Database Management Systems

Part III: Database Design

Lecture 11

Higher Normal Forms

Contents

- Introduction
- Multivalued Dependencies
- Fourth Normal Form
- Join Dependencies
- Fifth Normal Form
- Normalization Procedure Summarized
- Other Normal Forms

Introduction

- The concept of functional dependency is the most important type of dependency in relational database design theory and normal forms based on functional dependencies.
- However, in many cases, relations have constraints that cannot be specified as functional dependencies.
- This constraint is specified by a multivalued dependency.

Multivalued Dependencies

- Multivalued dependency occurs when the two attributes in a table are independent of each other but both depend on a third attribute.
- A multivalued dependency consists of at least two attributes that are dependent on a third attribute that is why it always requires at least three attributes.

Multivalued Dependencies (Cont.)

- Let R be a relation, and let A, B, and C be arbitrary subsets of the set of attributes of R.
- Then, we say that B is multidependent on A – in symbols,

$$A \twoheadrightarrow B$$

if and only if the set of B-values matching a given (A-value, C-value) pair in R depends only on the A-value and is independent of the C-value.

Multivalued Dependencies (Cont.)

Theorem (Fagin)

- Let R{A,B,C} be a relation, where A, B, and C are sets of attributes.
- Then, R is equal to the join of its projections on {A,B} and {A,C} if and only if R satisfies the MVD

$$A \twoheadrightarrow B \mid C$$

Multivalued Dependencies (Cont.)

Example

- Suppose there is a MEMBER relation as shown in the table.
- Here, columns AGE and CLUB are dependent on MEMBER_ID and independent of each other.
- In this case, these two columns can be called as multivalued dependents on MEMBER_ID.

MEMBER

MEMBER_ID	AGE	CLUB
1	18	Dancing
2	20	Dancing
3	16	Singing
4	18	Dancing
5	19	Singing

Multivalued Dependencies (Cont.)

Example

- The representation of these dependencies is shown below:

  MEMBER_ID $\twoheadrightarrow$ AGE

  MEMBER_ID $\twoheadrightarrow$ CLUB

- This can be read as MEMBER_ID multidetermined AGE and MEMBER_ID multidetermined CLUB.

- To represent both in a single joint statement using the notion is:

  MEMBER_ID $\twoheadrightarrow$ AGE | CLUB

Fourth Normal Form

- A relation will be in Fourth Normal Form (4NF) if it is in BCNF and has no multivalued dependency.
- For a dependency A → B, if for a single value of A, multiple values of B exists, then the relation will be a multivalued dependency.

Fourth Normal Form (Cont.)

- In the previous example, the given MEMBER table is in 3NF but there is no relationship between AGE and CLUB.
- Hence, AGE and CLUB are two independent entity but both are dependent on MEMBER_ID.
- So, there is a multivalued dependency on MEMBER_ID which leads to unnecessary repetition of data.

Fourth Normal Form (Cont.)

- So, to make the MEMBER relation into 4NF, we can decompose it into two tables:

MEMBER_AGE

MEMBER_ID	AGE
1	18
2	20
3	16
4	18
5	19

MEMBER_CLUB

MEMBER_ID	CLUB
1	Dancing
2	Dancing
3	Singing
4	Dancing
5	Singing

Join Dependency

- Join Dependency (JD) is a further generalization of multivalued dependencies.
- If a table can be recreated by joining multiple tables and each of the tables have a subset of the attributes of the original table, then the table is in join dependency (JD).
- In other words, if R1{A,B,C} and R2{C,D} are decompositions of a given relation R{A,B,C,D} and the join of R1 and R2 over C is equal to relation R, then a JD exists in R.
- Alternatively, R1 and R2 are nonloss decomposition of R.

Join Dependency (Cont.)

Example

- The relation MEMBER can be decomposed into the following two tables and it can be recreated by joining those tables; therefore, the relation MEMBER has join dependency.

MEMBER

ID	AGE	CLUB	TRAINER
1	18	Dancing	Alex
2	20	Dancing	Alex
3	16	Singing	Susan

ID_AGE_CLUB

ID	AGE	CLUB
1	18	Dancing
2	20	Dancing
3	16	Singing

CLUB_TRAINER

CLUB	TRAINER
Dancing	Alex
Dancing	Alex
Singing	Susan

Fifth Normal Form

- A relation is in Fifth Normal Form (5NF) if it is in 4NF and it does not contain any join dependency.
- 5NF is also known as Project-Join Normal Form (PJ/NF).
- 5NF is satisfied when all the tables are broken into as many tables as possible in order to avoid redundancy.

Fifth Normal Form (Cont.)

Example

- Consider the relation R below having the schema.
- The primary key is a combination of all three attributes of the relation.

Table 1

Supplier	Product	Consumer
S1	P1	C1
S1	P2	C1
S2	P1	C1
S3	P3	C3

Fifth Normal Form (Cont.)

Example

- As the original Table 1 can be recreated by joining the following three tables, Table 1 has join dependency and it is not in 5NF.
- However, Table 2, Table 3 and Table 4 satisfy 5NF as it has no multivalued dependency and cannot be decomposed further which means join dependency does not exists.

Table 2

Supplier	Product
S1	P1
S1	P2
S2	P1
S3	P3

Table 3

Product	Consumer
P1	C1
P2	C1
P3	C3

Table 4

Supplier	Consumer
S1	C1
S2	C1
S3	C3

Normalization Procedure Summarized

- The overall process can be state informally as a set of rules as follows.
- Take projections of the original 1NF relation to eliminate any functional dependencies that are not irreducible. This step will produce a collection of 2NF relations.
- Take projections of those 2NF relations to eliminate any transitive functional dependencies. This step will produce a collection of 3NF relations.

Normalization Procedure Summarized (Cont.)

- Take projections of those 3NF relations to eliminate any remaining functional dependencies in which the determinant is not a candidate key. This step will produce a collection of BCNF relations.
- Take projections of those BCNF relations to eliminate any multivalued dependencies that are not also functional dependencies. This step will produce a collection of 4NF relations.

Normalization Procedure Summarized (Cont.)

- Take projections of those 4NF relations to eliminate any join dependencies that are not implied by the candidate keys. This step will produce a collection of relations in 5NF.
- The process of taking projections at each step must be done in a nonloss way and in a dependency-preserving way.

Normalization Procedure Summarized (Cont.)

- Notice that there is a very attractive parallelism among the definitions of BCNF, 4NF and 5NF, viz.:
  - A relation R is in BCNF if and only if every FD in R is implied by the candidate keys of R
  - A relation R is in 4NF if and only if every MVD in R is implied by the candidate keys of R
  - A relation R is in 5NF if and only if every JD in R is implied by the candidate keys of R

Normalization Procedure Summarized (Cont.)

- The overall objectives of the normalization process are as follows:
  - To eliminate certain kinds of redundancy
  - To avoid certain update anomalies
  - To produce a design that is a good representation of the real world – one that is intuitively easy to understand and a good base for future growth
  - To simplify the enforcement of certain integrity constraints

Other Normal Forms

Domain-Key Normal Form

- Domain-key normal form (DK/NF) was proposed by Fagin.
- DK/NF is not defined in terms of FDs, MVDs, or JDs at all.
- Instead, a relation R is said to be in DK/NF if and only if every constraint on R is a logical consequence of the domain constraints and key constraints that apply to R.

Other Normal Forms (Cont.)

Domain-Key Normal Form

- A domain constraint is a constraint to the effect that values of a given attribute are taken from some prescribed domain.
- A key constraint is a constraint to the effect that a certain attribute or attribute combination constitutes a candidate key.

Other Normal Forms (Cont.)

Restriction-Union Normal Form

- Another direction for normalization research consists of examining the implications of decomposing relations by some operation other than projection.
- Here, the decomposition operator is restriction and the corresponding recomposition operator is union.
- Thus, it might be possible to construct a restriction-union normalization theory, analogous but orthogonal to the projection-join normalization theory.

Summary

- A relation is in 4NF if the only MVDs it satisfies are in fact FDs out of candidate keys.
- A relation is in 5NF (also called projection-join normal form, PJ/NF) if and only if the only JDs it satisfies are in fact FDs out of candidate keys.
- 5NF is the ultimate normal form with respect to projection and join.

Next Lecture

Part III: Database Design

The Entity/Relationship Model

- Introduction
- An Overview of the E/R Model
- E/R Diagrams

Textbook and References

Textbook

- C. J. Date, “An Introduction to Database Systems”, 6th Edition, 1994.

Additional References

- Abraham Silberschatz, Henry F. Korth, S. Sudarshan, “Database System Concepts”, 6th Edition, 2011.
- Ramez Elmasri, Shamkant B. Navathe, “Fundamentals of Database Systems”, 6th Edition, 2010.
- https:// www.tutorialspoint.com
- https://www.javatpoint.com
- https://www.geeksforfeeks.org