DATABASE - 1

3RD CLASS

COMPUTER SCIENCE DEPARTMENT

7th Lecture – Boyce-Codd Normal Form, Fourth Normal Form

Sunday 3rd of November 2024

LECTURER :

DR. RAYAN YOUSIF ALKHAYAT

BOYCE-CODD NORMAL FORM

1. When a relation has more than one candidate key, anomalies may result even though that relation is in 3NF.
2. For example, consider the STUDENT ADVISOR relation shown in Figure B-1. This relation has the following attributes: SID (student ID), Major,Advisor, and MajGPA.
3. Sample data for this relation are shown in Figure B-1a, and the functional dependencies are shown in Figure B-1b.

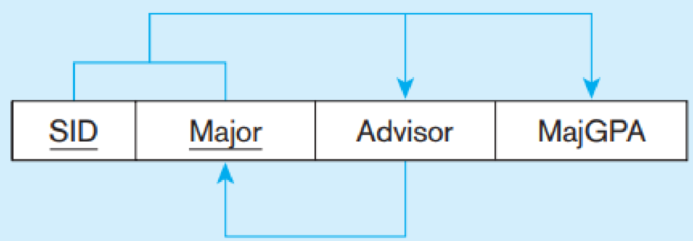
FIGURE B-1 Relation in 3NF but not in BCNF

(a) Relation with sample data

STUDENT ADVISOR

SID	Major	Advisor	MajGPA
123	Physics	Hawking	4.0
123	Music	Mahler	3.3
456	Literature	Michener	3.2
789	Music	Bach	3.7
678	Physics	Hawking	3.5

(b) Functional dependencies in STUDENT ADVISOR



- As shown in Figure B-1b, the primary key for this relation is the composite key consisting of SID and Major.
- Thus, the two attributes Advisor and MajGPA are functionally dependent on this key.

- This reflects the constraint that although a given student may have more than one major, for each major a student has exactly one advisor and one GPA.

- There is a second functional dependency in this relation: Major is functionally dependent on Advisor.
- That is, each advisor advises in exactly one major.
- Notice that this is not a transitive dependency, where a transitive dependency is a functional dependency between two nonekey attributes.
- In contrast, in this example a key attribute (Major) is functionally dependent on a nonekey attribute (Advisor).

Anomalies in STUDENT ADVISOR

The STUDENT ADVISOR relation is clearly in 3NF, because there are no partial functional dependencies and no transitive dependencies.

Nevertheless, because of the functional dependency between Major and Advisor, there are anomalies in this relation.

Consider the following examples:

1. Suppose that in Physics, the advisor Hawking is replaced by Einstein. This change must be made in two (or more) rows in the table (update anomaly).
2. Suppose we want to insert a row with the information that Babbage advises in Computer Science. This, of course, cannot be done until at least one student majoring in Computer Science is assigned Babbage as an advisor (insertion anomaly).
3. Finally, if student number 789 withdraws from school, we lose the information that Bach advises in Music (deletion anomaly).

DEFINITION OF BOYCE-CODD NORMAL FORM (BCNF)

- The anomalies in STUDENT ADVISOR result from the fact that there is a determinant (Advisor) that is not a candidate key in the relation. R. F. Boyce and E. F. Codd identified this deficiency and proposed a stronger definition of 3NF that remedies the problem.

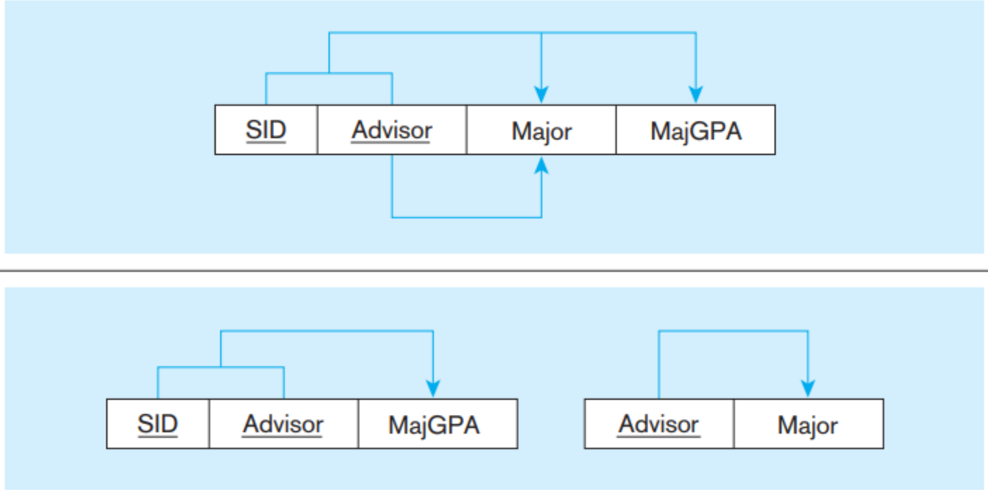
- We say a relation is in Boyce-Codd normal form (BCNF) if and only if every determinant in the relation is a candidate key. STUDENT ADVISOR is not in BCNF because although the attribute Advisor is a determinant, it is not a candidate key. (Only Major is functionally dependent on Advisor.)

- **Boyce-Codd normal form (BCNF):** A normal form of a relation in which every determinant is a candidate key

CONVERTING A RELATION TO BCNF

- A relation that is in 3NF (but not BCNF) can be converted to relations in BCNF using a simple two-step process. This process is shown in Figure B-2.
- **1. In the first step**, the relation is modified so that the **determinant** in the relation that is not a candidate key becomes a **component of the primary** key of the revised relation. The attribute that is functionally dependent on that determinant becomes a nonkey attribute.
- This is a legitimate restructuring of the original relation because of the functional dependency. The result of applying this rule to STUDENT ADVISOR is shown in Figure B-2a. **The determinant Advisor** becomes part of the **composite primary key**. The attribute Major, which is functionally dependent on Advisor, becomes a nonkey attribute.

Converting a Relation to BCNF



STUDENT

SID	Advisor	MajGPA
123	Hawking	4.0
123	Mahler	3.3
456	Michener	3.2
789	Bach	3.7
678	Hawking	3.5

ADVISOR

Advisor	Major
Hawking	Physics
Mahler	Music
Michener	Literature
Bach	Music

FIGURE B-2 Converting a relation to BCNF relations

(a) Revised STUDENT ADVISOR relations (1NF)

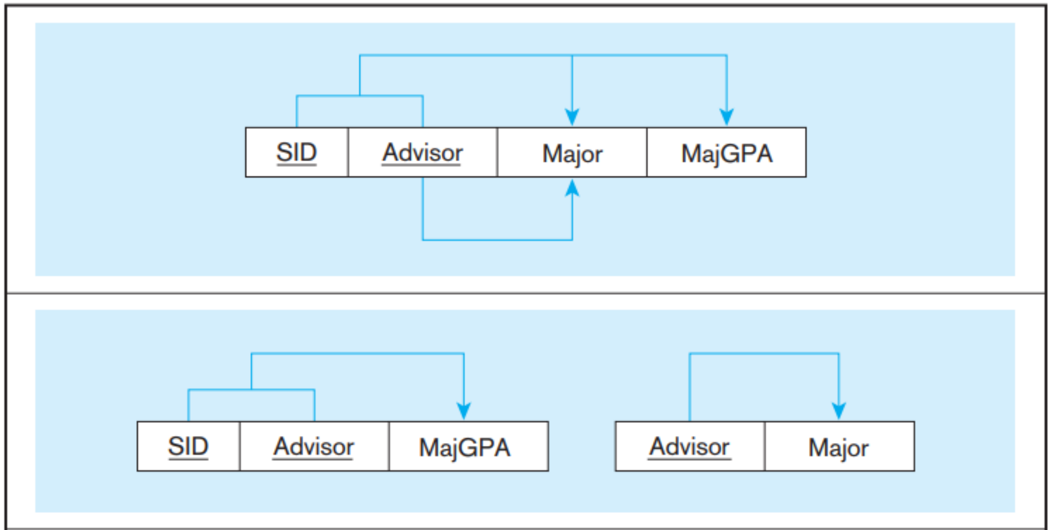
(b) Two relations in BCNF

(c) Relations with sample data

CONVERTING A RELATION TO BCNF

- If you examine Figure B-2a, you will discover that the new relation has a partial functional dependency. (Major is functionally dependent on Advisor, which is just one component of the primary key.) Thus the new relation is in first (but not second) normal form.
-
  2. **The second step** in the conversion process is to decompose the relation to eliminate the partial functional dependency, as we learned earlier .
- This results in two relations, as shown in Figure B-2b. These relations are in 3NF. In fact, the relations are also in BCNF because there is only one candidate key (the primary key) in each relation.

Converting a Relation to BCNF



STUDENT

SID	Advisor	MajGPA
123	Hawking	4.0
123	Mahler	3.3
456	Michener	3.2
789	Bach	3.7
678	Hawking	3.5

ADVISOR

Advisor	Major
Hawking	Physics
Mahler	Music
Michener	Literature
Bach	Music

FIGURE B-2 Converting a relation to BCNF relations

(a) Revised STUDENT ADVISOR relations (1NF)

(b) Two relations in BCNF

(c) Relations with sample data

CONVERTING A RELATION TO BCNF

- **Thus, we see that if a relation has only one candidate key (which therefore becomes the primary key), 3NF and BCNF are equivalent. The two relations (now named STUDENT and ADVISOR) with sample data are shown in Figure B-2c.**

FOURTH NORMAL FORM

The study of fourth normal form 4NF is focusing on what so called (Multivalued Dependency)

FOURTH NORMAL FORM

1. When a relation is in BCNF, there are no longer any anomalies that result from functional dependencies.
2. However, there may still be anomalies that result from multivalued dependencies.
3. For example, consider the user view shown in Figure B-4a.
4. This user view shows for each course the instructors who teach that course and the textbooks that are used. (These appear as repeating groups in the view.)

FIGURE B-4 **Data with multivalued dependencies**

(a) View of courses, instructors, and textbooks

COURSE STAFF AND BOOK ASSIGNMENTS

Course	Instructor	Textbook
Management	White Green Black	Drucker Peters
Finance	Gray	Jones Chang

(b) Relation in BCNF

OFFERING

Course	Instructor	Textbook
Management	White	Drucker
Management	White	Peters
Management	Green	Drucker
Management	Green	Peters
Management	Black	Drucker
Management	Black	Peters
Finance	Gray	Jones
Finance	Gray	Chang

FOURTH NORMAL FORM

In that table view, the following assumptions hold:

1. Each course has a well-defined set of instructors (e.g., Management has three instructors).
2. Each course has a well-defined set of textbooks that are used (e.g., Finance has two textbooks).
3. The textbooks that are used for a given course are independent of the instructor for that course (e.g., the same two textbooks are used for Management regardless of which of the three instructors is teaching Management).

In Figure B-4b, this table view has been converted to a relation by filling in all the empty cells.

This relation (named OFFERING) is in 1NF. Thus, for each course, all possible combinations of instructor and text appear in OFFERING.

Notice that the primary key of this relation consists of all three attributes (Course, Instructor, and Textbook). Because there are no determinants other than the primary key, the relation is actually in BCNF.

Yet it does contain much redundant data that can easily lead to update anomalies.

For example, suppose that we want to add a third textbook (author: Middleton) to the Management course.

This change would require the addition of three new rows to the relation in Figure B-4b, one for each Instructor (otherwise that text would apply to only certain instructors)

MULTIVALUED DEPENDENCIES

The type of dependency that called a multivalued dependency, it exists when there are at least three attributes (e.g., A, B, and C) in a relation, and for each value of A there is a well-defined set of values of B and a well-defined set of values of C.

However, the set of values of B is independent of set C,and vice versa.

To remove the multivalued dependency from a relation, we divide the relation into two new relations. Each of these tables contains two attributes that have a multivalued relationship in the original relation.

REMOVING MULTIVALUED DEPENDECIY

Figure B-5 shows the result of this decomposition for the OFFERING relation of Figure B-4b.

Notice that the relation called TEACHER contains the Course and Instructor attributes, because for each course there is a well-defined set of instructors.

Also, for the same reason, TEXT contains the attributes Course and Textbook.

However, there is no relation containing the attributes Instructor and Course because these attributes are independent.

FIGURE B-4 **Data with multivalued dependencies**

(a) View of courses, instructors, and textbooks

COURSE STAFF AND BOOK ASSIGNMENTS

Course	Instructor	Textbook
Management	White Green Black	Drucker Peters
Finance	Gray	Jones Chang

(b) Relation in BCNF

OFFERING

Course	Instructor	Textbook
Management	White	Drucker
Management	White	Peters
Management	Green	Drucker
Management	Green	Peters
Management	Black	Drucker
Management	Black	Peters
Finance	Gray	Jones
Finance	Gray	Chang

FIGURE B-5 **Relations in 4NF**

TEACHER

Course	Instructor
Management	White
Management	Green
Management	Black
Finance	Gray

TEXT

Course	Textbook
Management	Drucker
Management	Peters
Finance	Jones
Finance	Chang

FOURTH NORMAL FORM

We can define 4NF as :

- A normal form of a relation in which the relation is in BCNF and contains no multivalued dependencies.

END OF LECTURE 7

DATABASE - 1

3RD CLASS

COMPUTER SCIENCE DEPARTMENT

8th Lecture – Merging Relations

10 of November 2024

LECTURER :

DR. RAYAN YOUSIF ALKHAYAT