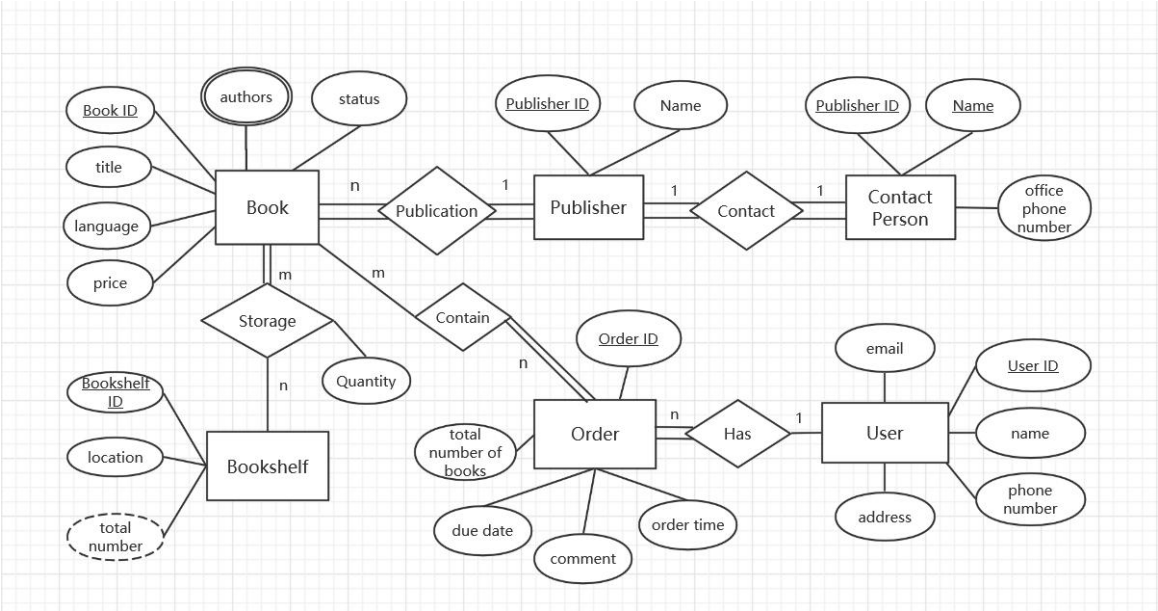
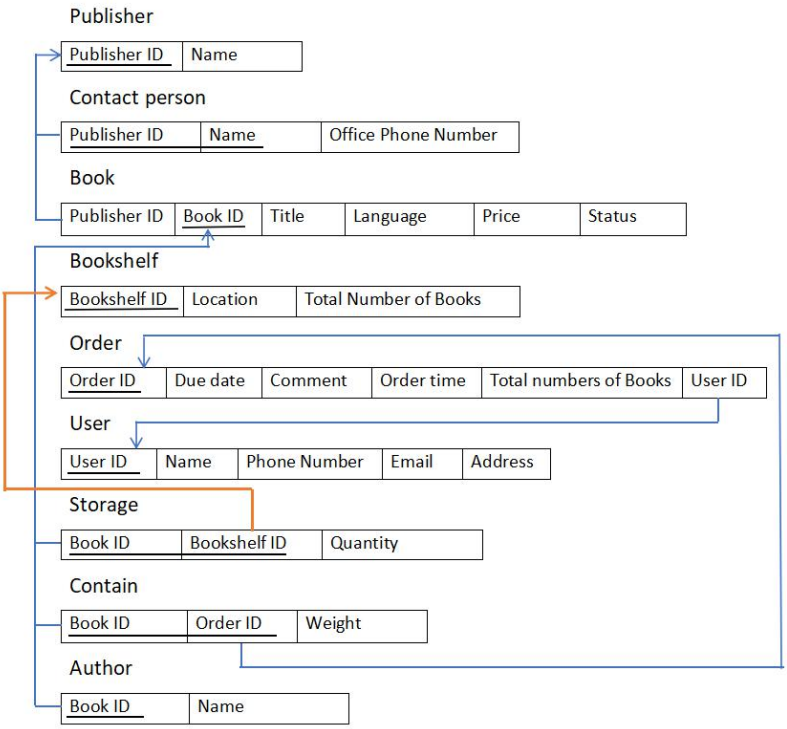
zid: z5239803 name: Zhengyang Jia

Question 1

(a)



(b)



Question 2

1) F = {B → CH, BCD → HI, EI → H, H → AB, I → E}.

Step 1: F'= {B →C, B →H, BCD →H, BCD →I, EI →H, H →A, H →B, I →E}

Step 2: check for each relation for its redundancy.

Because B →H, thus BCD →H is redundant.

Because B →C, thus BCD →H is equal to BD →I.

Because I →E, thus IE →H can be replaced to E →H.

Thus, F_{min} = {B →C, B →H, BD →I, E →H, H →A, H →B, I →E}

2) Because D only exists in the left side and G does not show up in F, the attributes D and G must be included in each candidate key.

{A, D, G}+={ A, D, G }=R. Thus, ADG is not a candidate key.

{B, D, G}+={A, B, C, D, E, G, H, I} =R. Thus, BDG is a candidate key.

{C, D, G}+={ C, D, G }=R. Thus, CDG is not a candidate key.

{E, D, G}+={ E, D, G }=R. Thus, EDG is not a candidate key.

{H, D, G}+={ A, B, C, D, E, G, H, I }=R. Thus, HDG is a candidate key.

{I, D, G}+={ A, B, C, D, E, G, H, I }=R. Thus, IDG is a candidate key.

Thus, the candidate keys are BDG, HDG, IDG.

3) It's lossless-join.

	A	B	C	D	E	G	H	I
R1	a	a	a	a	b	b	b	b
R2	b	b	b	a	a	a	a	a

	A	B	C	D	E	G	H	I
R1	a	a	a	a	a	b	a	a
R2	a	a	a	a	a	a	a	a

Because R2 can be changed to all 'a' which means it can restore R.

Also because R2 contains the candidate key DGH.

4)

R is 1NF, Because candidate keys are BDG, HDG and IDG, because B→CH, non-prime attribute C,H is partially functionally dependent on BDG, which violates 2NF. Because attributes of R are atomic, thus R is 1NF

Because candidate keys are BDG, HDG, IDG, all left parts of F are not super keys, so we can decompose R by sequence in F.

Consider B → CH, split R into $R1$= {B, C, H}, $R2$ = {A, B, D, E, G, I}

Consider H → C in $R1$, split $R1$ into $R11$ = {H, B}, $R12$ = {H, C}.

Consider BD → I in $R2$, split $R2$ into $R21$= {B, D, I}, $R3$ = {A, B, D, E, G}.

Consider BD → E in $R3$, split $R3$ into $R31$ = {B, D, E}, $R4$= {A, B, D, G}.

Consider B → A in $R4$, split $R4$ into $R41$ = {B, A}, $R42$ = {B, D, G}.

Because $R42$ is a candidate key, so the decomposition is lossless.

One possible lossless-decomposition into BCNF is :

$R11${H, B}, $R12${H, C}, $R21${B, D, I}, $R31${B, D, E}, $R41${B, A}, $R42${B, D, G}.

Question 3

1)

$A \leftarrow \sigma_{(age \geq 65 \text{ or } age \leq 24)} Visitor$

$B \leftarrow \pi_{\{pID\}}(A \bowtie Visit \bowtie Park)$

2)

$A \leftarrow Visit \div (\pi_{\{pID\}} Park)$

3)

$A \leftarrow \pi_{\{pID,\ location\}}(Park \bowtie Visit \bowtie (\sigma_{(name=\text{'Daniel'})} Visitor)$

$B \leftarrow \pi_{\{pID,\ location\}}(Park \bowtie Visit \bowtie (\sigma_{(name=\text{'James'})} Visitor)$

$C \leftarrow (\text{A} - \text{B}) \cup (B - A)$

4)

$A \leftarrow \pi_{\{vID,\ age\}}(Visitor \bowtie Visit \bowtie (\sigma_{(location=\text{'Hyde Park'})} Park)$

$B \leftarrow \pi_{\{vID,\ age\}}(Visitor \bowtie Visit \bowtie (\sigma_{(location=\text{'Hyde Park'})} Park)$

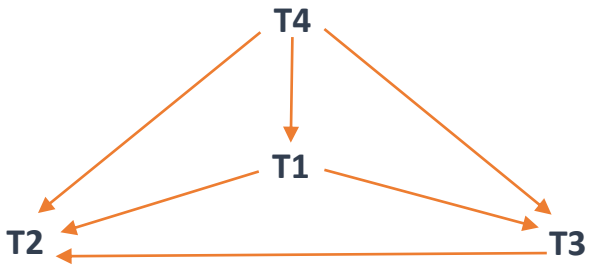
$C \leftarrow \text{A} \bowtie (\sigma_{(A.age > B.age)} B)$

$D \leftarrow \pi_{\{vID\}}(A) - \pi_{\{B.vID\}}(C)$

Question 4

(a)

1)



2) It is conflict serializable because the graph is not cyclic

Equivalent serial schedule:

Time	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
T_1			W(Y)	R(X)					
T_2								R(Y)	W(Z)
T_3					R(Z)	R(Y)	R(X)		
T_4	W(Y)	W(X)							

(b)

1) T3 and T4 will have a dead lock, T3 will wait for B to be unlocked while T4 will wait for C to be unlocked.

Question 5

(a)

Use a clustered B+ tree index on attributes (R.a, R.b) is the cheapest.

Because:

1). B+ tree is dynamic and don't has overflow pages, while hashing may have long overflow chains which degrade the efficiency.

2) B+ tree has better performance in range query when relation R is already sorted. 3).

3). As mentioned in question text, we only consider index-only plans, which means we can get results without visiting data records.

(b)

1) Consider the capacity of the buffer pool is 2 and the request frame sequence is 1232131, for FIFO it changes 2 times, for LRU is 3 times, and for MRU is also 3 times.

2) Consider the capacity of the buffer pool is 2 and the request frame sequence is 1213131, for LRU it changes 1 time, for FIFO it changes 2 times and for MRU it changes 4 times.

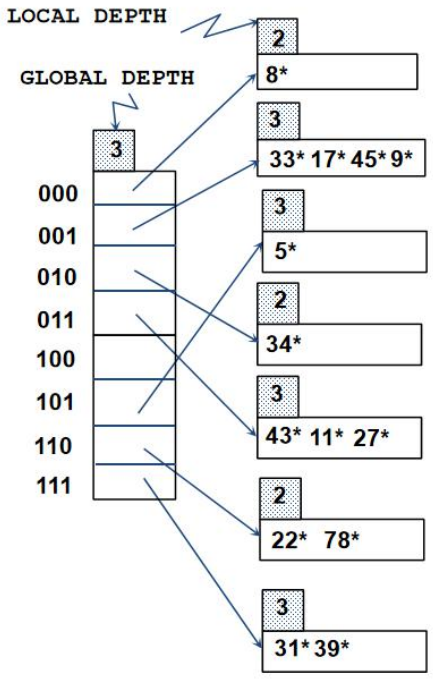
Question 6

(a)

1). The largest number less than 50 which will cause a split is 47.

According to the chart, only number in binary ending with 001 or 011 will cause a split, from 49 go down we can find 47 is the largest which ends with 011.

2). After inserting 9, 27, 78:

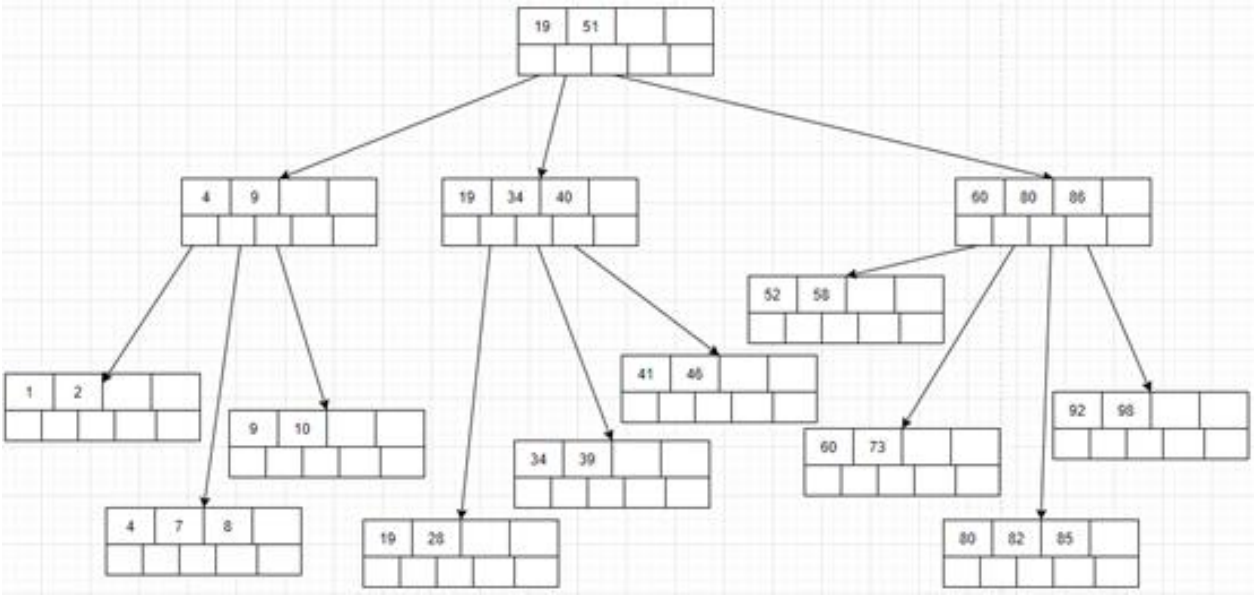


(b)

1). Minimum number= 14

Considering adding 3, 20, 21, 22, 23, 24, 25, 26, 27, 29, 30, 31, 32, 33, the B+ tree will increase depth.

2).



3).