Business School
UNIVERSITY OF COLORADO DENVER

Information Systems Program

Module 11
Normalization Concepts and Practice

Lesson 1: Modification Anomalies

Lesson Objectives

- Define modification anomaly
- Provide examples of modification anomalies
- Understand processing orientation for avoiding modification anomalies

Modification Anomaly

- Unexpected side effect from a row operation
- Must insert, modify, and delete more data than desired
- Caused by excessive redundancies
- Strive for one fact in one place

Big University Database Table

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

Anomaly Examples

- To insert a course (C4), must know student and offering
- Update multiple rows to change the description of course C2
- A row deletion can cause inadvertent removal of related entities. Deleting third enrollment row (S2, O3) loses details about O3 and C3.

Summary

- Modification anomaly: unwanted side effect from a row operation
- More rows impacted than anticipated
- Motivation for normalization process to remove excessive redundancies

Business School
UNIVERSITY OF COLORADO DENVER

Information Systems Program

Module 11
Normalization Concepts and Practice

Lesson 2: Functional dependencies

Lesson Objectives

- Define functional dependency
- Explain analogy of functional dependency to unique constraint
- Falsify functional dependencies in sample rows

Functional Dependency Basics

- Constraint on the possible rows in a table
- Value neutral like FKs and PKs
- Asserted
- Understand business rules

FD Definition

- Notation: X $\rightarrow$ Y
- X (functionally) determines Y
- For each X value, there is at most one Y value
- StdNo $\rightarrow$ StdCity if each StdNo value has at most one StdCity value
- X: left-hand side (LHS) or determinant
- Y: right-hand side (RHS)

Unique Constraint Analogy

- Similar to uniqueness constraint
- Place RHS and LHS in a table by themselves
- Examples
  - OfferNo → OffYear
  - OfferNo, StdNo → EnrGrade

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

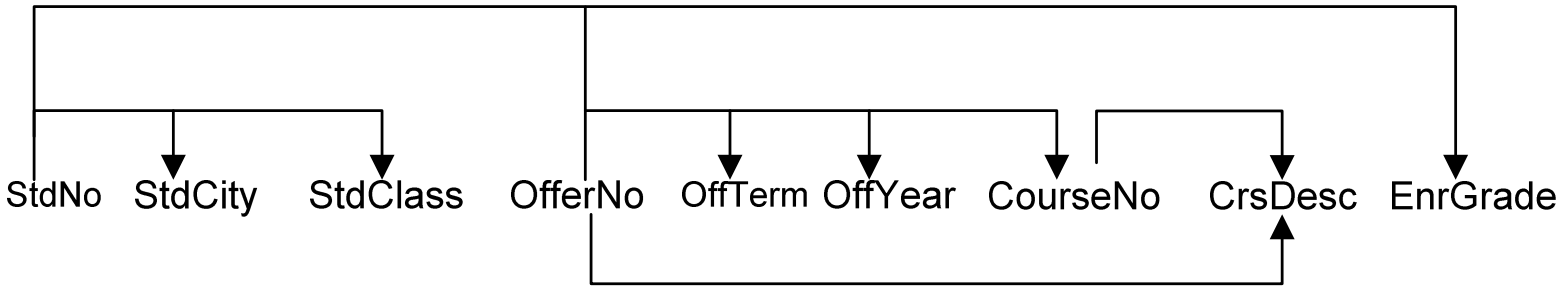
FD Lists and Diagrams

StdNo → StdCity, StdClass

OfferNo → OffTerm, OffYear, CourseNo, CrsDesc

CourseNo → CrsDesc

StdNo, OfferNo → EnrGrade



Falsification of FDs using Sample Rows

- Prove non existence (but not existence) by looking at data
- Two rows that have the same X value but a different Y value

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

Summary

- FDs are important constraints
- Asserting FDs is essential for removing unwanted redundancy
- Refinement activity

Business School
UNIVERSITY OF COLORADO DENVER

Information Systems Program

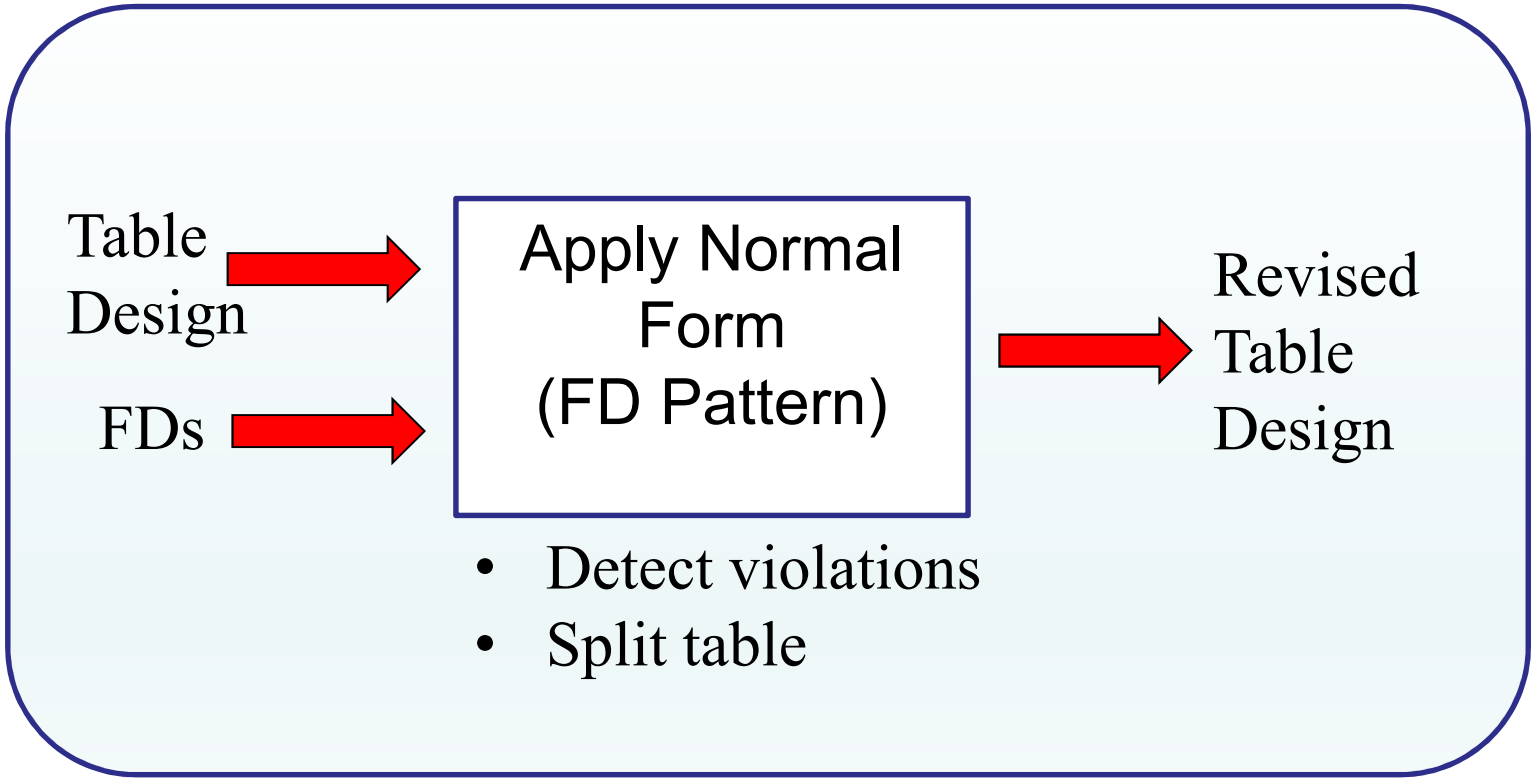
Module 11
Normalization Concepts and Practice

Lesson 3: Normal Forms

Lesson Objectives

- Understand the nature of normal forms
- Define Boyce-Codd Normal Form (BCNF)
- Apply BCNF to a list of functional dependencies

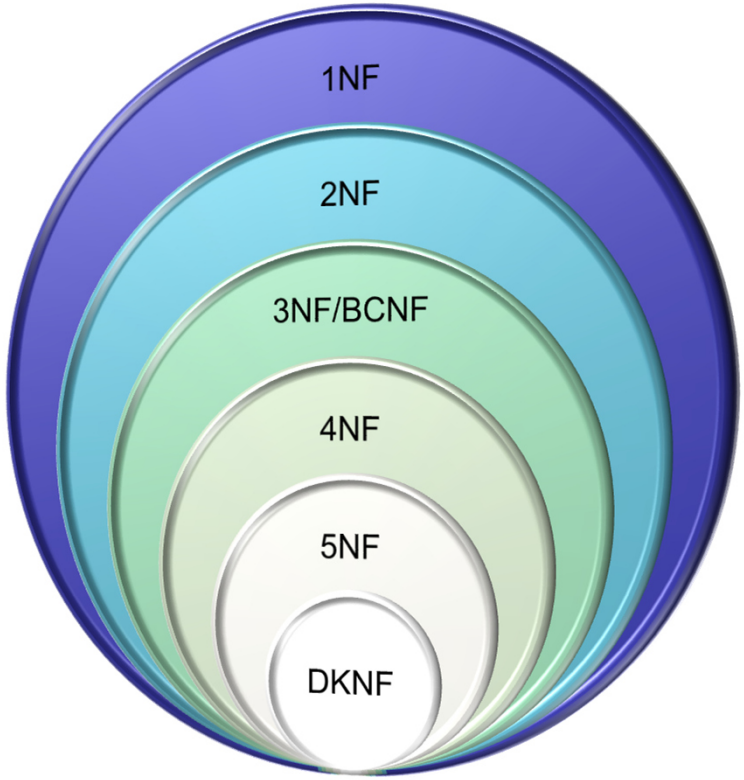
Normalization



Normalization Simplification

- Determination of a complete and minimal list of FDs
- Determination of unique columns from FDs
- Details too complex and specialized for this course

Relationships of Normal Forms



Boyce-Codd Normal Form (BCNF)

Simple definition

Every determinant must be unique.

Apply with BCNF procedure

Big University Database Table

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

BCNF Example

- Unique columns in the big university table
  - <StdNo, OfferNo>
  - <StdEmail, OfferNo>
- Many BCNF violations
  - ✓ StdNo → StdCity, StdClass, StdEmail
  - ✓ StdEmail → StdNo
  - ✓ OfferNo → OffTerm, OffYear, CourseNo
  - ✓ CourseNo → CrsDesc
  - – StdNo, OfferNo → EnrGrade

BCNF Procedure

Group FDs

- Same determinant

Define tables

- Each FD group
- Determinant as PK
- FKs

Merge tables

- Containment
- Prevents excessive splitting

BCNF Procedure Example

Group FDs

- StdNo
- StdEmail
- OfferNo
- CourseNo
- OfferNo, StdNo

Define tables

- Student
- StudentEmail
- Offering
- Course
- Enrollment

Merge tables

- Student
- StudentEmail

FKs in step 2

- Student.StdEmail,
- StudentEmail.StdNo
- Offering.CourseNo
- Enrollment.StdNo, Enrollment.OfferNo

Merging Tables

- Step 2 defines too many tables when two columns determine each other.
  - StdNo → StdEmail
  - StdEmail → StdNo
- Merge tables with a containment relationship
  - Student(StdNo, StdEmail, StdCity, StdClass)
  - StudentEmail(StdEmail, StdNo)
  - Merge tables because Student contains columns of StdEmail
- Multiple unique columns do not violate BCNF

Summary

- Covered general idea of normal forms and details of BCNF
- Know BCNF definition and simplified procedure
- Study examples for work on practice and graded problems

Business School
UNIVERSITY OF COLORADO DENVER

Information Systems Program

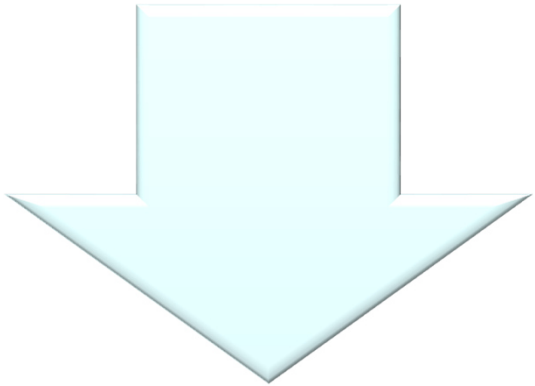
Module 11
Normalization Concepts and Practice

Lesson 4: Practical Concerns

Lesson Objectives

- Reflect on role of normalization
- Reflect on importance of normalization
- Reflect on situations to relax normalization requirements

Competing Roles of Normalization

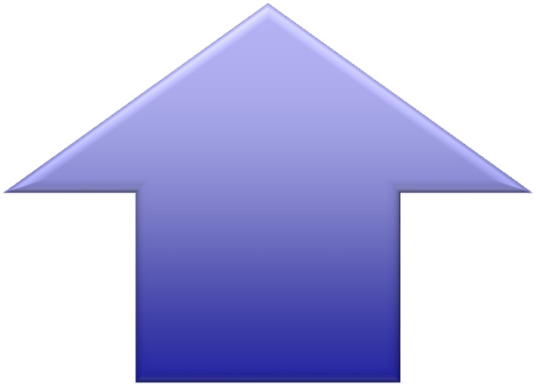


Refinement

- Use after ERD
- Few FDs

Initial Design

- FDs first
- ERD after FDs



Normalization Importance

- Update biased
- Not a major concern for databases without updates (data warehouses)
- Relax normalization sometimes

Denormalization

- Purposeful violation of a normal form
- Some FDs may not cause anomalies in practice
- May improve performance
- Common for data warehouses

Denormalization Example

- ZipCode → City, State
- Important for ecommerce business for sales tax
- May be important for ecommerce databases

Summary

- Covered practical issues
- Use normalization as a refinement approach
- Do not lose context of normalization when performing details

Business School
UNIVERSITY OF COLORADO DENVER

Information Systems Program

Module 11
Normalization Concepts and Practice

Lesson 5: Normalization problems

Lesson Objectives

- Gain confidence on practice problems
- Identify modification anomalies
- Identify sample rows that falsify FDs
- Apply both conversion rules and normalization

Modification Anomaly Problem

Big University Table

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

Problem requirements

- Specify one insert, update, and deletion anomaly
- Each anomaly should involve student representation in the table.

Modification Anomaly Problem Solution

Big University Table

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

Problem solution

- Insertion anomaly: cannot insert a student (S3) unless an OfferNo is provided.
- Update anomaly: must change multiple rows if S1 moves to a different city.
- Deletion anomaly: deleting third row also removes details about offering O3 and course C3.

FD Falsification Problem

Big University Table

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

Problem requirements

- List possible FDs with StdCity as determinant (LHS)
- Identify at least one falsification if it exists for each FD
  - Pair of sample rows for an FD falsification
  - Same LHS (determinant) value in each row but a different RHS value

FD Falsification Problem Solution

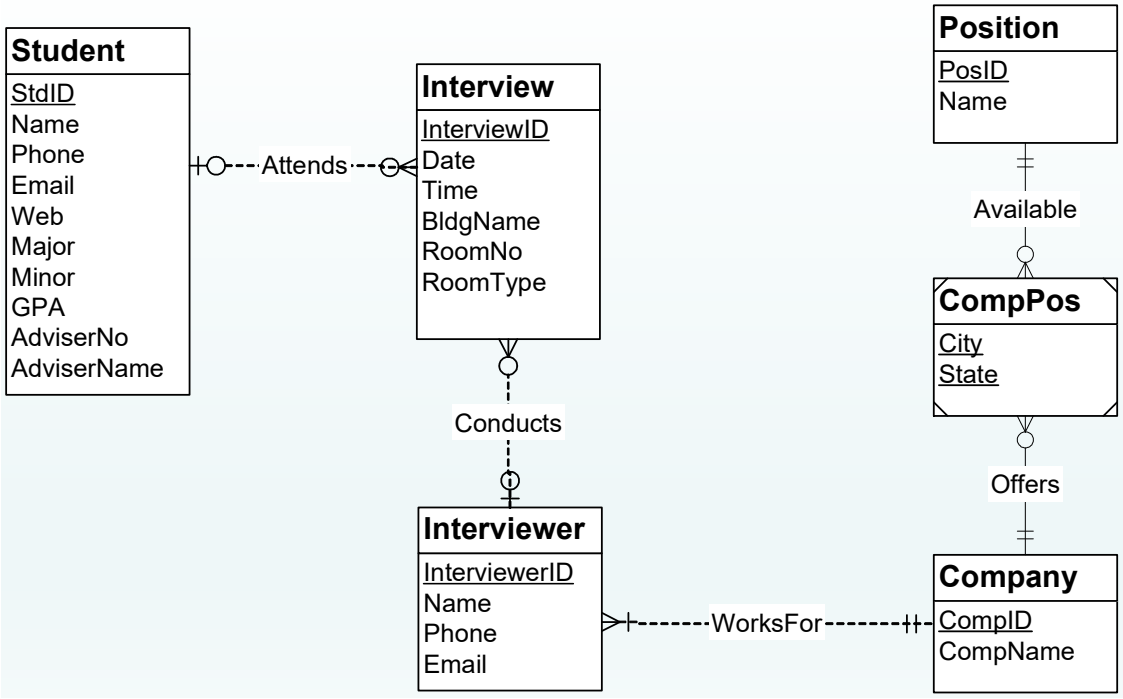
Big University Table

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

FD Falsification List

FD	Falsifications
$StdCity \rightarrow OfferNo$	(1,2), (3,4)
$StdCity \rightarrow OffTerm$	(3,4)
$StdCity \rightarrow EnrGrade$	?, ?
$StdCity \rightarrow CourseNo$	?, ?
$StdCity \rightarrow CrsDesc$	?, ?
$StdCity \rightarrow OffYear$	?, ?
$StdCity \rightarrow StdNo$	None
$StdCity \rightarrow StdClass$	?,?

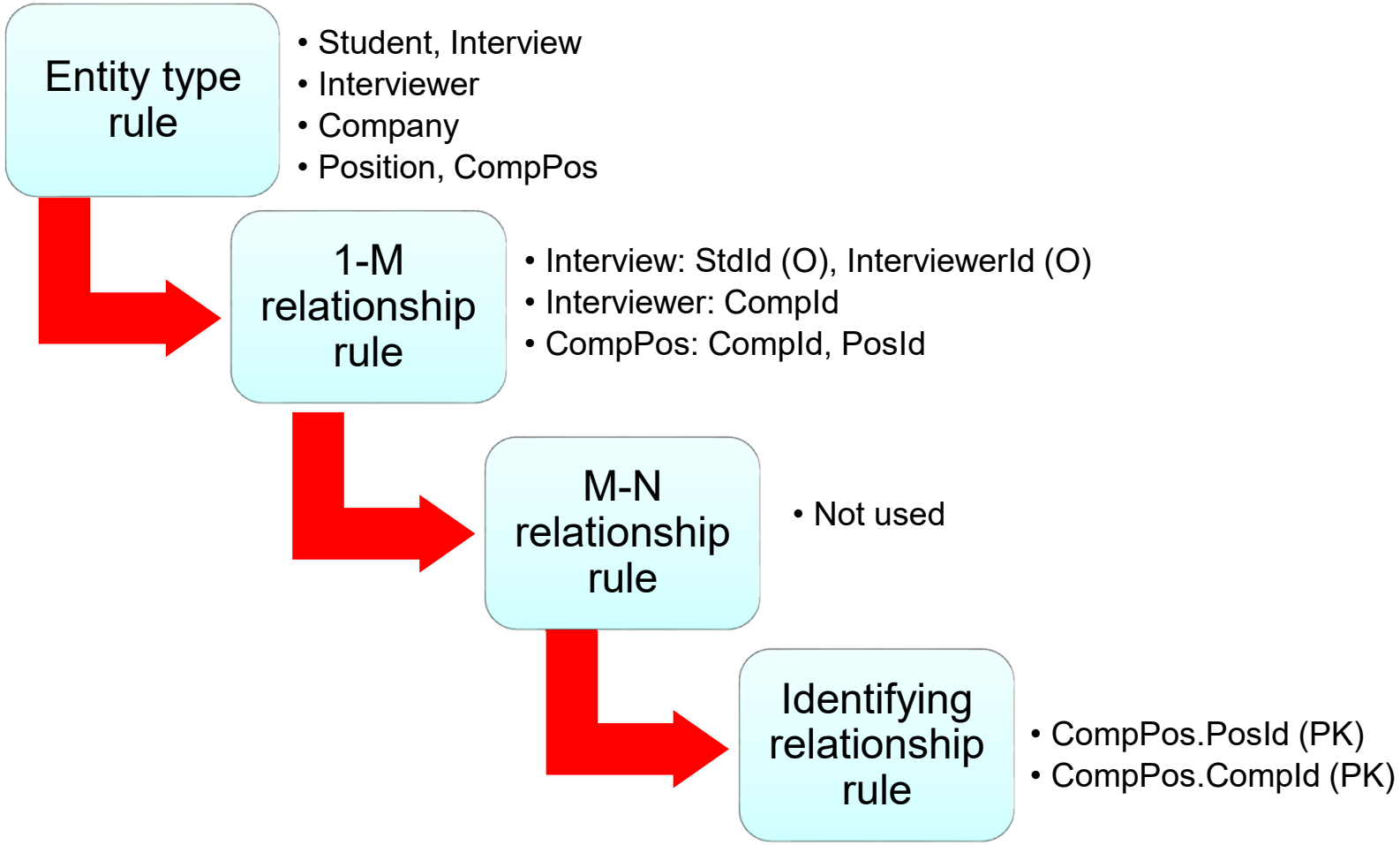
Conversion/Normalization Problem



Problem requirements

- Convert the ERD into tables using the conversion rules
- For each table, list FDs and split if the table violates BCNF.

Conversion Rule Application



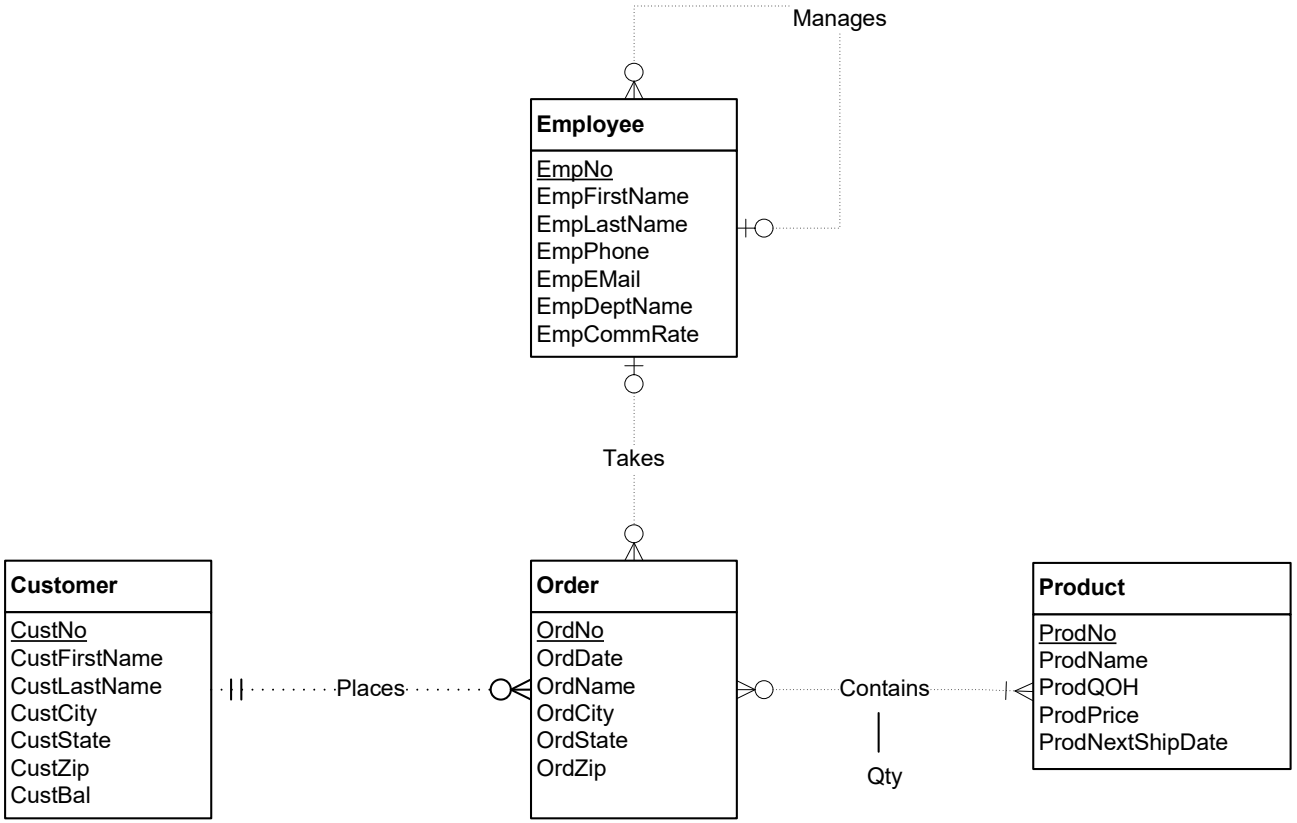
Additional Normalization

- Only list FDs not implied by PKs
- Additional FDs
  - AdviserNo → AdviserName
  - Possible FD: BldgName, RoomNo → RoomType
  - Possible FD: RoomNo → BldgName, RoomType

Summary

- Practice using sample rows to falsify FDs
- Practice combining conversion and normalization
- Useful practical skills

Practice Conversion Problem



Conversion Rule Application