Database Management Essentials

Normalization Concepts and Practice

Summarized by
github.com/dgkim5360
dgkim5360.tistory.com

Modification Anomalies

- Unexpected side effects from a row operation
- Insert / update / delete more data than desired
- Caused by excessive redundancies
- 이를 해결하기 위해 strive for "one fact in one place"

Modification Anomalies

StdNo	StdClass	OfferNo	OffYear	EnrGrade	CourseNo	CrsDesc
S1	JUN	O1	2017	3.5	C1	DB
S1	JUN	O2	2017	3.3	C2	VB
S2	JUN	O3	2018	3.1	C3	OO
S2	JUN	O2	2017	3.4	C2	VB

- Big university table: PK (StdNo ,OfferNo)

- Insertion anomaly
  Course C4를 넣기 위해서는 student와 offer를 알아야 한다 (PK니까).

- Update anomaly
  C2 course의 description을 변경하려면 여러 row를 업데이트 해야 함.

- Deletion anomaly
  한 줄에 student, offering, course 정보가 모두 있으므로, 한 줄을 지우면 모두 날라감.

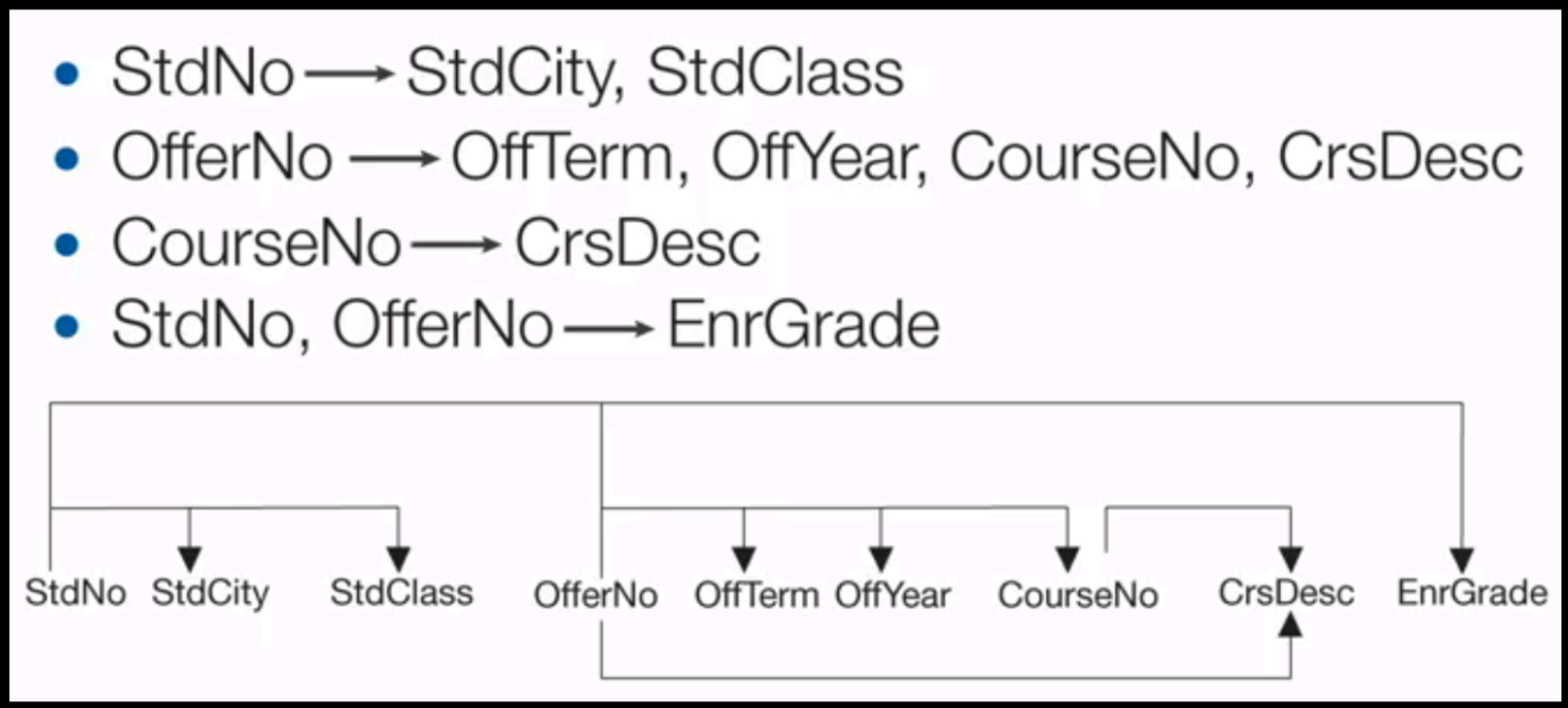
Functional Dependencies

Definition

- X determines Y
- X → Y
- For each X value, there is at most one Y value
- e.g., StdNo → StdCity

Unique Constraint Analogy

FD X → Y일 때, X는 unique한 값을 가져야 한다는 것과 비슷하다.



Functional Dependencies

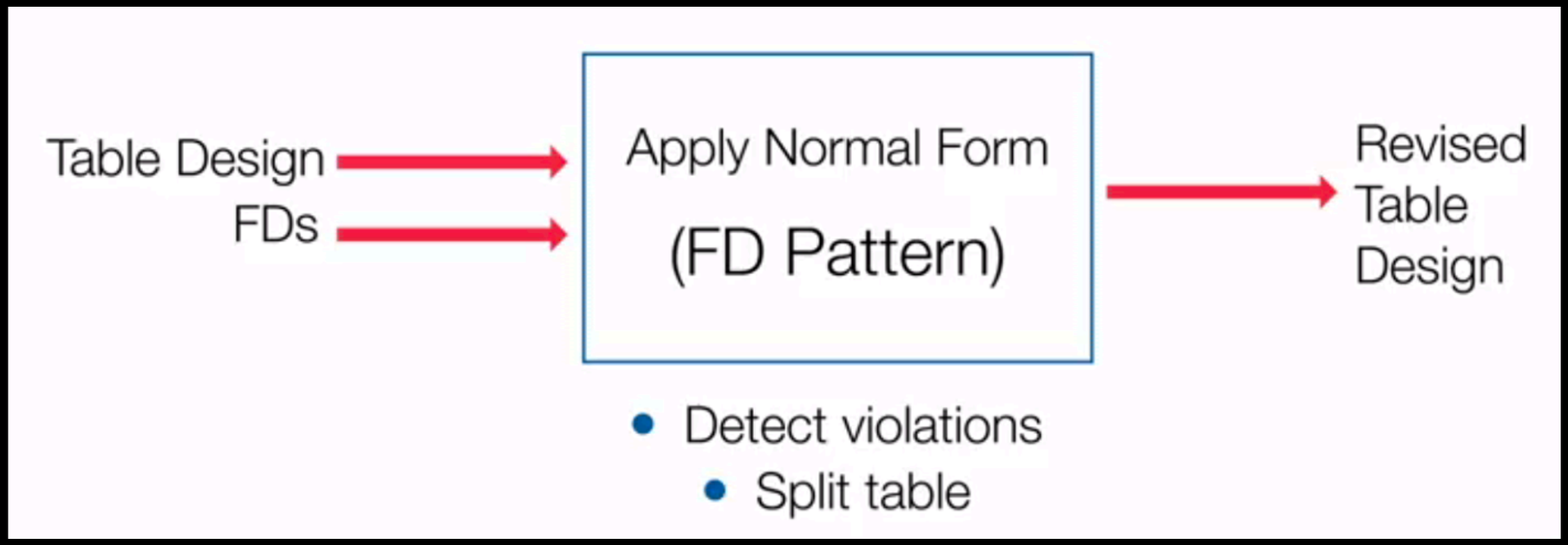
Falsification

- FD를 만족함을 증명할 수는 없고, 그 대신 반례를 찾아서 FD를 충족하지 않음을 증명
- Find two rows that have the same X values but different Y values!

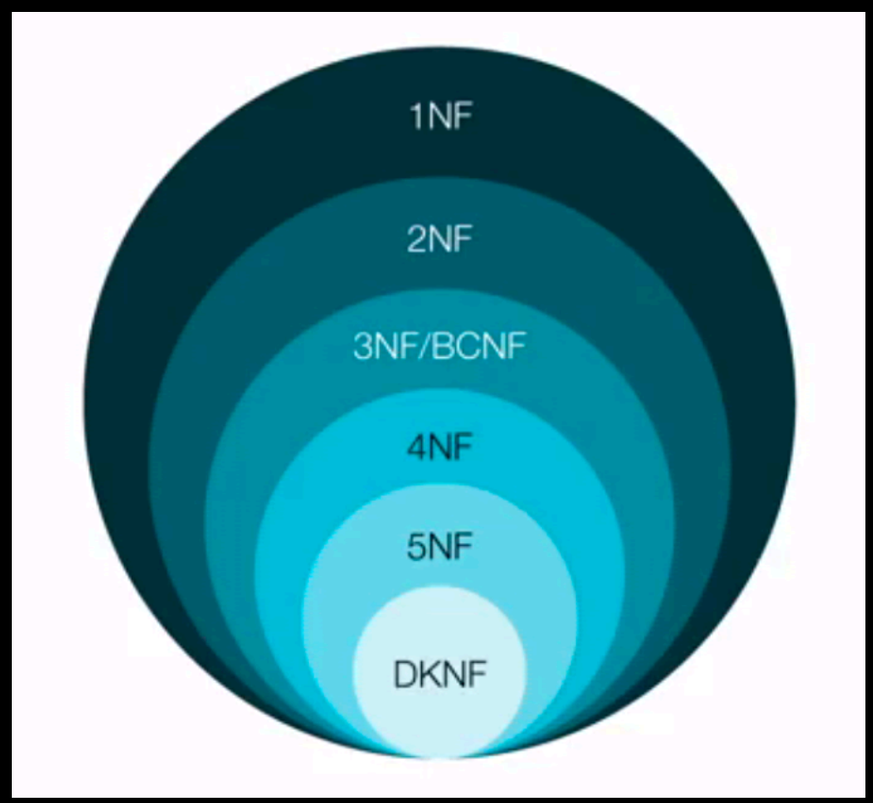
StdNo	StdClass	OfferNo	OffYear	EnrGrade	CourseNo	CrsDesc
S1	JUN	O1	2017	3.5	C1	DB
S1	JUN	O2	2017	3.3	C2	VB
S2	JUN	O3	2018	3.1	C3	OO
S2	JUN	O2	2017	3.4	C2	VB

- StdNo -X-> OfferNo by (#1, #2), (#3, #4)
- StdNo -X-> EnrGrade by (#3, #4)
- StdNo -?-> StdClass

Normal Forms



- First normal form: Starting point
- Second: Stronger than the first
  i.e., Only a subset of first normal form table are in second
- Each successive normal form refines the previous one to remove additional kinds of redundancies.
- BCNF is revised and stronger definition for third one.
  가장 중요한 constraint를 담당한다.



Boyce-Codd Normal Form

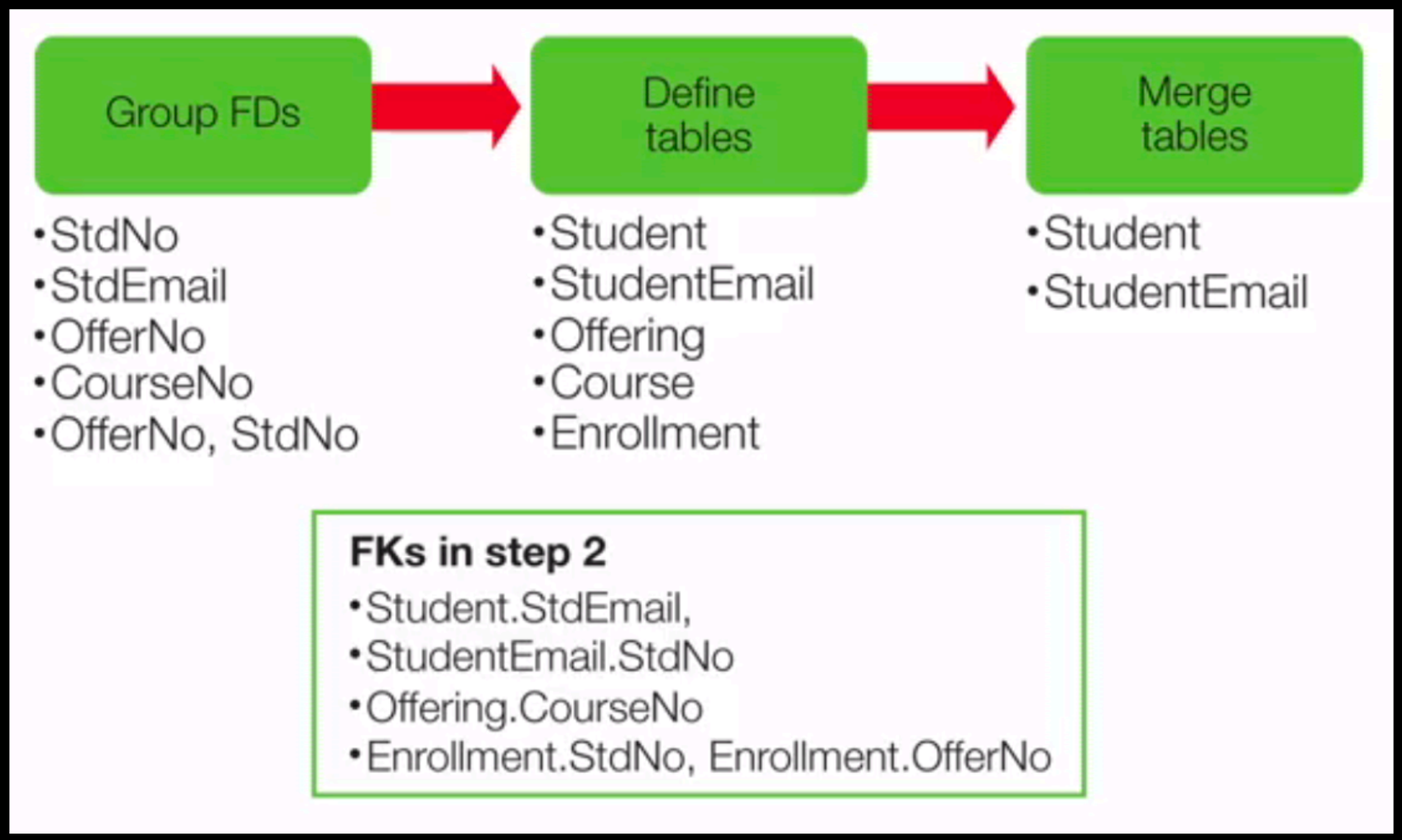
- Originally conceived as a simpler definition of the third normal form.
- BCNF requires that every determinant must be unique in a table. A FD violates BCNF if the determinate in the FD is not unique in a table.
- After demonstrating violations of BCNF, a simple procedure will be presented to refine a table design so that it conforms to BCNF.

Exercise

StdNo	StdEmail	StdClass	OfferNo	OffYear	EnrGrade	CourseNo	CrsDesc
S1	joe@bigu.edu	JUN	O1	2017	3.5	C1	DB
S1	sue@bigu.edu	JUN	O2	2017	3.3	C2	VB
S2	mj@bigu.edu	JUN	O3	2018	3.1	C3	OO
S2	tom@bigu.edu	JUN	O2	2017	3.4	C2	VB

- PKs: StdNo + OfferNo
- Unique columns: (StdNo, OfferNo), (StdEmail, OfferNo)
- BCNF violations: LHS of FD is not a determinant.
    - StdNo → StdCity, StdClass, StdEmail
      StdNo is a subset of a determinant (StdNo, OfferNo), but not a determinant itself.
    - StdEmail → StdNo
    - OfferNo → OffTerm, OffYear, CourseNo
    - CourseNo → CrsDesc

BCDF Procedure

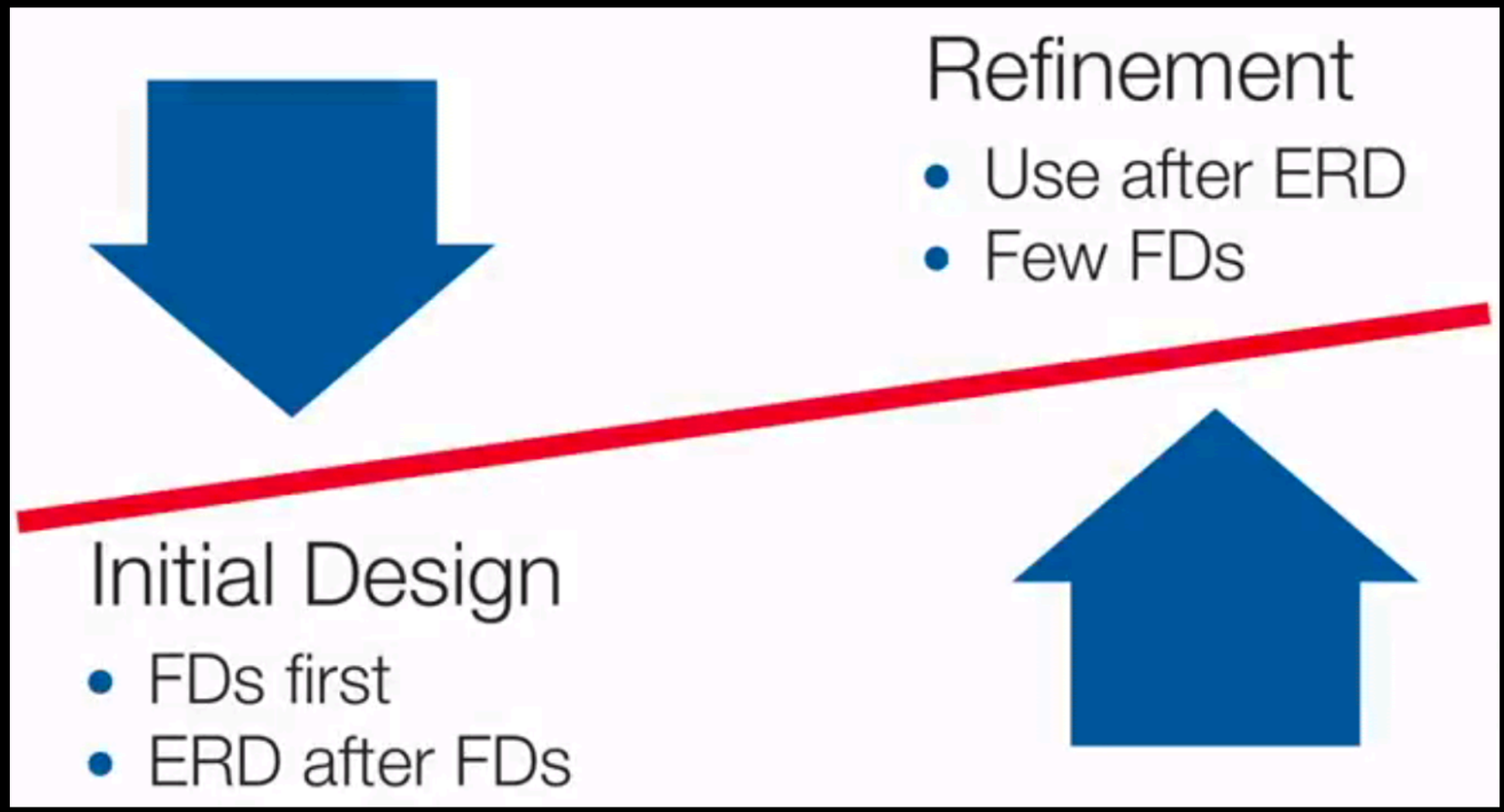


1. Organize FDs by determinant: Group FDs by the same determinant
2. Define tables
   - One for each FD group
   - Determinant = PK
   - Also add FKs
3. Merge tables to prevent excessive splitting
   - If one table contains a subset of columns in the other table

Practical Concerns

Normalization as Refinement Tool vs. Initial Design Tool

- 교수님 가라사대 refinement tool로써 사용하는 것이 바람직
- 왜냐하면 ERD를 그리는 것만으로, FD를 기록하는 지겨운 일 없이
  많은 normalization이 달성되기 때문



Denormalization

- Purposeful violation of a normal form
- Some FDs may not cause anomalies in practice
  (명확한 business rule에 의해)
- May improve performance (많은 join을 하지 않아도 되니까)
- Common for DW (무거운 query가 많아서)