15CSE302 Database Management Systems
Lecture 14 **Normalization**

B.Tech /III Year CSE/V Semester   **L T P C 2 0 2 3**

DBMS Team
Dr G Jeyakumar
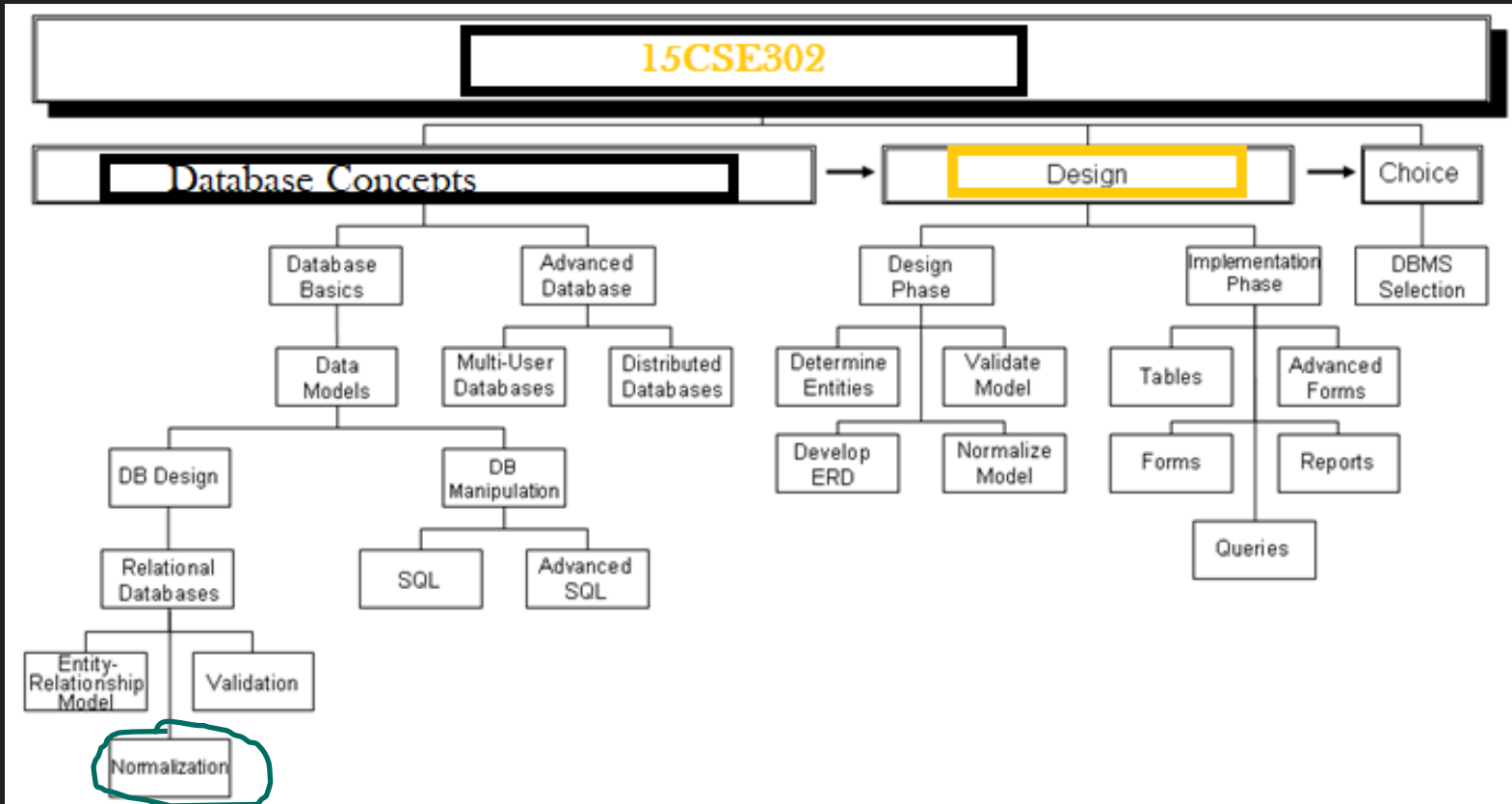**Bindu K R**
Dr Priyanka Kumar
R. Manjusha
Department of CSE
Amrita School of Engineering

Slides Courtesy : Carlos Alvarado,
San Jose State University

Syllabus

Brief Recap of Previous Lecture

- Enhanced ER Diagram
- Examples

Normalization

Overview

Today we'll discuss

- Database Normalization
    - Data Anomalies Caused by:
        - Update, Insertion, Deletion
- Brief History/Overview
    - 1st Normal Form
    - 2nd Normal Form
    - 3rd Normal Form
    - Conclusion

Why Database Normalization?

The main goal of Database Normalization is to restructure the logical data model of a database to:

- Eliminate redundancy

- Organize data efficiently

- Reduce the potential for data anomalies

NORMALIZATION

- Normalization is **formal method to check tables for potential data storage problems termed anomalies.**
- In order to clarify the discussion, we will formalize the definition of several terms.
  - **KEYS:** The term KEY is often confusing because it has different meanings during design and implementation of a system.

  - **DESIGN: During design, KEY means a combination of one or more attributes (columns) of a relational table that uniquely identify rows in the table.**

  - **KEY guarantees uniqueness; no two rows can be identical.**

.

NORMALIZATION

IMPLEMENTATION:

- During implementation, the term **KEY** is a column on which the DBMS builds an index or other data structure, to allow quick access to rows.

- Such keys need not be unique - they may be secondary keys enabling access to a SET of rows.

.

Data Anomalies

- **Data anomalies** are inconsistencies in the data stored in a database as a result of an operation such as update, insertion, and/or deletion.

- Such inconsistencies may arise when have a particular record stored in multiple locations and not all of the copies are updated.

- We can prevent such anomalies by implementing different level of normalization called Normal Forms (NF)

Student

ROLLNO	NAME	BRANCH	HOD	OFFICE_TEL
401	Asha	CSE	Mr. Sam	53337
402	Babu	CSE	Mr. Sam	53337
403	Cini	CSE	Mr. Sam	53337
404	Dilip	CSE	Mr. Sam	53337

Employee

EmpId	Name	Dept	Salary	Course	DateTook	Fee
130	Margaret	Math	45,000	Calculus	01/15	150
130	Margaret	Math	45,000	Biology	02/15	200
200	Susan	Sci	38,000	Biology	01/15	200
250	Chris	Math	52,000	Calculus	03/15	150
250	Chris	Math	52,000	Biology	03/15	200
425	Bill	Math	48,000	Algebra	03/15	200
425	Bill	Math	48,000	Calculus	04/15	

Normalisation

Problems With This Table:

- Redundancy of data storage.

- Potential inconsistencies on updating data.

What is the Primary Key of this table?

- To determine an appropriate key, you should first examine the **FDs**.
- **EmpId -> Name, Dept, Salary**
  Course -> Fee
  EmpId, Course -> DateTook

- Assuming employees only take a course once (no time dependencies), then uniqueness for the rows in the table is ensured by a composite key of **EmpId + Course**.

Data Anomalies

Data Anomalies are problems with data storage caused by poorly structured tables.

Insertion Anomaly

- If the primary key is **EmpId + Course**, to add a new employee, the employee must first be enrolled in a course.
- If an employee is not enrolled in a course, then the **COURSE** column that is part of the composite primary key will be **null**, and null key values are **not** allowed.

Data Anomalies

Deletion Anomaly.

▸ Deleting data for Employee #425 (Bill) causes us to lose data about Algebra and the course fee for Algebra because Bill is the only employee who has enrolled in Algebra.

Modification Anomaly.

▸ If the fee for Calculus is increased, the data must be updated for more than one row.

▸ Note there is also a time-sensitivity between **EmpId** and **Course** since an employee could take a course many times, but the table does not track this fact

Brief History/Overview

- Database Normalization was first proposed by Edgar F. Codd.

- Codd defined the first three Normal Forms, which we'll look into, of the 7 known Normal Forms.

- In order to do normalization we must know what the requirements are for each of the three Normal Forms that we'll go over.

- One of the key requirements to remember is that Normal Forms are progressive.

- That is, in order to have 3rd NF we must have 2nd NF and in order to have 2nd NF we must have 1st NF.

First Normal Form (1NF)-The Requirements

The requirements to satisfy the First Normal Form (1NF):

➢ **Each table has a primary key:** minimal set of attributes which can uniquely identify a record

➢ **The values in each column of a table are atomic** (No multi-valued attributes allowed).

➢ **There are no repeating groups:** two columns do not store similar information in the same table.

First Normal Form(1NF) - Example

Un-normalized Students table:

Student#	AdvID	AdvName	AdvRoom	Class1	Class2
123	123A	James	555	102-8	104-9
124	123B	Smith	467	209-0	102-8

First Normal Form(1NF) - Example

Un-normalized Students table:

Student#	AdvID	AdvName	AdvRoom	Class1	Class2
123	123A	James	555	102-8	104-9
124	123B	Smith	467	209-0	102-8

Normalized Students table:

Student#	AdvID	AdvName	AdvRoom	Class#
123	123A	James	555	102-8
123	123A	James	555	104-9
124	123B	Smith	467	209-0
124	123B	Smith	467	102-8

Second Normal Form(2NF) -The Requirements

The requirements to satisfy the Second Normal Form

- All requirements for First Normal Form must be met.

- Redundant data across multiple rows of a table must be moved to a separate table.

  - The resulting tables **must be related to each other by use of foreign key**.

Second Normal Form(2NF) -Example

Student Table			
Student#	AdvID	AdvName	AdvRoom
123	123A	James	555
124	123B	Smith	467

Registration Table	
Student#	Class#
123	102-8
123	104-9
124	209-0
124	102-8

References

- Hillyer Mike, MySQL AB. An Introduction to Database Normalization, http://dev.mysql.com/tech-resources/articles/intro-to-normalization.html, accessed October 17, 2006.

- Microsoft. Description of the database normalization basics, http://support.microsoft.com/kb/283878 , accessed October 17, 2006.

- Wikipedia. Database Normalization. http://en.wikipedia.org/wiki/Database_normalization.html , accessed October 17, 2006. https://www.db-book.com/db6/index.html
- https://www.youtube.com/watch?v=mfVCesoMaGA&list=PLroEs25KGvwzmvIxYHRhoGTz9w& LeXekO&index=22

Summary

- **Normalization basics**
- **Anomalies**

Next Lecture
Functional dependency

Thank You

Happy to answer any questions ! ! !