UNIT 2

Lecture 36

Normalization

Lossless join Decomposition and Difference between 3NF and BCNF

Loss less Join Decomposition

Q. 1 Consider R = (A, B, C, D, E) with

F = { AD → C,

CD → A,

B → D,

D → BE }.

Determine whether {ABC, BCD, DE} is loss-less. Find all keys for R.

Algorithm 15.3. Testing for Nonadditive Join Property

Input: A universal relation R, a decomposition $D = \{R_1, R_2, \ldots, R_m\}$ of R, and a set F of functional dependencies.

Note: Explanatory comments are given at the end of some of the steps. They follow the format: (**comment**).

1. Create an initial matrix S with one row i for each relation R_i in D, and one column j for each attribute A_j in R.
2. Set $S(i, j) := b_{ij}$ for all matrix entries. (**Each b_{ij} is a distinct symbol associated with indices (i, j)**)
3. For each row i representing relation schema R_i
   {for each column j representing attribute A_j
   {if (relation R_i includes attribute A_j) then set $S(i, j) := a_j$;};}; (**Each a_j is a distinct symbol associated with index (j)**)
4. Repeat the following loop until a *complete loop execution* results in no changes to S
   {for each functional dependency $X \rightarrow Y$ in F
   {for all rows in S *that have the same symbols* in the columns corresponding to attributes in X
   {make the symbols in each column that correspond to an attribute in Y be the same in all these rows as follows: If any of the rows has an a symbol for the column, set the other rows to that *same* a symbol in the column. If no a symbol exists for the attribute in any of the rows, choose one of the b symbols that appears in one of the rows for the attribute and set the other rows to that same b symbol in the column ;} ; } ;};
5. If a row is made up entirely of a symbols, then the decomposition has the nonadditive join property; otherwise, it does not.

Loss less Join Decomposition

Q. 2 Consider R = (A, B, C, D) with

F = { A → B,

C → D }.

Determine whether {AB, CD, AC} is loss-less.

Why BCNF is considered as a stronger normal form of 3NF

- BCNF is considered as a stronger normal form of 3NF because 3NF relations still have insertion, deletion and update anomalies but BCNF relations doesn’t have any insertion, deletion and update anomalies.
- In BCNF every dependency is on the candidate key which is our goal of normalization and in 3NF it is not required that every dependency is on candidate key.

Why BCNF is considered as a stronger normal form of 3NF

- For e.g., consider a relation Stud_Course with the following instance with the fact that a particular student takes a course and has one instructor and a particular instructor teaches one course only.

Sid	Cid	Tid
S1	C1	T1
S1	C2	T2
S2	C1	T3
S2	C2	T2

- So, from the instance we have

  Sid, Cid → Tid

  Tid → Cid

Why BCNF is considered as a stronger normal form of 3NF

- So, the relation Stud_Course is in 3NF, because there is no transitive dependencies exist in the relation. (As there are no non – prime attributes are there in the relation)
- The relation Stud_Course in not BCNF, because it has a functional dependency **Tid → Cid**, and Tid is not super key or candidate key of the relation.
- Now, The relation Stud_Course still have insertion, deletion and update anomalies, so we need to further normalized to minimize these anomalies.

Why BCNF is considered as a stronger normal form of 3NF

- So, we decompose Stud_Course relation into Stud and Teacher as follows.

Stud_Course

Sid	Cid	Tid
S1	C1	T1
S1	C2	T2
S2	C1	T3
S2	C2	T2

Stud

Sid	Tid
S1	T1
S1	T2
S2	T3
S2	T2

Teacher

Cid	Tid
C1	T1
C2	T2
C1	T3

- Now these relations don't have any insertion, deletion and update anomalies because every dependency is on the super key.
- The natural join between these two relations produce the original relations.

Why BCNF is considered as a stronger normal form of 3NF

- Suppose we decompose Stud_Course relation into Stud and Teacher as follows.

Stud_Course

Sid	Cid	Tid
S1	C1	T1
S1	C2	T2
S2	C1	T3
S2	C2	T2

Stud

Sid	Cid
S1	C1
S1	C2
S2	C1
S2	C2

Teacher

Cid	Tid
C1	T1
C2	T2
C1	T3

- The natural join between these two relations do not produce the original relation.
- Some extra tuples (Spurious Tuples) may be generated which cause loss of information.

Q.1 Consider the universal relation R = {A, B, C, D, E, F, G, H, I, J} and the set of FDs F = {AB→C, BD→EF, AD→GH, A→I, H→J}. What is the key for R? Decompose R into 2NF and then 3NF relations.

Sol : Given Relation

R = {A, B, C, D, E, F, G, H, I, J}

with a set of FDs

F = { AB → C,
BD → EF,
AD → GH,
A → I,
H → J }

Super key = ABCDEFGHIJ

Essential attribute for candidate key = ABD

Now, Check whether ABD is a candidate key or not, So we have to find ABD^+.

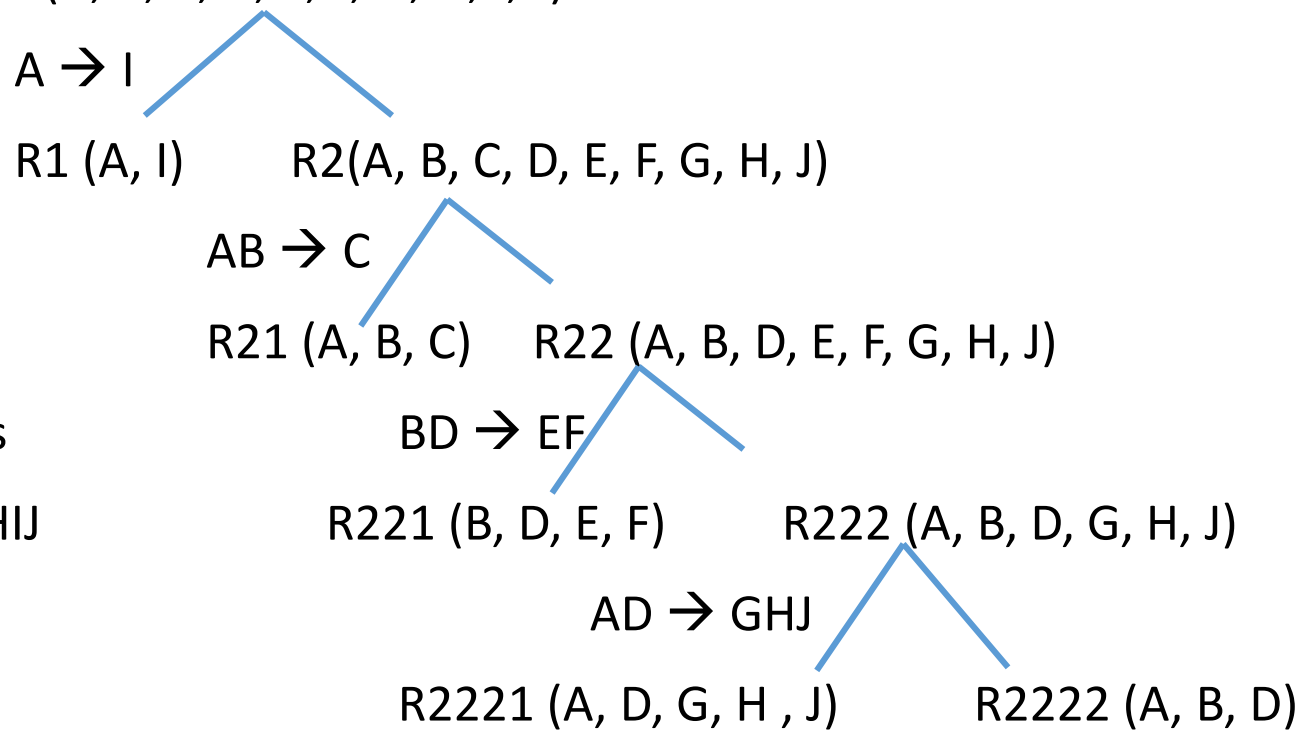
So, ABD^+ = ABCDEFGHIJ

Since ABD^+ contains all the attributes of relation R, So ABD is the only candidate key of this relation.

Q.1 Consider the universal relation R = {A, B, C, D, E, F, G, H, I, J} and the set of FDs F = {AB→C, BD→EF, AD→GH, A→I, H→J}. What is the key for R? Decompose R into 2NF and then 3NF relations.

To decompose R into 2NF first we need to find F^+.

F^+ = {
AB → CI,
BD → EF,
AD → GHIJ,
A → I,
H → J }



So, we have 4 partial dependencies

A → I, AB → CI, BD → EF, AD → GHIJ

So, R1 (A, I), R2 (A, B, C), R3 (B, D, E, F), R4 (A, D, G, H, J), and R5 (A, B, D) is the required 2NF decompositions of R.

Q.1 Consider the universal relation R = {A, B, C, D, E, F, G, H, I, J} and the set of FDs F = {AB→C, BD→EF, AD→GH, A→I, H→J}. What is the key for R? Decompose R into 2NF and then 3NF relations.

Now,

R1 (A, I) with F1 = {A → I}

R2 (A, B, C) with F2 = {AB → C}

R3 (B, D, E, F) with F3 = {BD → EF}

R5 (A, B, D) with F5 = { ø } is in 3NF, but

R4 (A, D, G, H, J) with

F4 = {AD → GHJ, H → J}

is not in 3NF, because

it has a transitive dependency

H → J, with candidate key AD.

To decompose R4 into 3NF, we need to find canonical cover $F4_C$.

So, $F4_C$ = {AD → GH, H → J}

So, R41 (A, D, G, H) and R42 (H, J) are the 3NF decomposition of R4.

So, finally R1 (A, I), R2 (A, B, C),

R3 (B, D, E, F), R4 (A, B, D), R5 (A, D, G, H) and R6 (H, J) are the required 3NF decompositions of R.

Question

Q.2 Consider the universal relation R = {A, B, C, D, E, F, G, H, I, J} and the set of FDs F = {AB→C, A→DE, B→F, F→GH, D→IJ}. What is the key for R? Decompose R into 2NF and then 3NF relations.

Difference between 3NF and BCNF

SNO.	3NF	BCNF
1	A relation schema R is in third normal form (3NF) if, whenever a nontrivial functional dependency X→A holds in R, either (a) X is a super key of R, or (b) A is a prime attribute of R.	A relation schema R is in BCNF if whenever a nontrivial functional dependency X→ A holds in R, then X is a super key of R.
2	If the relation do not contain any non prime attribute then the relation is in 3NF.	If the relation do not contain any non prime attribute then the relation may or may not in BCNF.
3	3NF is considered as weaker normal form of BCNF.	BCNF is considered as stronger normal form of 3NF.
4	Relations in 3NF have redundancies due to FDs.	Relations in BCNF do not have redundancies due to FDs.
5	It is always possible to find a dependency preserving decomposition with respect to a F such that the resulting relations are in 3NF.	In general, there may not be a dependency preserving decomposition that also decomposes relations in BCNF.
6	3NF relations still have anomalies.	BCNF relations do not have anomalies.
7	3NF relation can have non key attributes as determinants.	BCNF relation cannot have non key attributes as determinants.

GATE Questions

Relation R with an associated set of functional dependencies, F, is decomposed into BCNF. The redundancy (arising out of functional dependencies) in the resulting set of relations is

(A) Zero

(B) More than zero but less than that of an equivalent 3NF decomposition

(C) Proportional to the size of F^+

(D) Indetermine

[GATE 2002]

For Video lecture on this topic please subscribe to my youtube channel.

The link for my youtube channel is

https://www.youtube.com/channel/UCRWGtE76JlTp1iim6aOTRuw?sub_confirmation=1