DB Normalization 7.3& 7.7

Yupei Sun

Some Important Terms

- **Superkey:** A superkey is a set of one or more attributes (columns) that can uniquely identify records in a relational database table.
- **Candidate Key:** A candidate key is a minimal set of attributes (columns) within a table that can uniquely identify any table record without referring to any other data.
- **Primary Key:** A candidate key that is selected when there are multiple of them.
- **Primary Attribute:** An attribute is considered a primary attribute if it appears in any candidate key.
- **Non-Primary Attribute:** An attribute is considered a non-primary attribute if it does not appear in any candidate key.
- **Partial Functional Dependency:** Given a relation R(A,B,C), if AB→C and A→C, then C is partially functionally dependent on AB.
- **Full Functional Dependency:** Given a relation R(A,B,C), if AB→C and neither A→C nor B→C, then C is fully functionally dependent on AB.
- **Transitive Functional Dependency:** Given a relation R(A,B,C), if A→B and B→C, then C is transitively functionally dependent on A.

Normal Forms

- Normal forms are used to classify database schemas based on their levels of redundancy and the presence of undesirable characteristics like update anomalies, insertion anomalies, and deletion anomalies.
- E.g. first-fifth normal forms, BCNF
- Many examples of simple databases will be given for easier understanding but the data in those databases are NOT necessarily True.

Decomposing Schemas in Normal Forms

- To decompose a schema not in BCNF. Let R be a schema that is not in BCNF. Then there is at least one nontrivial functional dependency $\alpha \rightarrow \beta$ such that α is not a superkey for R. We replace R in our design with two schemas:
- - $(\alpha \cup \beta)$
  - $(R - (\beta - \alpha))$
- Works for decomposing other schemas too
- I will show you some easier way to do that.

Level of the normal forms (1NF)

- Every attribute in the relation is indivisible.
- Minimum Requirement to be a relational database.

Band_Name	Annual Income
Whispered	$100,000, $50,000
Estatic Fear	$50,000, $30,000
Epica	$200,000, $200,000
Dream Theatre	$1,000,000, $2,000,000

Band_Name	Basic_Salary	Commission
Whispered	$100,000	$50,000
Estatic Fear	$50,000	$30,000
Epica	$200,000	$200,000
Dream Theatre	$1,000,000	$2,000,000

Level of the normal forms (2NF)

- Based on 1NF, it eliminates partial functional dependency of non-primary attributes on candidate keys.
- **Revision of Partial Functional Dependency:** Given a relation R(A,B,C), if AB→C and A→C, then C is partially functionally depend

Band_Name	Concert_Location	Ticket_Price	Drummer_Name	Music_Style
Whispered	Haverford	$100	Jaakko Nylund	Melodic Death Metal
Estatic Fear	Bryn Mawr	$110	Markus Pointner	Doom Metal
Epica	Haverford	$120	Jeroen Simons	Symphonic Metal
Dream Theatre	Bryn Mawr	$130	Mike Portnoy	Progressive Metal
Within Temptation	Haverford	$140	Mike Coolen	Gothic Metal
Whispered	Bryn Mawr	$150	Jaakko Nylund	Melodic Death Metal
Estatic Fear	Haverford	$160	Markus Pointner	Doom Metal
Epica	Bryn Mawr	$170	Jeroen Simons	Symphonic Metal
Dream Theatre	Haverford	$180	Mike Portnoy	Progressive Metal
Within Temptation	Bryn Mawr	$190	Mike Coolen	Gothic Metal

Correct Example of 2NF

Band_Name	Concert_Location	Ticket_Price
Whispered	Haverford	$100
Estatic Fear	Bryn Mawr	$110
Epica	Haverford	$120
Dream Theatre	Bryn Mawr	$130
Within Temptation	Haverford	$140
Whispered	Bryn Mawr	$150
Estatic Fear	Haverford	$160
Epica	Bryn Mawr	$170
Dream Theatre	Haverford	$180
Within Temptation	Bryn Mawr	$190

Band_Name, Concert_Location→Ticket_Price

Band_Name	Drummer_Name	Music_Style
Whispered	Jaakko Nylund	Melodic Death Metal
Estatic Fear	Markus Pointner	Doom Metal
Epica	Jeroen Simons	Symphonic Metal
Dream Theatre	Mike Portnoy	Progressive Metal
Within Temptation	Mike Coolen	Gothic Metal

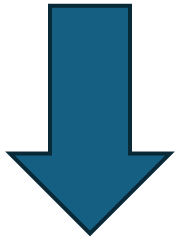
Band_Name->Drummer, Music_Style

Level of the normal forms (3NF)

- Definition from the book:
- A relation schema R is in **third normal form** with respect to a set F of functional dependencies if, for all functional dependencies in F^+ of the form $\alpha \rightarrow \beta$, where $\alpha \subseteq R$ and $\beta \subseteq R$, at least one of the following holds:
- - $\alpha \rightarrow \beta$ is a trivial functional dependency.
  - α is a superkey for R.
  - Each attribute A in $\beta - \alpha$ is contained in a candidate key for R.
- My definition:
- Based on 2NF, it eliminates transitive functional dependency of non-primary attributes on candidate keys.

Band_Name →Drummer_Name
Drummer →Drummer_Hometown
Band_Name →Drummer_Hometown

Band_Name	Drummer_Name	Drummer_Hometown
Whispered	Jaakko Nylund	Finland
Estatic Fear	Markus Pointner	Austria
Epica	Jeroen Simons	Netherland
Dream Theatre	Mike Portnoy	United States
Within Temptation	Mike Coolen	Netherland



Band_Name	Drummer_Name
Whispered	Jaakko Nylund
Estatic Fear	Markus Pointner
Epica	Jeroen Simons
Dream Theatre	Mike Portnoy
Within Temptation	Mike Coolen

Band_Name →Drummer_Name

Band_Name	Drummer_Hometown
Whispered	Finland
Estatic Fear	Austria
Epica	Netherland
Dream Theatre	United States
Within Temptation	Netherland

Drummer_Name →Drummer_Hometown

Boyce-Codd Normal Form(BCNF)

- The Normal Form majorly talked by the book
- Introduced by R.F. Boyce and E.F. Codd in the 1970s
- Eliminates all redundancy that can be discovered based on functional dependencies

Definition By the Book

- A relation schema R is in BCNF with respect to a set F of functional dependencies if, for all functional dependencies in F+ of the form $\alpha \rightarrow \beta$, where $\alpha \subseteq R$ and $\beta \subseteq R$, at least one of the following holds:
- $\alpha \rightarrow \beta$ is a trivial functional dependency (i.e., $\beta \subseteq \alpha$).
- α is a superkey for schema R.
- Revision of definition of superkey:A combination of attributes that ensures each record in a database table can be uniquely identified, even if the combination includes more attributes than necessary to establish that uniqueness.

My Definition of BCNF

- Based on 3NF, it eliminates partial and transitive functional dependencies of **primary attributes** on candidate keys.
- *3NF only eliminates partial and transitive functional dependencies of **non-primary attributes** on candidate keys.

Example From Book

- in_dep (ID, name, salary, dept_name, building, budget)

ID	name	salary	dept_name	building	budget
1	Alex, Adam	$80,000	Computer Science	Park Science	$500,000
2	Bob, Byran	$75,000	Spanish	Taylor	$600,000

A simple illustration of the dataset I built based on the description from book.

Answering (in_dep (ID, name, salary, dept_name, building, budget))

- Functional dependency Dept_name → budget holds
- Dept_name is not a superkey
- Candidate key(ID, Dept_name)
- Not even satisfying 2NF

ID	name	salary	dept_name
1	Alex, Adam	$80,000	Computer Science
2	Bob, Byran	$75,000	Spanish

dept_name	building	budget
Computer Science	Park Science	$500,000
Spanish	Taylor	$600,000

One More Example Question about BCNF

- Since that example is actually not even satisfying 2NF!
- I will give an example satisfying 2NF and 3NF but not BCNF to really see what is the difference and uniqueness of it.

One More Example Question about BCNF

Band_Name	Drummer_Name	Album_Name	Album_rate
Whispered	Jaakko Nylund	Thousand Swords	10
Estatic Fear	Markus Pointner	Somnium Obmutum	7
Epica	Jeroen Simons	The Phantom Agony	6
Dream Theatre	Mike Portnoy	A View from the Top of the World	9
Within Temptation	Mike Coolen	The Heart of Everything	5
Whispered	Jaakko Nylund	Shogunate Macabre	9
Estatic Fear	Markus Pointner	A Sombre Dance	9
Epica	Jeroen Simons	Consign to Oblivion	6
Dream Theatre	Mike Portnoy	Metropolis Pt. 2: Scenes from a Memory	10
Within Temptation	Mike Coolen	Bleed Out	4
GraveLand	M. Ahrin	Thousand Swords	3

Simply based on my personal preference, don't have any other meaning

What are the candidate keys?
Satisfying 2NF and 3NF?
Album_Name →Album_rate?

(Band_Name, Album_Name) or (Drummer_Name, Album_Name) , both could be candidate key

Band_Name	Album_Name	Album_rate
Whispered	Thousand Swords	10
Estatic Fear	Somnium Obmutum	7
Epica	The Phantom Agony	6
Dream Theatre	A View from the Top of the World	9
Within Temptation	The Heart of Everything	5
Whispered	Shogunate Macabre	9
Estatic Fear	A Sombre Dance	9
Epica	Consign to Oblivion	6
Dream Theatre	Metropolis Pt. 2: Scenes from a Memory	10
Within Temptation	Bleed Out	4
GraveLand	Thousand Swords	3

Band_Name	Drummer_Name
Whispered	Jaakko Nylund
Estatic Fear	Markus Pointner
Epica	Jeroen Simons
Dream Theatre	Mike Portnoy
Within Temptation	Mike Coolen
Whispered	Jaakko Nylund
GraveLand	M. Ahrin

No partial and transitive functional dependencies of **primary attributes** on candidate keys

Comparing BCNF and 3NF

- Goals of database design with functional dependencies are:
- 1. BCNF.
  2. Losslessness.
  3. Dependency preservation.
- Trade off between BCNF and dependency preservation.

Higher Normal Forms

- Sometimes even databases satisfying BCNF is not good enough.
- Example from book
- (*ID*, *child name*)
- (*ID*, *phone number*)
- Combining them we get:
- (*ID*, *child name*, *phone number*)
- Bad Idea(Repetition)
- (99999, David, 512-555-1234)
- (99999, David, 512-555-4321)
- (99999, William, 512-555-1234)
- (99999, William, 512-555-4321)

Basic Information About Higher Normal forms

- 4NF
- 5NF(Project-Join Normal Form)
- Domain Key Normal Form
- Mentioned but not discussed in 7.7

References

- https://blog.csdn.net/ffzbknd/article/details/78484974
- https://blog.csdn.net/Kzj_19/article/details/116308277?utm_medium=distribute.pc_relevant.none-task-blog-2~default~baidujs_baidulandingword~default-0-116308277-blog-106980598.235^v43^pc_blog_bottom_relevance_base9&spm=1001.2101.3001.4242.1&utm_relevant_index=3
- https://blog.csdn.net/qq_42225047/article/details/106980598
- https://chat.openai.com/c/7e48ab0e-4020-41e5-bd54-9fdef22defa0
- https://translate.google.com