \bowtie_{E.Sss=S.Super\_ssn} S)$

The initial (canonical) query tree:

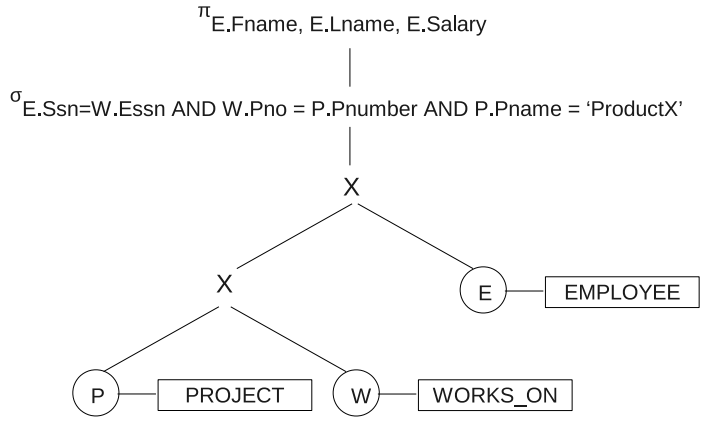


Another (slightly optimised) query tree:



(b) The relational algrebra expression for the second SQL query:
$E \leftarrow EMPLOYEE$
$P \leftarrow \sigma_{Pname='ProductX'}(PROJECT)$
$\pi_{Fname,Lname,Salary}((P \bowtie_{Pnumber=Pno} WORKS\_ON) \bowtie_{Essn=Ssn} E)$

The initial (canonical) query tree:



Another (slightly optimised) query tree: