

Beyond Third Normal Form

(Chapter 7)

Advanced Normalization

- Three additional forms:
 - Boyce Codd normal form (BCNF)
 - Fourth normal form (4NF)
 - Fifth normal form (5NF)
- Others have been proposed, but none gained much support
- In 2003, Date et al proposed a Sixth normal form related to time-dependent data

Common Misconceptions

- Advanced forms are too difficult to understand (more often just badly explained)
- If you avoid resolving distinct M:M relationships with a single entity (a good rule), further normalization is not needed
- 3NF is enough for business applications

BCNF

- Every determinant must be a candidate key.
- 3NF states that every determinant of a *nonkey* attribute must be a candidate key.
- Relations can be in 3NF but not BCNF when there are *overlapping* candidate keys

BCNF

- BRANCH_CUSTOMER: CUST_NO, BRANCH_NO, ..., SALES_ID

Can be changed to:

- BRANCH_CUSTOMER: CUST_NO, SALES_ID, ..., BRANCH_NO

But now we see that SALES_ID determines BRANCH_NO, so this should be split into two relations.

BCNF

- We often must trade enforcement of a particular rule for the advantages of normalization
- In the last example, the DBMS cannot enforce a rule that only one sales rep per branch can serve a customer
- Changes to data structures usually change the rules that can be enforced

Domain Key Normal Form (DKNF)

- Requires that all constraints are a consequence of domains or keys
- The notion that constraints are a consequence of keys is not obvious
- Not used much by practitioners
- Academic interest has also faded

4NF and 5NF

- Once in BCNF, remaining normalization problems deal almost exclusively with relations where every attribute is part of the primary key.
- 4NF is tricky to recognize, and since 5NF covers 4NF, we jump to it.
- Also, resolving two M:M relationships with two separate tables avoids 4NF problems

Business 5NF

- Keep splitting relations, stopping only when one of the following is true:
 - Any further splitting would lead to relations where the original view cannot be reconstructed with joins
 - The only splits left are trivial
- Trivial splits are those where resulting relations have the primary key or a candidate key of the other relation

Source: "Data Modeling Essentials", Graeme Simsion and Graham Whitt

Business 5NF

- Some three-way relationships are legitimate. Problems arise only where they can be split into simpler, more fundamental relationships.
- 5NF usually synonymous with “fully normalized”

Normalization and Redundancy

- Relations can be normalized and still permit redundancy:
 - Overlapping relations (e.g. Teacher and Student)
 - Derivable data – specifically data derived from attributes in other relations

Reference Tables and Normalization

- Normalization does not generate all the reference tables we may require
- Also normalizing out the last non-key attribute could lead to 4NF or 5NF issues that we would recognize if the non-key attribute were left alone

Normalization and Primary Keys

- Most primary key selection errors are made when moving from unnormalized structures to 1NF
- There are cases where the primary key of original view is not necessary as part of the primary key of a relation split out to satisfy 1NF (Agent and Insurance Policy example where Policy Number is a stand-alone key)

Modeling Exercise

Employee ID	Language Skill	Office Skill
10001	Spanish	Microsoft Office
10001	French	10-key
10002	Mandarin	Microsoft Office
10002	(null)	Shorthand
10003	Spanish	Typing 65+ WPM
10003	Italian	(null)