

# Storage and Data Analysis

CAC220 Spring 2024

---

R. Steven Wingo, Ph.D.

March 12, 2024

Birmingham Southern College

1. Quiz
2. Database Design<sup>2</sup>
  - First, Second, and Third Normal Form
  - Boyce-Codd Normal Form
3. Assignment

## QUIZ

---

# **DATABASE DESIGN<sup>2</sup>**

---

# TERMINOLOGY

## Functional Dependence

Dependence of one column on another is called functional dependence.

## Multivalued Dependence / Join Dependence

Multivalued dependence and join dependence entail dependencies between three or more columns.

## Redundancy

Redundancy is the repetition of related values in a table.

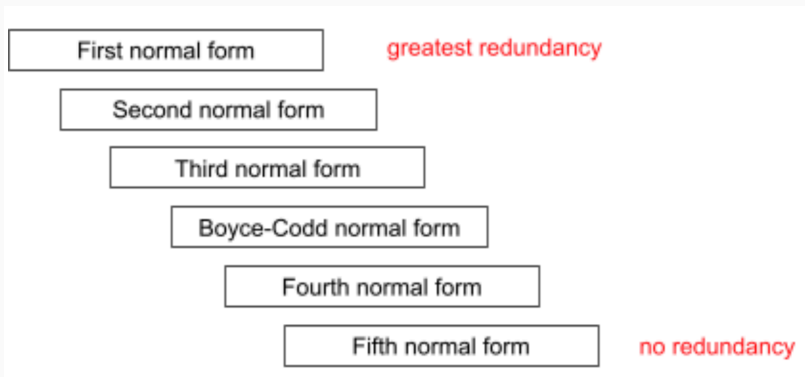
# PARTICIPATION ACTIVITY 4.10.1 AND 4.10.2

## Functional Dependence

Booking					
● PassengerNumber	→ PassengerName	● FlightCode	FareClass	→ BoardingZoneNumber	
222	Elvira Yin	AZ312	First	1	
222	Elvira Yin	BF999	Economy	3	
222	Elvira Yin	GC848	Business	3	
333	Deepak Chopra	GC848	First	1	
444	Mary Hutcher	GC848	First	1	

# NORMAL FORMS

Normal forms are rules for designing tables with less redundancy.



## NORMALIZATION<sup>1</sup>

### First Normal Form (1NF)

1. Using row order to convey information is not permitted
2. Mixing data types within the same column is not permitted
3. Having a table without a primary key is not permitted
4. Repeating groups are not permitted

### Second Normal Form (2NF)

Each non-key attribute in the table must be dependent on the entire primary key.

### Third Normal Form (3NF)

Each non-key attribute in the table must depend on the key, the whole key, and nothing but the key.

### Boyce-Codd Normal Form (BCNF)

Each attribute in the table must depend on the key, the whole key, and nothing but the key.

### Fourth Normal Form (4NF)

The only kinds of multivalued dependency allowed in a table are multivalued dependencies on the key.

### Fifth Normal Form (5NF)

It must not be possible to describe the table as being the logical result of joining some other tables together.



## FIRST NORMAL FORM

Every cell of a table contains exactly one value. A table is in first normal form when, in addition, the table has a primary key.

Passenger	
● PassengerNumber	PassengerName
222	Elvira Yin
829	John Miller
333	Deepak Chopra
444	Mary Hatcher

First normal form

1. The Passenger table has a primary key and therefore is in first normal form.
2. Since the primary key is unique, each primary key value appears on one row only.
3. Since each cell contains one value, each primary key value is related to one PassengerName value.
4. Therefore, PassengerName depends on PassengerNumber.

**PARTICIPATION ACTIVITY 4.10.6**

# First Normal Form

☐ True

☐ False

## SECOND NORMAL FORM

A table is in second normal form when all non-key columns depend on the whole primary key.

● PassengerNumber	PassengerName	● FlightCode	FareClass	BoardingZoneNumber
222	Elvira Yin	AZ312	First	1
222	Elvira Yin	BF999	Economy	3
222	Elvira Yin	GC848	Business	3
333	Deepak Chopra	GC848	First	1
444	Mary Hutcher	GC848	First	1

First normal form

- PassengerName depends on PassengerNumber.
- Dependence on part of the primary key causes repetition of (222, Elvira Yin).

## SECOND NORMAL FORM

A table is in second normal form when all non-key columns depend on the whole primary key.

Booking				Passenger	
● PassengerNumber	● FlightCode	FareClass	BoardingZoneNumber	● PassengerNumber	PassengerName
222	AZ312	First	1	222	Elvira Yin
222	BF999	Economy	3	333	Deepak Chopra
222	GC848	Business	3	444	Mary Hatcher
333	GC848	First	1		
444	GC848	First	1		

Second normal form

- PassengerName moves to the Passenger table.
- The Passenger table has no redundancies, since PassengerName depends on the whole primary key.

# PARTICIPATION ACTIVITY 4.10.8

## Second Normal Form

StudentRecord				
• StudentNumber	• CourseNumber	ScoreNumber	EmailAddress	GradeLetter
8034	Math 100	95	john@gmail.com	A
2111	Math 100	73	sammy@icloud.com	C
9930	Spanish 22A	89	abc@hotmail.com	B
8034	History 11	89	john@gmail.com	B
5091	Biology 200B	41	maria@sbcc.net	F

## THIRD NORMAL FORM

A table is in third normal form when all non-key columns depend on the key, the whole key, and nothing but the key.

● PassengerNumber	● FlightCode	FareClass	BoardingZoneNumber
222	AZ312	First	1
222	BF999	Economy	3
222	GC848	Business	3
333	GC848	First	1
444	GC848	First	1

Second normal form

- BoardingZoneNumber depends on FareClass.
- Dependence on a non-key column causes the redundancy (First, 1).
- Removing BoardingZoneNumber from Booking eliminates redundancy.

## THIRD NORMAL FORM

A table is in third normal form when all non-key columns depend on the key, the whole key, and nothing but the key.

Booking		
● PassengerNumber	● FlightCode	FareClass
222	AZ312	First
222	BF999	Economy
222	GC848	Business
333	GC848	First
444	GC848	First

Third normal form

Fare	
● FareClass	BoardingZoneNumber
First	1
Economy	3
Business	3

Third normal form

- BoardingZoneNumber moves to the Fare table.
- The Fare table has no redundancies since BoardingZoneNumber depends on the primary key.

# PARTICIPATION ACTIVITY 4.10.10

## Third Normal Form

StudentRecord

• StudentNumber	• CourseNumber	ScoreNumber	GradeLetter
8034	Math 100	95	A
2111	Math 100	88	B
9930	Spanish 22A	72	C
8034	History 11	88	B
5091	Biology 200B	41	F



## **TERMINOLOGY**

### **Candidate Key / Minimal**

A candidate key is a simple or composite column that is unique and minimal. Minimal means all columns are necessary for uniqueness.

### **Non-Key**

A non-key column is a column that is not contained in a candidate key.

### **Third Normal Form**

A table is in third normal form if, whenever a non-key column A depends on column B, then B is unique. Columns A and B may be simple or composite.

## BOYCE-CODD NORMAL FORM

A table is in Boyce-Codd normal form if, whenever column A depends on column B, then B is unique. Columns A and B may be simple or composite

EmployeeProjectTask		
EmployeeID	ProjectName	TaskName
489	Performance reviews	Review Jiho Chen
489	Training	Corporate ethics
600	Training	Security basics
600	Performance reviews	Review Jiho Chen
777	Performance reviews	Review Jiho Chen
835	Performance reviews	Review Maria Rodriguez

Third normal form

- Each employee is assigned at most one task on each project.

## BOYCE-CODD NORMAL FORM

A table is in Boyce-Codd normal form if, whenever column A depends on column B, then B is unique. Columns A and B may be simple or composite

EmployeeID	ProjectName	TaskName
489	Performance reviews	Review Jiho Chen
489	Training	Corporate ethics
600	Training	Security basics
600	Performance reviews	Review Jiho Chen
777	Performance reviews	Review Jiho Chen
835	Performance reviews	Review Maria Rodriguez

- TaskName depends on the primary key (EmployeeID, ProjectName) but not on EmployeeID or ProjectName.
- Since every non-key column depends on a unique column, EmployeeProjectTask is in third normal form.

## BOYCE-CODD NORMAL FORM

A table is in Boyce-Codd normal form if, whenever column A depends on column B, then B is unique. Columns A and B may be simple or composite

EmployeeProjectTask		
● EmployeeID	● ProjectName	← TaskName
489	Performance reviews	Review Jiho Chen
489	Training	Corporate ethics
600	Training	Security basics
600	Performance reviews	Review Jiho Chen
777	Performance reviews	Review Jiho Chen
835	Performance reviews	Review Maria Rodriguez

Third normal form

- In this company, TaskNames are never repeated on different projects.
- So ProjectName depends on TaskName, which is not unique.
- Since ProjectName depends on a non-unique column, EmployeeProjectTask is not in Boyce-Codd normal form and contains redundancy.

## BOYCE-CODD NORMAL FORM

A table is in Boyce-Codd normal form if, whenever column A depends on column B, then B is unique. Columns A and B may be simple or composite

EmployeeTask		TaskProject	
EmployeeID	TaskName	TaskName	ProjectName
489	Review Jiho Chen	Review Jiho Chen	Performance reviews
489	Corporate ethics	Review Maria Rodriguez	Performance reviews
600	Security basics	Corporate ethics	Training
600	Review Jiho Chen	Security basics	Training
777	Review Jiho Chen		
835	Review Maria Rodriguez		

- Decomposing to EmployeeTask and TaskProject eliminates the redundancy.

## BOYCE-CODD NORMAL FORM

A table is in Boyce-Codd normal form if, whenever column A depends on column B, then B is unique. Columns A and B may be simple or composite

EmployeeTask		TaskProject	
●EmployeeID	●TaskName	●TaskName	ProjectName
489	Review Jiho Chen	Review Jiho Chen	Performance reviews
489	Corporate ethics	Review Maria Rodriguez	Performance reviews
600	Security basics	Corporate ethics	Training
600	Review Jiho Chen	Security basics	Training
777	Review Jiho Chen		
835	Review Maria Rodriguez		

Boyce-Codd normal form

- EmployeeTask contains no dependencies. TaskProject contains one dependency, ProjectName on the unique column TaskName.

## PARTICIPATION ACTIVITY 4.12.4

Match the activity with the resulting normal form.

First or second normal form	Boyce-Codd normal form	Second normal form
		Consider denormalizing tables in reporting databases.
		Eliminate dependencies on columns contained within a composite unique column.
		Eliminate dependencies on non-unique columns.

# ASSIGNMENT

---



# UPDATED ASSIGNMENTS

**Project:**

For March 14<sup>th</sup> turn in the following for grading

1. SQL Scripts to create and populate database tables.
2. SQL Queries to perform actions for the selected Story Cards
3. Next Story Cards selected for implementation.

# CONTINUED ASSIGNMENTS

## Reading:

For March 19 Midterm<sup>th</sup>

Read Chapters 1–4 in our course textbook

(<https://learn.zybooks.com/zybook/BSCCAC220WingoSpring2024>)

There will be a midterm review in our March 19<sup>b</sup> class.

## Essay:

One thousand word essay about how the use of data is changing a field in which you are interested (i.e., stock market, sports, healthcare).

**Topic Choice:** March 14<sup>th</sup>

**Due Date:** April 23<sup>rd</sup>

### REFERENCES

---

- [1] Decomplexify. ***Learn Database Normalization - 1NF, 2NF, 3NF, 4NF, 5NF***. Nov. 2021. URL: [https://www.youtube.com/watch?v=GFQaEYE8\\_8](https://www.youtube.com/watch?v=GFQaEYE8_8).
- [2] zybooks. ***CAC 220: Storage and Analysis of Data***. zyBooks, a Wiley brand, 2024. ISBN: 9798203315762.

Material for slides uses the course textbook<sup>2</sup>