CSC343	Introduction to Database

Assignment 3:	database (re)design
Group member:	Weizhou Wang 1004421262
	Lingwei Sun 1004723276

1. Relation Reservation is meant to keep track of which skipper reserves which craft on which date, but it's design has some redundancy:

$$Reservation(sID,\ age,\ length,\ sName,\ day,\ cName,\ rating,\ cID)$$

. . . where sID identifies the skipper, sName is the skipper's name, whereas rating and age record the skipper's skill (a number between 0 and 5, inclusive) and age (a number greater than 0). The reserved craft is identified by cID, its name is cName, length is in feet, and the date and time the craft is reserved for by the skipper is day. The following dependencies hold:

$$S = \{sID \rightarrow sName,\ rating,\ age\ ;\ cID \rightarrow cName, length\}$$

(a) Give one example of redundancy that relation Reservation, combined with FDs S, allow.

Reservation

sID	age	length	sName	day	cName	rating	cID
1	18	10	Adam	2021-02-24 12:12:12	Audi	3	1
1	18	20	Adam	2021-02-25 13:12:12	BMW	3	2
1	18	30	Adam	2021-02-26 14:12:12	Porsche	3	3

Redundant data is data that can be inferred from other tuples in the relation based on the FDs. We can see that there are redundancies on attributes age, sName and rating. Since there is an FD of $sID \rightarrow sName,\ rating,\ age$, if the sIDs of some tuples are the same, their ages, sNames and ratings must also be the same. Therefore, there is no need to present age, sName and rating multiple times for the same sID, and the above relations have redundancies.

b) Use BCNF decomposition to reduce the redundancies.

First, we found that in the FD, $sID \rightarrow sName,\ rating,\ age$,
$sID^{+} = \{sID,\ sName,\ rating,\ age\}$. Since sID is not a superkey, this FD violates BCNF.

- Decompose Relation using FD $sID \rightarrow sName,\ rating,\ age$.
  $sID^{+} = \{sID,\ sName,\ rating,\ age\}$, so this yields two relations:

$sID^{+} = \{sID,\ sName,\ rating,\ age\}$

$Reservation - (sID^{+} - sID)$

$= \{sID,\ age,\ length,\ sName,\ day,\ cName,\ rating,\ cID\} - \{sID,\ sName,\ rating,\ age\} + \{sID\}$

$= \{sID,\ day,\ cID,\ cName,\ length\}$

So we get: $R1(sID,\ sName,\ rating,\ age)$ and $R2(sID,\ day,\ cID,\ cName,\ Length)$

- Project the FDs onto $R1(sID,\ sName,\ rating,\ age)$

sID	sName	rating	age	closure	FDs
✓				$sID^+ = \{sID,\ sName,\ rating,\ age\}$	$sID \rightarrow sName,\ rating,\ age$; sID as a super key of R1
	✓			$sName^+ = sName$	nothing
		✓		$rating^+ = rating$	nothing
			✓	$age^+ = age$	nothing
supersets of sID				irrelevant	can only generate weaker FDs than what we already have
	✓	✓		$\{sName, rating\}^+ = \{sName,\ rating\}$	nothing
	✓		✓	$\{sName, age\}^+ = \{sName,\ age\}$	nothing
		✓	✓	$\{rating, age\}^+ = \{rating,\ age\}$	nothing
	✓	✓	✓	$\{sName, rating, age\}^+ = \{sName,\ rating,\ age$	nothing

This relation satisfies BCNF.

- Project the FDs onto $R2(sID,\ day,\ cID,\ cName,\ length)$

sID	day	cID	cName	length	closure	FDs
		✓			$cID^+ = \{cID,\ cName,\ length\}$	$cID \rightarrow cName, length$: violates BCNF; abort the projection

We must decompose R2 further.

- Decompose Relation using FD $cID \rightarrow cName, length. cID^+ = \{cID,\ cName,\ length\}$, so this yields two relations:

$cID^+ = \{cID,\ cName,\ length\}$

$$
\begin{aligned}
R2 - (cID^+ - cID) &= \{sID,\ day,\ cID,\ cName, length\} - \{cID,\ cName,\ length\} + \{cID\} \\
&= \{sID,\ day,\ cID\}
\end{aligned}
$$

So we get: $R3(cID,\ cName,\ length)$ and $R4(sID,\ day,\ cID)$

- Project the FDs onto $R3(cID,\ cName,\ length)$

cID	cName	length	closure	FDs
✓			$cID^+ =$ $\{cID,\ cName,\ length\}$	$cID \rightarrow cName, length$; cID is a superkey of R3
	✓		$cName^+ = cName$	nothing
		✓	$length^+ = length$	nothing
supersets of cID			irrelevant	can only generate weaker FDs than what we already have
	✓	✓	$\{cName, length\}^+ =$ $\{cName,\ length\}$	nothing

This relation satisfies BCNF.

- Project the FDs onto $R4(sID,\ day,\ cID)$

sID	day	cID	closure	FDs
✓			$sID^+ = \{sID,\ sName,\ rating,\ age\}$ sName, rating, age not in R4	nothing
	✓		$day^+ = day$	nothing
		✓	$cID^+ = \{cID,\ cName,\ length\}$ cName, length not in R4	nothing
✓	✓		$\{sID, day\}^+ = \{sID,\ sName,\ rating,\ age,\ day\}$ sName, rating, age not in R4	nothing
✓		✓	$\{sID, cID\}^+ = \{sID,\ sName,\ rating,\ age,$ $cID,\ cName,\ length\}$ sName, rating, age, cName, length not in R4	nothing
	✓	✓	$\{cID, day\}^+ = \{cID,\ cName,\ length,\ day\}$ cName, length not in R4	nothing

This relation satisfies BCNF.

- Final decomposition:

(a) $R1(sID,\ sName,\ rating,\ age)$ with FD $sID \rightarrow sName,\ rating,\ age$, named R1 as skipper: $skipper(sID,\ sName,\ rating,\ age)$

(b) $R3(cID,\ cName,\ length)$ with FD $cID \rightarrow cName, length$, named R2 as craft: $craft(cID,\ cName,\ length)$

(c) $R4(sID,\ day,\ cID)$ with no FDs, named R4 as reserve: $reserve(sID,\ day,\ cID)$

- BCNF decomposition method guarantees redundancy-prevention and lossless join. So the schema created based on the final decomposition results will have “as few redundancies as possible” as required.

skipper(<u>sID</u>, sName, rating, age)
craft(<u>cID</u>, cName, length)
reserve(sID, day, cID)

reserve[sID]⊆skipper[sID]
reserve[cID]⊆craft[cID]

Please see reservation.ddl about the implementation of the schema.

2. Relation F has attributes $KLMNOPQRS$ and functional dependencies G:

$$G = \{KOQ \rightarrow PS,\ L \rightarrow KN,\ KQ \rightarrow RS\}$$

(a) Which FDs in G violates BCNF? List them.

BCNF requires that the LHS of an FD be a superkey.

- $KOQ^+ = \{K,\ O,\ Q,\ P,\ S,\ R\} = KOPQRS$, so KOQ is not a super key as KOQ^+ does not contain LMN, and $KOQ \rightarrow PS$ violates BCNF.
- $L^+ = \{L,\ K,\ N\} = KLN$, so L is not a super key as L^+ does not have MOPQRS and $L \rightarrow KN$ violates BCNF.
- $KQ^+ = \{K,\ Q,\ R,\ S\} = KQRS$, so KQ is not a super key and $KQ \rightarrow RS$ violates BCNF.

(b) Use the BCNF decomposition method to derive a redundancy-preventing, lossless, decomposition of F into a new schema consisting of relations that are in BCNF. Be sure to project the FDs from G onto the relations in your final schema. There may be more than one correct answer possible since there are choices possible at steps in the decomposition. List your final relations alphabetically, and order the attributes within each relation alphabetically (this avoids combinatorial explosion of the number of alternatives we have to check).

- Decompose F using FD $KOQ \rightarrow PS$. $KOQ^+ = \{K,\ O,\ Q,\ P,\ S,\ R\} = KOPQRS$, so this yields two relations:

$KOQ^+ = \{K,\ O,\ Q,\ P,\ S,\ R\} = KOPQRS$

$F - (KOQ^+ - KOQ) = \{KLMNOPQRS\} - \{KOPQRS\} + \{KOQ\} = KLMNOQ$

So we get: $R1(K, O, P, Q, R, S)$ and $R2(K, L, M, N, O, Q)$

- Project the FDs onto $R1(K, O, P, Q, R, S)$

K	O	P	Q	R	S	closure	FDs
✓			✓			$KQ^+ = \{K,\ Q,\ R,\ S\}$ $= KQRS$	$KQ \rightarrow RS$: violates BCNF; abort the projection

We must decompose R1 further.

- Decompose R1 using FD $KQ \rightarrow RS$. $KQ^+ = \{K,\ Q,\ R,\ S\} = KQRS$, so this yields two relations:

$KQ^+ = \{K, Q, R, S\} = KQRS$

$R1 - (KQ^+ - KQ) = \{KOPQRS\} - \{KQRS\} + \{KQ\} = KOPQ$

So we get: $R3(K, Q, R, S)$ and $R4(K, O, P, Q)$

- Project the FDs onto $R3(K, Q, R, S)$

K	Q	R	S	closure	FDs

✓				$K^+ = K$	nothing
	✓			$Q^+ = Q$	nothing
		✓		$R^+ = R$	nothing
			✓	$S^+ = S$	nothing
✓	✓			$KQ^+ = KQRS$	$KQ \to RS$; KQ is a superkey of R3
✓		✓		$KR^+ = KR$	nothing
✓			✓	$KS^+ = KS$	nothing
	✓	✓		$QR^+ = QR$	nothing
	✓		✓	$QS^+ = QS$	nothing
		✓	✓	$RS^+ = RS$	nothing
supersets of KQ				irrelevant	can only generate weaker FDs than what we already have
✓		✓	✓	$KRS^+ = KRS$	nothing
	✓	✓	✓	$QRS^+ = QRS$	nothing

This relation satisfies BCNF.

- Project the FDs onto $R4(K, O, P, Q)$

K	O	P	Q	closure	FDs
✓				$K^+ = K$	nothing
	✓			$O^+ = O$	nothing
		✓		$P^+ = P$	nothing
			✓	$Q^+ = Q$	nothing
✓	✓			$KO^+ = KO$	nothing
✓		✓		$KP^+ = KP$	nothing
✓			✓	$KQ^+ = KQRS$, RS not in R4	nothing
	✓	✓		$OP^+ = OP$	nothing

	✓		✓	$OQ^+ = OQ$	nothing
		✓	✓	$PQ^+ = PQ$	nothing
✓	✓	✓		$KOP^+ = KOP$	nothing
✓	✓		✓	$KOQ^+ = KOPQRS$	$KOQ \rightarrow P$; KOQ is a superkey of R4
✓		✓	✓	$KPQ^+ = KPQ$	nothing
	✓	✓	✓	$OPQ^+ = OPQ$	nothing

This relation satisfies BCNF.

- Project the FDs onto $R2(K, L, M, N, O, Q)$

K	L	M	N	O	Q	closure	FDs
	✓					$L^+ = \{L,\ K,\ N\} = KLN$	$L \rightarrow KN$: violates BCNF; abort the projection

We must decompose R2 further.

- Decompose R2 using FD $L \rightarrow KN$. $L^+ = \{L,\ K,\ N\} = KLN$, so this yields two relations:

$R5 = L^+ = \{L,\ K,\ N\} = KLN$

$R6 = R2 - (L^+ - L) = \{KLMNOQ\} - \{KLN\} + \{L\} = LMOQ$

So we get: $R5(K, L, N)$ and $R6(L, M, O, Q)$

- Project the FDs onto $R5(K, L, N)$

K	L	N	closure	FDs
✓			$K^+ = K$	
	✓		$L^+ = \{L,\ K,\ N\} = KLN$	$L \rightarrow KN$: L is a superkey of R5
		✓	$N^+ = N$	
supersets of L			irrelevant	can only generate weaker FDs than what we already have
✓		✓	$KN^+ = KN$	

This relation satisfies BCNF.

- Project the FDs onto $R6(L, M, O, Q)$

L	M	O	Q	closure	FDs

✓				$L^+ = \{L, K, N\} = KLN$, KN not in R6	nothing
	✓			$M^+ = M$	nothing
		✓		$O^+ = O$	nothing
			✓	$Q^+ = Q$	nothing
✓	✓			$LM^+ = KLMN$, KN not in R6	nothing
✓		✓		$LO^+ = KLNO$, KN not in R6	nothing
✓			✓	$LQ^+ = KLNQ$, KN not in R6	nothing
	✓	✓		$MO^+ = MO$	nothing
	✓		✓	$MQ^+ = MQ$	nothing
		✓	✓	$OQ^+ = OQ$	nothing
✓	✓	✓		$LMO^+ = KLMNO$, KN not in R6	nothing
✓	✓		✓	$LMQ^+ = KLMNQ$, KN not in R6	nothing
✓		✓	✓	$LOQ^+ = KLNOQ$, KN not in R6	nothing
	✓	✓	✓	$MOQ^+ = MOQ$	nothing

This relation satisfies BCNF.

- Final decomposition (in alphabetical order):

(a) $R5(K, L, N)$ with FD $L \rightarrow KN$
(b) $R4(K, O, P, Q)$ with FD $KOQ \rightarrow P$
(c) $R3(K, Q, R, S)$ with FD $KQ \rightarrow RS$
(d) $R6(L, M, O, Q)$ with no FDs.

- BCNF decomposition method guarantees redundancy-prevention and lossless join.

(c) Does your final schema preserve dependencies? Explain why you answer yes or no.

The FDs in the above decomposition relations are $G' = \{KQ \rightarrow RS,\ KOQ \rightarrow P,\ L \rightarrow KN\}$, and the FDs in the original relations are $G = \{KOQ \rightarrow PS,\ L \rightarrow KN,\ KQ \rightarrow RS\} = \{KOQ \rightarrow P, KOQ \rightarrow S, L \rightarrow KN, KQ \rightarrow R, KQ \rightarrow S\}$. Note that $KOQ \rightarrow S$ is a weaker FD than $KQ \rightarrow S$, so it is equivalent to say $G = \{KOQ \rightarrow P, L \rightarrow KN, KQ \rightarrow R, KQ \rightarrow S\} = \{KQ \rightarrow RS,\ KOQ \rightarrow P,\ L \rightarrow KN\} = G'$. Therefore, these two sets of FDs are the same and the final schema preserves dependencies.

(d) BCNF guarantees a lossless join. However, demonstrate this to a possibly-skeptical observer using the Chase Test.

Suppose $t = (k, l, m, n, o, p, q, r, s) \in R3 \bowtie R4 \bowtie R5 \bowtie R6$, so we get a tableau of:

K	L	M	N	O	P	Q	R	S
k	l1	m1	n1	o1	p1	q	r	s
k	l2	m2	n2	o	p	q	r2	s2
k	l	m3	n	o3	p3	q3	r3	s3
k4	l	m	n4	o	p4	q	r4	s4

Since $L \rightarrow KN$, the tableau should be changed to:

K	L	M	N	O	P	Q	R	S
k	l1	m1	n1	o1	p1	q	r	s
k	l2	m2	n2	o	p	q	r2	s2
k	l	m3	n	o3	p3	q3	r3	s3
k	l	m	**n**	o	p4	q	r4	s4

Since $KQ \rightarrow RS$:

K	L	M	N	O	P	Q	R	S
k	l1	m1	n1	o1	p1	q	r	s
k	l2	m2	n2	o	p	q	**r**	**s**
k	l	m3	n	o3	p3	q3	r3	s3
k	l	m	n	o	p4	q	**r**	**s**

Since $KOQ \rightarrow PS$:

K	L	M	N	O	P	Q	R	S
k	l1	m1	n1	o1	p1	q	r	s
k	l2	m2	n2	o	p	q	r	s
k	l	m3	n	o3	p3	q3	r3	s3
k	l	m	n	o	**p**	q	r	s

We got a completely unsubscripted row, so $t \in F(K, L, M, N, O, P, Q, R, S)$, and the decomposition is lossless.

3. Relation R has attributes $ABCDEFGH$ and functional dependencies S:

$$S = \{ACDE \to B,\ B \to CF,\ CD \to AF,\ BCF \to AD,\ ABF \to H\}$$

(a) Find a minimal basis for S. Your final answer must put the FDs in ascending alphabetical order, and the attributes within the LHS and RHS of each FD into alphabetical order.

- Split RHS and name the set S1:

1. $ACDE \to B$
2. $B \to C$
3. $B \to F$
4. $CD \to A$
5. $CD \to F$
6. $BCF \to A$
7. $BCF \to D$
8. $ABF \to H$

- Reduce LHS:

FD	Closure	Decision
1	$A^+ = A$ $C^+ = C$ $D^+ = D$ $E^+ = E$ $AC^+ = AC$ $AD^+ = AD$ $AE^+ = AE$ $CD^+ = ACDF$ $CE^+ = CE$ $DE^+ = DE$ $ACD^+ = weak\ FD$ $ACE^+ = ACE$ $ADE^+ = ADE$ $CDE^+ = ABCDEF$	$CDE \to B$
2	only one attribute on LHS	keep
3	only one attribute on LHS	keep
4	$C^+ = C$ $D^+ = D$	keep
5	$C^+ = C$ $D^+ = D$	keep

6	$B^{+} = \{B, C, F, A, D, H\} = ABCDFH$ $C^{+} = C$ $F^{+} = F$ $BC^{+} = weak\ FD$ $BF^{+} = weak\ FD$ $CF^{+} = CF$	$B \rightarrow A$
7	$B^{+} = \{B, C, F, A, D, H\} = ABCDFH$ $C^{+} = C$ $F^{+} = F$ $BC^{+} = weak\ FD$ $BF^{+} = weak\ FD$ $CF^{+} = CF$	$B \rightarrow D$
8	$A^{+} = A$ $B^{+} = \{B, C, F, A, D, H\} = ABCDFH$ $F^{+} = F$ $AB^{+} = weak\ FD$ $AF^{+} = AF$ $BF^{+} = weak\ FD$	$B \rightarrow H$

- Updated FDs and name the set S2:

1. $CDE \rightarrow B$
2. $B \rightarrow C$
3. $B \rightarrow F$
4. $CD \rightarrow A$
5. $CD \rightarrow F$
6. $B \rightarrow A$
7. $B \rightarrow D$
8. $B \rightarrow H$

- Look for redundant FDs to eliminate.

FD	Closure	Decision
1	$CDE^{+}_{S2-(1)} = \{A, C, D, E, F\} = ACDEF$	keep
2	$B^{+}_{S2-(2)} = \{B, F, A, D, H\} = ABDFH$	keep
3	$B^{+}_{S2-(3)} = \{B, C, A, D, H, F\} = ABCDFH$	remove (3)
4	$CD^{+}_{S2-(3)-(4)} = \{C, D, F\} = CDF$	keep

5	$CD^{+}_{S2-(3)-(5)} = \{C, D, A\} = ACD$	keep
6	$B^{+}_{S2-(3)-(6)} = \{B, C, D, H, A, F\} = ABCDFH$	remove (6)
7	$B^{+}_{S2-(3)-(6)-(7)} = \{B, C, H\} = BCH$	keep
8	$B^{+}_{S2-(3)-(6)-(8)} = \{B, C, D, A, F\} = ABCDF$	keep

- The minimal basis for S, called set S3, is:
  1. $CDE \rightarrow B$
  2. $B \rightarrow C$
  4. $CD \rightarrow A$
  5. $CD \rightarrow F$
  7. $B \rightarrow D$
  8. $B \rightarrow H$

(b) Find all the keys for R using your solution for a minimal basis.

Attribute	Appears on		Conclusion
	LHS	RHS	
A	–	✓	must not in any key
B	✓	✓	must check
C	✓	✓	must check
D	✓	✓	must check
E	✓	–	must be in every key
F	–	✓	must not in any key
G	–	–	must be in every key
H	–	✓	must not in any key

1. $BEG^{+} = \{B, E, G, C, D, H, A, F\} = ABCDEFGH$ So, BEG is a key.
2. $CEG^{+} = \{C, E, G\} = CEG$ This is not a key.
3. $DEG^{+} = \{D, E, G\} = DEG$ This is not a key.
4. $BCEG^{+} = weaker\ key$
5. $BDEG^{+} = weaker\ key$
6. $CDEG^{+} = \{C, D, E, G, A, F, B, H\} = ABCDEFGH$ So, CDEG is a key.

7. $BCDEG^{+} = weaker\ key$

All the keys for R using the solution for a minimal basis are BEG and $CDEG$.

(c) Use the 3NF synthesis algorithm to find a lossless, dependency-preserving decomposition of relation R into a new schema consisting of relations that are in 3NF. Your final answer should combine FDs with the same LHS to create a single relation. If your schema has a relation that is a subset of another, keep only the larger relation.

- Revised FD and name as S4 (in alphabetical order)
  $B \rightarrow CDH$
  $CD \rightarrow AF$
  $CDE \rightarrow B$

- The set of relations that would result to have these attributes:

$R1(B, C, D, H),\ R2(A, C, D, F),\ R3(B, C, D, E)$

- Since there is no relation as a superkey of R, so we add a relation BEG. The final set of relations is:

$R1(B, C, D, H),\ R2(A, C, D, F),\ R3(B, C, D, E),\ R4(B, E, G)$

- 3NF synthesis algorithm guarantees lossless join and dependency preservation.

(d) Does your solution allow redundancy? Explain how (with an example), or why not.

According to the week9_3NF worksheet, "Because we formed each relation from an FD, the LHS of those FDs are indeed superkeys for their relations. However, there may be other FDs that violate BCNF and therefore allow redundancy. The only way to find out is to project the FDs onto each relation."

To project FDs on R3, we need to compute the closure for each subset of {B,C,D,E}. When computing the closure of B, B^{+}, we found that $B^{+} = \{A, B, C, D, F, H\}$, so the projection onto R3 is $B \rightarrow CD$. However, B is not a super key of R3 (its closure does not have E), so the schema violates BCNF and allows redundancy.