



Search content

Search

View content explorer

Configure zyBook

### Table of contents

[About this material](#)

1. SET 1  

2. SET 1 - 1:INTRODUCTION  

3. SET 1 - 2:DATA MODELING - WHAT - ERD MODEL   

4. SET 1 - 3:DATA MODELING HOW - RELATIONAL MO...   

5. SET 1 - 4:ERD to RELATIONAL   

6. SET 1 - 5:NORMALIZATION     

## COSC 3380: Database Systems

Spring 2024

### Class information

115 students subscribed [Subscription instructions](#)

**Manage**

Add teaching assistants, sections, email domains, & drop/manage student sections

### Instructors

Victoria Hilford

### Teaching assistants

Jordan Yu  
Alvaro Urtaza  
Fernando Ramirez

### Class sections

No class sections

02.07.2024

**TA Download**

(7 - We)

**ZyBook SET 1 Sections**

**(4 PM)**

**(PART of 30 points)**

**EXAM 1 Review**

**(PART of 20 points)**

**DATABASE SYSTEMS - ZYBOOK [30 points]**

30% of Total



**SET 1: Sections 1 - 6 [ZyBook] (100 points) (complete it anytime before EXAM REVIEW class)**

Available until Feb 7 at 4:00pm | Due Feb 7 at 4pm | 100 pts



**Only 115 out of 120 registered.**

**COSC 3380 Spring 2024**

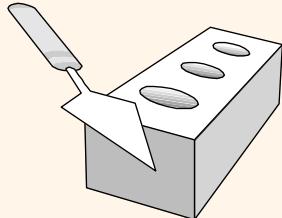
**Database Systems**

**M & W 4:00 to 5:30 PM**

Prof. **Victoria Hilford**

**PLEASE TURN your webcam ON (must have)**

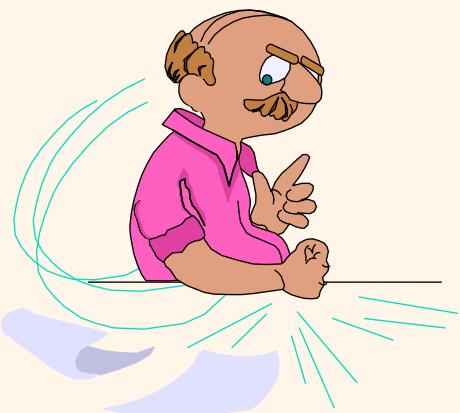
**NO CHATTING during LECTURE**



# COSC 3380

## 4 to 5:30

**PLEASE  
LOG IN  
CANVAS**



Please close all other windows.

01.31.2024

ZyBook SET 1 - 5

(5 - We)

Set 1

LECTURE 5 NORMALIZATION

02.05.2024

(6 - Mo)

EXAM 1 Practice

(PART of 20 points)

02.07

(7)

TA Download

ZyBook SET 1 Sections

(4 PM)

(PART of 30 points)

EXAM 1 Review

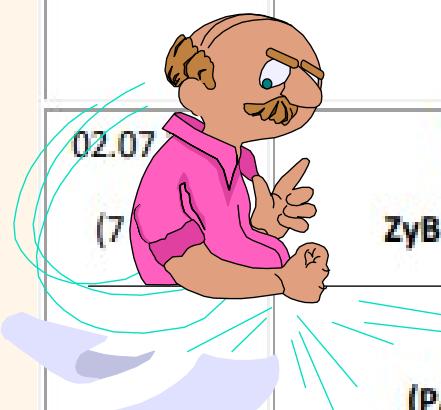
(PART of 20 points)

02.12.2024

(8 - Mo)

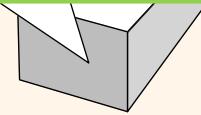
EXAM 1

(PART of 50 points)



ta

# From 4:00 to 4:07 PM – 5 minutes.



01.31.2024  
(5 - We)

## ZyBook SET 1 - 5

Set 1

### LECTURE 5 NORMALIZATION

CLASS PARTICIPATION 20 points

20% of Total + :

## CLASS 5



### CLASS 5 BEGIN PARTICIPATION

Not available until Jan 31 at 4:00pm | Due Jan 31 at 4:07pm | 100 pts

VH, publish



This is a synchronous online class.

Attendance is required.

Recording or distribution of class materials is prohibited.

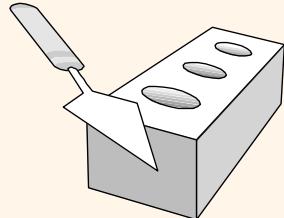
1. At the beginning of selected classes there is an assessment in the first 10 minutes. (beige BOX in the Detailed Syllabus)

2. At the end of selected classes there is an assessment in the last 10 minutes. (blue BOX in the Detailed Syllabus)

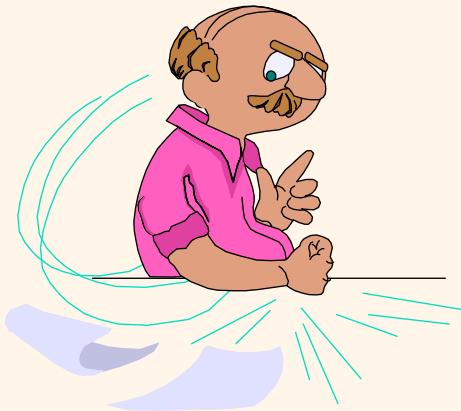
3. ZyBook sections will be downloaded and used for 30% of Total Score on the dates specified in the Detailed Syllabus.

4. EXAMS are in CANVAS. No late EXAMS.

5. I have to be present in TEAMS in order to take any graded assignment assigned during that class.

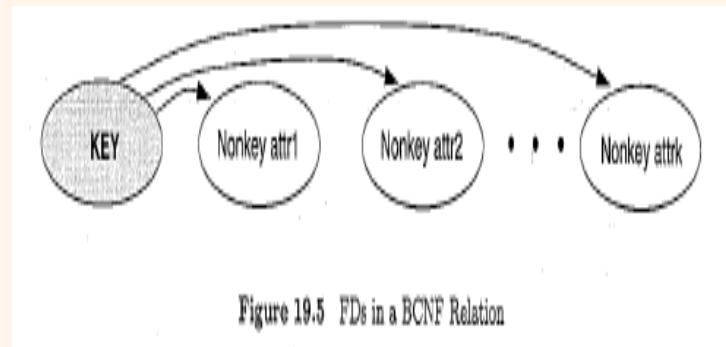


# COSC 3380



## Class 5

01.31.2024	ZyBook SET 1-5	Set 1
(5 - We)		LECTURE 5 NORMALIZATION



From 4:07 to 5:00 PM – 50 minutes.

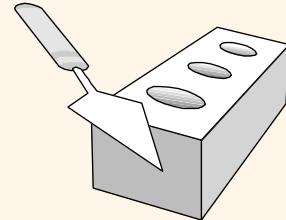


## Lecture 5

# *Relations Normalization*

6. SET 1 - 5:NORMALIZATION

# Relation



- ❖ A **Database** is a collection of one or more **Relations**, where each **Relation** is a **Table** with rows and columns.

SUPPLIER						Columns (Attributes, Fields)
Supplier_Number	Supplier_Name	Supplier_Street	Supplier_City	Supplier_State	Supplier_Zip	Rows (Records, Tuples)
8259	CBM Inc.	74 5 <sup>th</sup> Avenue	Dayton	OH	45220	
8261	B. R. Molds	1277 Gandolly Street	Cleveland	OH	49345	
8263	Jackson Composites	8233 Micklin Street	Lexington	KY	56723	
8444	Bryant Corporation	4315 Mill Drive	Rochester	NY	11344	

Key Field  
(Primary Key)

# Functional Dependency ( $X \rightarrow Y$ )

## Relation

**Hourly\_Emps** (ssn, name, lot, *rating*, *hrly\_wages*, *hrs\_worked*)

S	N	L	R	W	H
123-22-3666	Attishoo	48	8	10	40
231-31-5368	Smiley	22	8	10	30
131-24-3650	Smethurst	35	5	7	30
434-26-3751	Guldu	35	5	7	32
612-67-4134	Madayan	35	8	10	40

$$R \rightarrow W$$

## 6.1 Normal form

# Functional Dependency ( $X \rightarrow Y$ )

## Relation

Hourly\_Emps (ssn, name, lot, rating, hrly\_wages, hrs\_worked)

S	N	L	R	W	H
123-22-3666	Attishoo	48	8	10	40
231-31-5368	Smiley	22	8	10	30
131-24-3650	Smethurst	35	5	7	30
434-26-3751	Guldu	35	5	7	32
612-67-4134	Madayan	35	8	10	40

R → W

### Redundancy and dependence

**Redundancy** is the repetition of related values in a table. Redundancy causes database management problems. When related values are updated or deleted, all copies must be changed, which makes queries slow and complex. If copies are not updated or deleted uniformly, the copies become inconsistent and the correct version is uncertain.

Column A **depends on** column B means each B value is related to at most one A value. Columns A and B may be simple or composite. 'A depends on B' is denoted ' $B \rightarrow A$ '

PARTICIPATION  
ACTIVITY

6.1.1: Examples of redundancy and dependence.

Booking					
PassengerNumber	→	PassengerName	FlightCode	→	BoardingZoneNumber
222		Elvira Yin	AZ312		1
222		Elvira Yin	BF999		3
222		Elvira Yin	GC848		3
333		Deepak Chopra	GC848		1
444		Mary Hatcher	GC848		1

# (CA 6.2.1 – Step 1 – Functional Dependencies)



CHALLENGE  
ACTIVITY

6.2.1: Normal form.



• ISO	CountryName	ContinentCode	ContinentName
TM	Turkmenistan	AS	Asia
MA	Morocco	AF	Africa
LC	Saint Lucia	NA	NAmerica
PT	Portugal	EU	Europe
ID	Indonesia	AS	Asia

What columns depend on ISO?



CountryName



ContinentCode



ContinentName

ISO is the primary key of Country, so all columns depend on ISO

1

?????

# TA time (Alvaro) – 4 minutes

## (CA 6.2.1 – Step 1 – Functional Dependencies)

CHALLENGE  
ACTIVITY

6.2.1: Normal form.



Country			
• ISO	CountryName	CapitalName	ContinentName
NI	Nicaragua	Managua	NAmerica
MZ	Mozambique	Maputo	Africa
NO	Norway	Oslo	Europe
SR	Suriname	Paramaribo	SAmerica
SY	Syria	Damascus	Asia

What columns depend on ISO?

- CountryName     CapitalName     ContinentName

What columns depend on ContinentName?

- ISO     CountryName     CapitalName

1



ta

hrke

A key for a relation R is a minimal set of attributes that uniquely determines all other attributes.

Think of each tuple as being identified by a key and described by the remaining attributes.



Figure 19.5 FDs in a BCNF Relation

Convention:  $X, Y$  are sets of attributes  
 $A, B$  are single attributes

$X$  is a superkey if it contains a key for  $R$ .

$A$  is part of a candidate key if  $A$  is part of some key for  $R$ .

Attribute closure of  $X$  (denoted  $X^+$ ):

closure =  $X$ ;

repeat until there is no change:

{  
if there is an FD  $U \rightarrow V$  in  $F$  such that  $U \subseteq \text{closure}$ ,  
then set closure = closure  $\cup$   $V$

}  
The closure of a candidate key will return all the attributes in the relation.  
The closure of a primary key will return all the attributes in the relation.  
The closure of a key will return all the attributes in the relation.

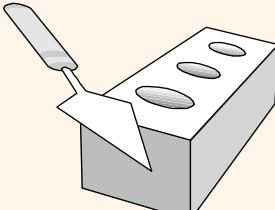
2. Given relation R with four attributes ABCD and the set of FD s:

$$C \rightarrow D$$

$$C \rightarrow A$$

$$B \rightarrow C$$

is B a candidate key?

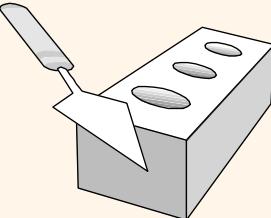


Primary key

>	sid	name	email	login	age	gpa
50000	Dave	dave@cs			19	3.3
53666	Jones	jones@cs			18	3.4
53688	Smith	smith@ee			18	3.2
53650	Smith	smith@math			19	3.8
53831	Madayan	madayan@music			11	1.8
53832	Guldu	guldu@music			12	2.0

Students (Referenced relation)





Attribute closure of X (denoted  $X^+$ ):

closure = X;

repeat until there is no change:

{

if there is an FD  $U \rightarrow V$  in F such that  $U \subseteq \text{closure}$ ,  
then set  $\text{closure} = \text{closure} \cup V$

}

The closure of a candidate key will return **all the attributes** in the relation.

The closure of a primary key will return **all the attributes** in the relation.

The closure of a key will return **all the attributes** in the relation.

2. Given relation R with four attributes ABCD and the set of FD s:

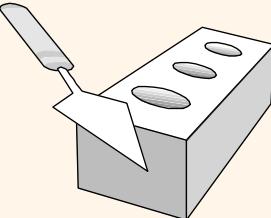
- (1)  $C \rightarrow D$
- (2)  $C \rightarrow A$
- (3)  $B \rightarrow C$

is B a candidate key?

Use this template

	A	B	C	D
$\{B\}^+$				

$$\{B\}^+ = \{ , , , \}$$



Attribute closure of X (denoted  $X^+$ ):

closure = X;

repeat until there is no change:

{

if there is an FD  $U \rightarrow V$  in F such that  $U \subseteq \text{closure}$ ,  
then set  $\text{closure} = \text{closure} \cup V$

}

The closure of a candidate key will return **all the attributes** in the relation.

The closure of a primary key will return **all the attributes** in the relation.

The closure of a key will return **all the attributes** in the relation.

2. Given relation R with four attributes ABCD and the set of FD s:

- (1)  $C \rightarrow D$
- (2)  $C \rightarrow A$
- (3)  $B \rightarrow C$

is B a candidate key?

Use this template

	A	B	C	D
$\{B\}^+$	$A_{(2)}$	B	$C_{(3)}$	$D_{(1)}$

yes

$$\{B\}^+ = \{A, B, C, D\}$$

**B is a candidate key**

# Normal Forms



## Normal Forms

NORMAL FORM	CHARACTERISTIC
First normal form (1NF)	Table format, no repeating groups, and PK identified
Second normal form (2NF)	1NF and no partial dependencies
Third normal form (3NF)	2NF and no transitive dependencies
Boyce-Codd normal form (BCNF)	Every determinant is a candidate key (special case of 3NF)

**Let's Start with BCNF**

# Boyce Codd Normal Form (BCNF)

Given **Relation** R with FDs (1)  $C \rightarrow D$  (2)  $C \rightarrow A$  (3)  $B \rightarrow C$

- ❖ **Relation** R with FDs F is in **BCNF** if, for all X
  - A ⊂ X (called a *trivial* FD), or
  - X is a key for R.
- ❖ Think of each **tuple** as being identified by a key and described by the remaining attributes. One

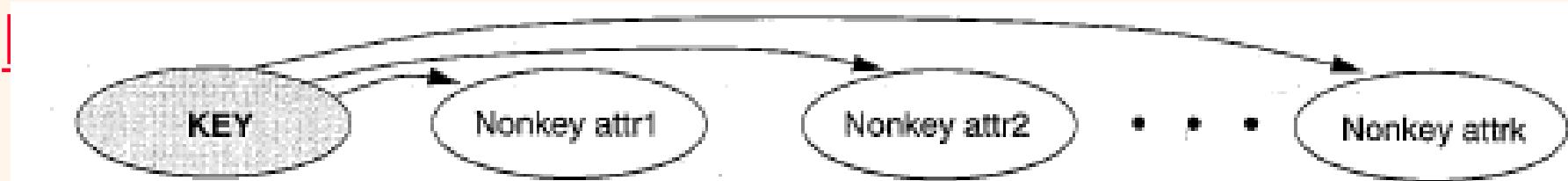


Figure 19.5 FDs in a BCNF Relation

# Boyce Codd Normal Form (BCNF)

Given Relation R with FDs (1)  $C \rightarrow D$  (2)  $C \rightarrow A$  (3)  $B \rightarrow C$

❖ Relation R with FDs F is in BCNF if, for all  $X \rightarrow A$  in F

- $A \in X$  (called a *trivial* FD), or
- $X$  is a key for R.

❖ Think of each tuple as being identified by a key and described by the remaining attributes. One key for simplicity.

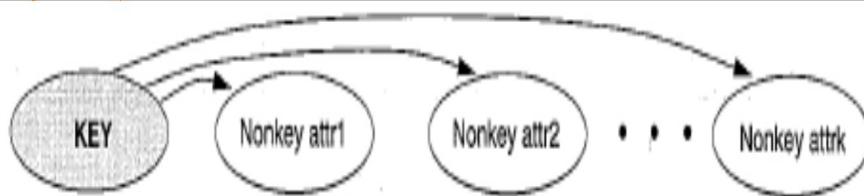


Figure 19.5 FDs in a BCNF Relation

## Boyce-Codd normal form

The definition of third normal form applies to non-key columns only, which allows for occasional redundancy. Boyce-Codd normal form applies to all columns and eliminates this redundancy.

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. This definition is identical to the definition of third normal form with the term 'non-key' removed.

Boyce-Codd normal form is considered the gold standard of table design. Although fourth and fifth normal forms remove additional types of redundancy, these redundancies are uncommon and of little practical concern.

PARTICIPATION ACTIVITY

6.1.10: Boyce-Codd normal form eliminates all common redundancy.



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

BCNF?

No

# Is R in BCNF?

Given **Relation** R with FDs  $(1) C \rightarrow D$      $(2) C \rightarrow A$      $(3) B \rightarrow C$

$\{X\}^+$  key , i.e., must contain all attributes of the Relation R

	A	B	C	D
X				
$(1) \{C\}^+ \rightarrow D$	$A_{(2)}$		C	D
$(2) \{C\}^+ \rightarrow A$				
$(3) \{B\}^+ \rightarrow C$				

A **key** for a relation R is a minimal set of attributes that uniquely determines all other attributes.  
Think of each tuple as being identified by a **key** and described by the remaining attributes.



Figure 10-8 FDs in a BCNF Relation

Convention: X<sub>Y</sub> are sets of attributes  
A<sub>B</sub> are single attributes  
X is a **superkey** if it contains a **key** for R.  
A is part of a **candidate key** if A is part of some **key** for R.

Attribute closure of X (denoted  $X^+$ ):  
 $closure = X$   
repeat until there is no change:  
if there is an FD  $U \rightarrow V$  in F such that  $U \subseteq closure$ ,  
then set  $closure = closure \cup V$

The closure of a candidate key will return all the attributes in the relation.  
The closure of a primary key will return all the attributes in the relation.  
The closure of a foreign key will return all the attributes in the relation.

BCNF?

NO!

WHY?

C is not a key!

X

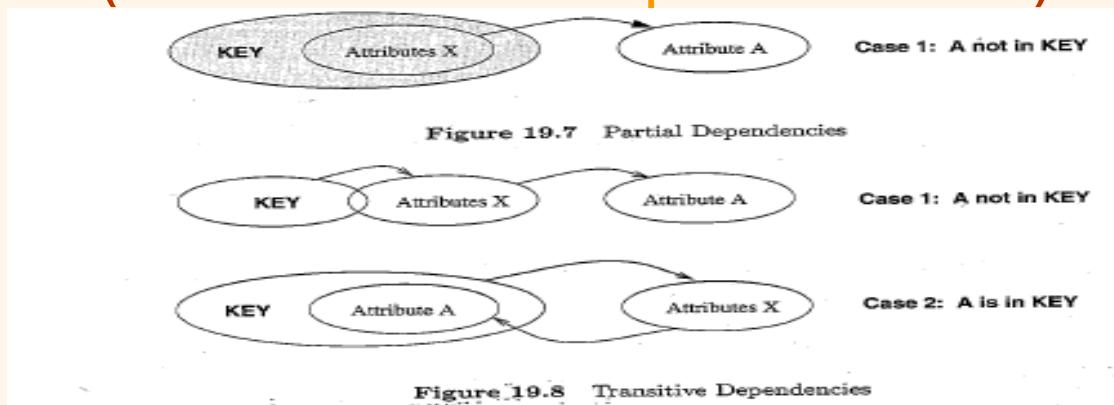
❖ Relation R with FDs F is in BCNF if, for all  $X \rightarrow A$  n F

- $A \in X$  (called a *trivial* FD), or
- $X$  is a **key** for R.

# Third **Normal Form** (3NF)

Given **relation** R with FDs (1)  $C \rightarrow D$  (2)  $C \rightarrow A$  (3)  $B \rightarrow C$

- ❖ **Relation R with FDs F is in 3NF if, for all A in  $\mathcal{F}$** 
  - $X$  is a **key** for R, or
  - $A$  contains a **key** ( $X$  is a **superkey**) for R, or
  - $A$  is part of some **key** for R. ( $A$  is part of a **candidate key**)  
**(NO transitive dependencies )**





Given relation  $R(A,B,C,D)$  and the set of FDs  $F$ :

$$\begin{array}{l} ABD \rightarrow C \\ CD \rightarrow A \\ AC \rightarrow B \\ AC \rightarrow D \end{array}$$

Answer:

Is  $R$  in 3NF?

YES

NO

A	B	C	D
A	B	C	D
A	B	C	D
A	B	C	D
A	B	C	D

Explain why you chose the answer:

YES! \_\_\_\_\_  
There is NO BCNF violation! \_\_\_\_\_

- ❖ Relation  $R$  with FDs  $F$  is in 3NF if, for all  $X \rightarrow A$  in  $F$ 
  - $A \in X$  (called a *trivial FD*), or
  - $X$  is a **key** for  $R$ , or
  - $A$  is part of some **key** for  $R$ .  $A$  is part of a candidate key)

# *Third **N**ormal **F**orm (3**NF**)*



If R is in **BCNF**, obviously is in  
**3NF**.

Given relation  $R(A,B,C)$  and the set of FDs  $F$ :



Answer:

$$\begin{array}{l} AB \rightarrow C \\ C \rightarrow B \end{array}$$

Is R in 3NF?

Explain Your answer.

Candidate keys are  $AB$  and  $AC$ .

YES!

$$\begin{array}{l} \{AB\}^+ \rightarrow C \\ \{C\}^+ \rightarrow B \end{array}$$

A B C  
B C

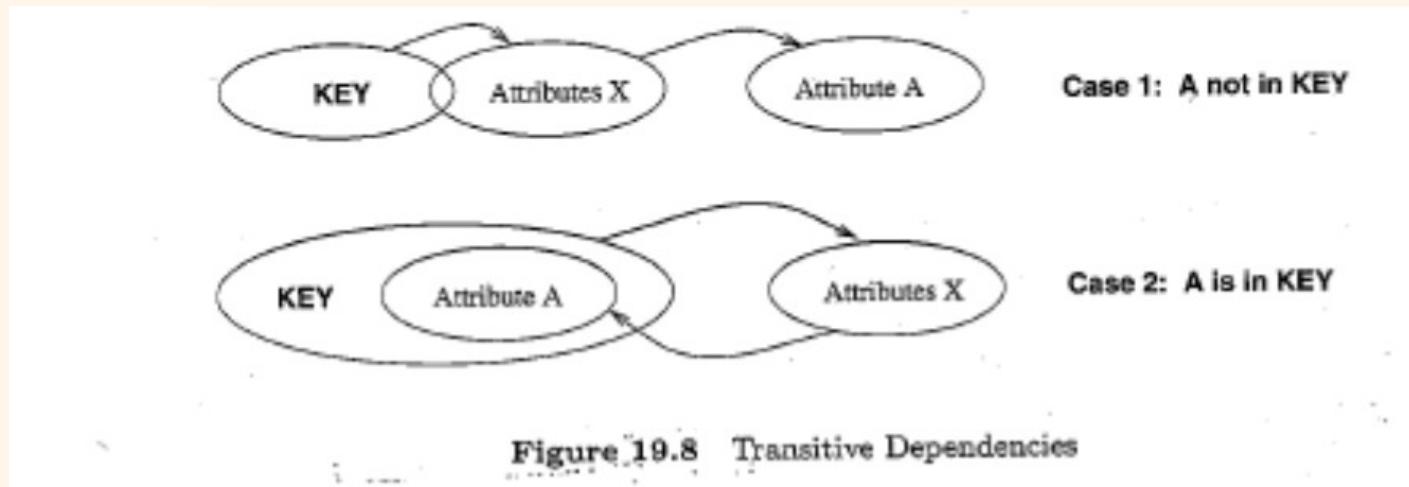
$B$  is part of  $AB$  candidate key

- ❖ Relation  $R$  with FDs  $F$  is in 3NF if, for all  $X \rightarrow A$  in  $F$ 
  - $A \in X$  (called a *trivial FD*), or
  - $X$  is a key for  $R$ , or
  - $A$  is part of some key for  $R$ . ( $A$  is part of a candidate key)

# **3NF** - Third Normal Form



- ❖ In **3NF** transitive dependencies are not allowed.



# (CA 6.2.1 – Step 2 – Normal form)

CHALLENGE  
ACTIVITY

6.2.1: Normal form.



	A	B	C	D
--	---	---	---	---



A table has:

- columns A, B, C, D

**B** not a KEY – not BCNF

**C** part of key **ACD**

What is the normal form of the table?

Ex: If a table is in first and second but not third normal form, select Second.

Pick ▼

**3NF**

# TA time (Alvaro) – 4 minutes

## (CA 6.2.1 – Step 2 – Normal form)

CHALLENGE  
ACTIVITY

6.2.1: Normal form.



A table has:

- columns A, B, C, D
- candidate keys A and (B, C)
- primary key (B, C)
- dependency  $D \rightarrow C$  (in addition to dependencies on candidate keys)

What is the normal form of the table?

Ex: If a table is in first and second but not third normal form, select Second.

Pick

Pick

Boyce-Codd

First

Second

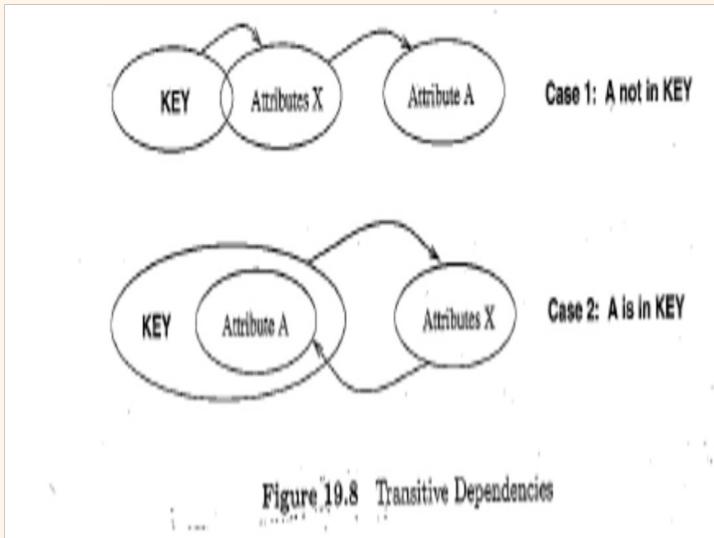
Third

2



# 3NF - Third Normal Form

- ❖ In 3NF transitive dependencies are not allowed.



## Third normal form

Redundancy can occur in a second normal form table when a non-key column depends on another non-key column. Informally, a table is in third normal form when all non-key columns depend on the key, the whole key, and nothing but the key.

PARTICIPATION  
ACTIVITY

6.1.8: Third normal form eliminates most redundancy.



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

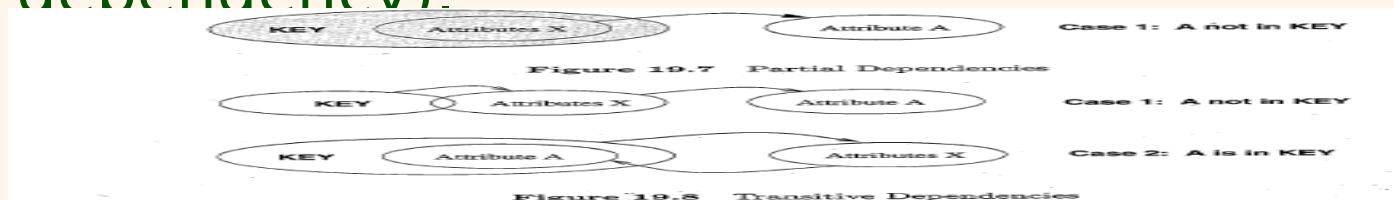
3NF?

No

# Second Normal Form (2NF)

Given **relation** R with FDs (1)  $C \rightarrow D$  (2)  $C \rightarrow A$  (3)  $B \rightarrow C$

- ❖ **Relation** R with FDs F is in 2NF if, for all X in  $F \in$ 
  - X is a key for R, or
  - X contains a key for R (X is a superkey), or
  - A is part of some key (A is part of a **candidate key**) for R, or
  - X is **NOT** a proper subset of the key (no partial dependency).



# **2NF** - Second Normal Form



- ❖ In **2NF** partial dependencies are not allowed.
- ❖ If a relation is in **3NF** (which precludes both **partial** and **transitive dependencies**), it is also in **2NF**.



Figure 19.7 Partial Dependencies



Case 1: A not in KEY

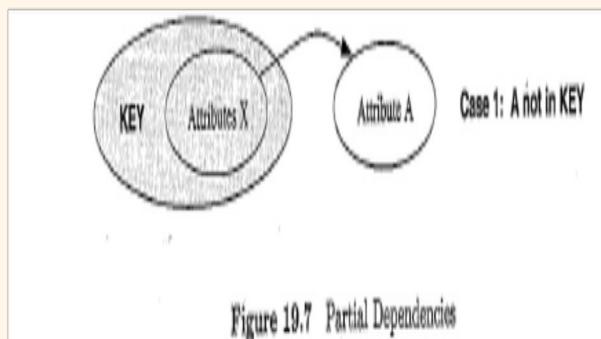


Case 2: A is in KEY

Figure 19.8 Transitive Dependencies

# 2NF - Second Normal Form

- ❖ In 2NF partial dependencies are not allowed.
- ❖ If a relation is in 3NF (which precludes both partial and transitive dependencies), it is also in 2NF.



## Second normal form

A table is in **second normal form** when all non-key columns depend on the whole primary key. In other words, a non-key column cannot depend on part of a composite primary key. A table with a simple primary key is automatically in second normal form.

PARTICIPATION  
ACTIVITY

6.1.6: Second normal form eliminates some redundancy.

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 Hatcher	GC848	First	1

2NF?

No

# Decomposition

## Relation

Hourly\_Eml

S	N	L	R	W	H
123-22-3666	Attishoo	48	8	10	40
231-31-5368	Smiley	22	8	10	30
131-24-3650	Smethurst	35	5	7	30
434-26-3751	Guldu	35	5	7	32
612-67-4134	Madayan	35	8	10	40

- ❖ Problems due to  $\text{R} \rightarrow \text{W}$  (**NOT 3NF**)

Will 2 smaller **Relations** be better?

# Decomposition



Hourly\_Emps (ssn, name, lot, rating, hrly\_wages, hrs\_worked)

Hourly\_Emps2 (ssn, name, lot, rating, hrs\_worked)

Hourly\_Emps2

S <u>R</u>	N	L	R	H
123-22-3666	Attishoo	48	8	40
231-31-5368	Smiley	22	8	30
131-24-3650	Smethurst	35	5	30
434-26-3751	Guldu	35	5	32
612-67-4134	Madayan	35	8	40

Wages

<u>R</u>	W
8	10
5	7

Wages (rating, hrly\_wages)

Will 2 smaller Relations be better?



# Decomposition

A **Decomposition** of a **Relation Schema**  $R$  consists of replacing the **Relation Schema** by two (or more) **Relation Schemas** that each contain a subset of the attributes of  $R$  **and** together include all attributes of  $R$ .

Intuitively, we want to store the information in any given **instance** of  $R$  by storing projections ( $\pi$ ) of the **instance**.

**TA, Alvaro Urtaza (A – L).**

**TA, Fernando Rodriguez (M – Z).**

**Please compare CANVAS vs. TEAMS Attendance.  
Print screens of students in CANVAS but not in the TEAMS meeting.  
(1.31.2024 Attendance X missing LastName.docx)**

## 6. SET 1 - 5:NORMALIZATION

---



  6.1 First, second, and third normal form 

  6.2 Boyce-Codd normal form 

  6.3 Applying normal form

  6.4 LAB - Implement strong entity (Sakila) 

  6.5 LAB - Implement supertype and subtype entities (Sakila) 

## Structural rules

Tables obey three structural rules:

1. Tables are **normalized** – exactly one value exists in each cell.
2. No duplicate column names – duplicate column names are not allowed in one table. However, the same column name can appear in different tables.
3. No duplicate rows – no two rows may have identical values in all columns.

In addition to the three rules, relational databases obey the principle of **data independence**, which states that rows and columns of a table have no inherent order. Although values in rows and columns are stored sequentially on a storage device, such as a disk drive, the sequence is arbitrary and does not affect the results of a database query. Data independence allows database administrators to tune the database for fast insertion and retrieval of data without affecting database query results.

PARTICIPATION  
ACTIVITY

4.2.5: Rules governing tables.



Employee

ID	Name	Salary
2538	Lisa Ellison	45000
5384	Sam Snead	30500
6381	Maria Rodriguez	92300 70000

Employee

ID	Name	Salary	Salary
2538	Lisa Ellison	45000	35000
5384	Sam Snead	30500	25600
6381	Maria Rodriguez	92300	88100

Employee

ID	Name	Salary
2538	Lisa Ellison	45000
5384	Sam Snead	30500
5384	Sam Snead	30500



1. A table cannot have multiple values in the same cell.
2. A table cannot have multiple columns with the same name.
3. A table cannot have multiple rows with identical data.

PARTICIPATION  
ACTIVITY

4.2.6: Rules governing tables.

1) How many values may appear in each cell?

- Any number of values
- Exactly one value
- No values or one value

Correct

Tables must be normalized, so only one value may appear in each cell.

?????

# 6.1 Normal form

## Functional Dependency ( $X \rightarrow Y$ )

### Redundancy and dependence

**Redundancy** is the repetition of related values. If a value in one column is updated or deleted, all copies must be changed. If one copy is changed, other copies become inconsistent and the correct value is lost.

Column A **depends on** column B means each value in A depends on B. 'B  $\rightarrow$  A'

#### Relation

**Hourly\_Emps** (ssn, name, lot, rating, hrly\_wages, hrs\_worked)

S	N	L	R	W	H
123-22-3666	Attishoo	48	8	10	40
231-31-5368	Smiley	22	8	10	30
131-24-3650	Smethurst	35	5	7	30
434-26-3751	Guldu	35	5	7	32
612-67-4134	Madayan	35	8	10	40

In related values are updated uniformly, the example or composite. 'A

$R \rightarrow W$

PARTICIPATION  
ACTIVITY

6.1.1: Examples of redundancy and 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



# 6.1 Normal form

## Redundancy and dependence

**Redundancy** is the repetition of related values in a table. Redundancy causes database management problems. When related values are updated or deleted, all copies must be changed, which makes queries slow and complex. If copies are not updated or deleted uniformly, the copies become inconsistent and the correct version is uncertain.

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

### PARTICIPATION ACTIVITY

#### 6.1.2: Redundancy and dependence.

Column names refer to the Booking table in the animation above.

1) Redundancy is the repetition of a value.

True

False

Correct

Redundancy is the repetition of related values, such as (222, Elvira Yin).

?????

## Normal forms

**Normal forms** are rules for designing tables with less redundancy. Normal forms are numbered, first normal form through fifth normal form. An additional normal form, named Boyce-Codd normal form, is an improved version of third normal form. The six normal forms comprise a sequence, with each normal form allowing less redundancy than the prior normal form.

As a practical matter, Boyce-Codd normal form is most important. Boyce-Codd normal form is unique, and all associated redundancy are not unique, and all associated redundancy.

Dependence of one column on another is formally called **functional dependence**. Functional dependence. **Multivalued dependence** and **join dependence** entail dependencies between three or more columns. Multivalued and join dependencies are complex, uncommon, and primarily of theoretical interest.

Fourth and fifth normal forms eliminate multivalued and join dependencies, respectively, and associated redundancy. Fourth and fifth normal forms are not discussed in this material.

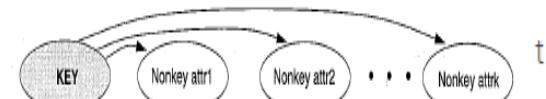
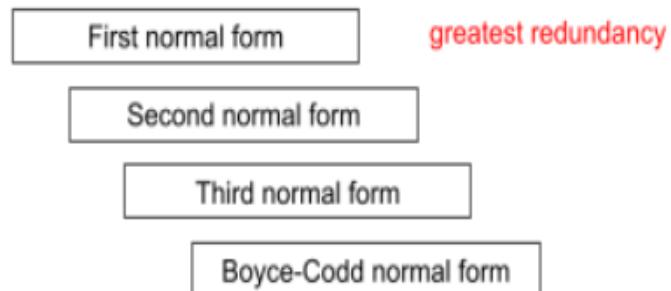


Figure 19.5 FDs in a BCNF Relation

Figure 6.1.1: Normal forms are successively more restrictive.



## First normal form

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

If a table has no duplicate rows, the composite of all columns is unique, so either all columns or some subset comprise a primary key. A table with a primary key is in first normal form, as illustrated in the animation below. Thus, a table with no duplicate rows is in first normal form.

PARTICIPATION  
ACTIVITY

6.1.4: Tables with a primary key are in first normal form.



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

First normal form

## Second normal form

## Decomposition



A table is in **second normal form** when all non-key columns depend on part of a composite primary key. A table with a simple primary key is in first normal form.

PARTICIPATION  
ACTIVITY

6.1.6: Second normal form eliminates some redundancy

Hourly\_Emps2 (*ssn, name, lot, rating,*  
*hrs\_worked*)

Hourly\_Emps2

S	N	L	R	H
123-22-3666	Attishoo	48	8	40
231-31-5368	Smiley	22	8	30
131-24-3650	Smethurst	35	5	30
434-26-3751	Guldu	35	5	32
612-67-4134	Madayan	35	8	40

Wages

R	W
8	10
5	7

Wages (*rating, hrly\_wages*)

Will 2 smaller Relations be better?

X → A

relation1

XA

relation2

R - A

41

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

First normal form



## Second normal form

A table is in **second normal form** when all non-key columns depend on the whole primary key. In other words, a non-key column cannot depend on part of a composite primary key. A table with a simple primary key is automatically in second normal form.

### PARTICIPATION ACTIVITY

#### 6.1.7: Second normal form.

Refer to the StudentRecord table, listing all courses taken by students.

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@sbc.net	F

1) Which fact is repeated?

- (8034, john@gmail.com)
- (2111, Math 100)
- (95, A)

Correct

The fact that StudentNumber 8034 has email address john@gmail.com appears twice.

?????

R	W
8	10
5	7

## Third normal form

Wages (rating, hrly\_wages)

Will 2 smaller Relations be better?

$X \rightarrow A$  relation1

$XA$  relation2  $R - A$

Redundancy can occur in a second normal form table when a non-key column depends on another non-key column. Informally, a table is in third normal form when all non-key columns depend on the key, the whole key, and nothing but the key.

PARTICIPATION  
ACTIVITY

6.1.8: Third normal form eliminates most redundancy.

X A

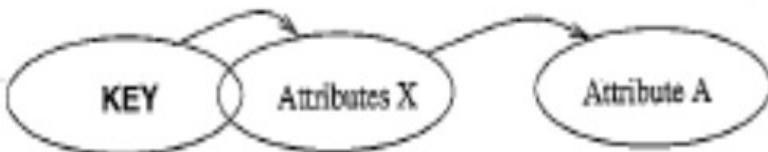
Booking

● 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

R - A

X A



Case 1: A not in KEY

Figure 19.8 Transitive Dependencies

# Third normal form



Redundancy can occur in a second normal form table when a non-key column depends on another non-key column. Informally, a table is in third normal form when all non-key columns depend on the key, the whole key, and nothing but the key.

## PARTICIPATION ACTIVITY

### 6.1.9: Third normal form.

Refer to the following table. Each department has one chair, but the same person occasionally chairs several departments. Course names can be repeated in different departments, but never within the same department.

Course

CourseCode	DepartmentCode	CourseName	DepartmentChair
8451	MATH	Discrete Mathematics	Azim Rafiq
8452	CS	Discrete Mathematics	Je Soomin
2391	SPAN	Introduction to Spanish	Susan Williams
5505	BIO	Genetics and Evolution	Susan Williams
8449	MATH	Calculus I	Azim Rafiq
8036	MATH	Differential Equations	Azim Rafiq

1) Which fact is repeated?

- Discrete Mathematics
- (MATH, Differential Equations)
- (MATH, Azim Rafiq)

Correct

The fact that Azim Rafiq is chair of the Math department is repeated for every math course in the table.

?????

Wages	
R	W
8	10
5	7

## Boyce-Codd normal form

The definition of third normal form applies to non-key columns only, which allows for occasional redundancy. Boyce-Codd normal form applies to all columns and eliminates this redundancy.

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. This definition is identical to the definition of third normal form with the term 'non-key' removed.

Boyce-Codd normal form is considered the gold standard of table design. Although fourth and fifth normal forms remove additional types of redundancy, these redundancies are uncommon and of little practical concern.

PARTICIPATION  
ACTIVITY

6.1.10: Boyce-Codd normal form eliminates all common redundancy.



A      X

EmployeeProjectTask

EmployeeID	ProjectName	TaskName
489	Performance reviews	Review Jino 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

R - A

X

X

A

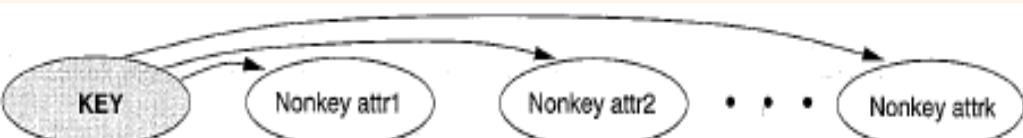


Figure 19.5 FDs in a BCNF Relation

## Boyce-Codd normal form



The definition of third normal form applies to non-key columns only, which allows for occasional redundancy. Boyce-Codd normal form applies to all columns and eliminates this redundancy.

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. This definition is identical to the definition of third normal form with the term 'non-key' removed.

Boyce-Codd normal form is considered the gold standard of table design. Although fourth and fifth normal forms remove additional types of redundancy, these redundancies are uncommon and of little practical concern.

### PARTICIPATION ACTIVITY

#### 6.1.11: Boyce-Codd normal form.

Refer to the following table. Assume all licenses and postal codes are in the United States.

Employee

•EmployeeID	DriversLicenseNumber	DriversLicenseState	Name	PostalCode
489	AB7325	IL	Lisa Ellison	60415
517	N3259211	CA	Sam Snead	90295
600	B16629045	CA	Malia Efrenza	90295
777	8242103	TX	Nadia Shah	75185
929	8242103	FL	Maria Rodriguez	32099
933	AX493200	CA	Jiho Chen	94701

- 1) EmployeeID is the only candidate key.

- True
- False

Correct

(DriversLicenseNumber, DriversLicenseState) is unique and minimal, and therefore also a candidate key.

?????

# 6.2 Normalization

## Normalization

Implementing entities, relationships, and attributes usually generates tables with no redundancy. Occasionally, however, implementation results in redundant tables. This redundancy is eliminated with normalization, the last step of logical design.

?????

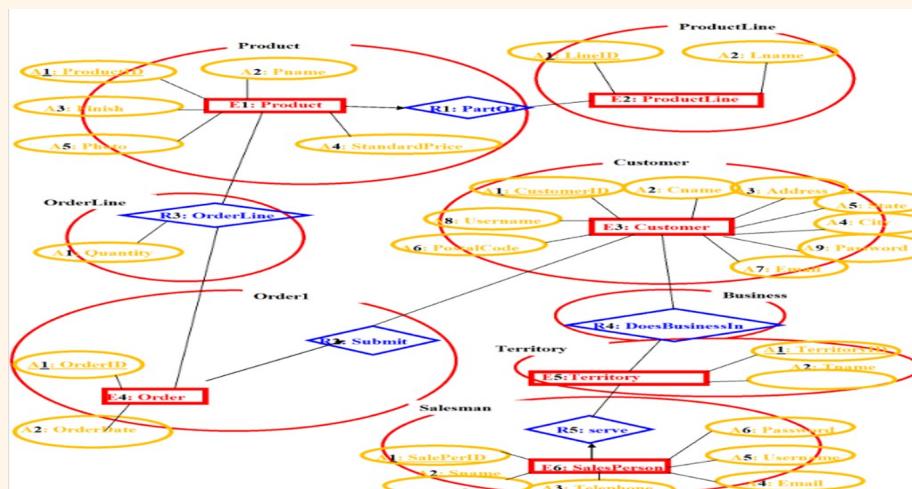
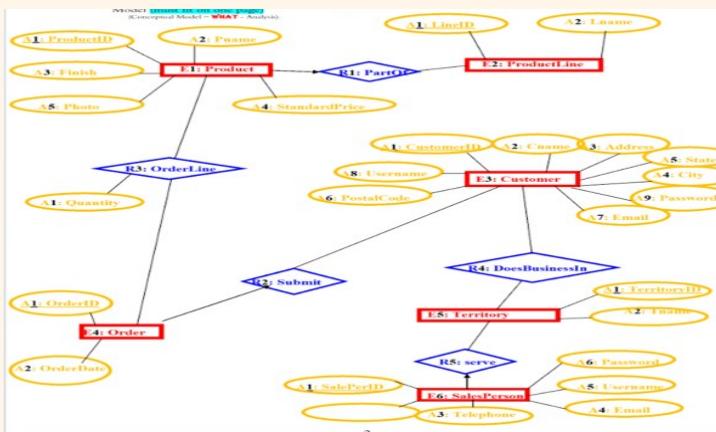


TABLE format

```
Create TABLE Customer (CustomerID CHAR(20), Cname CHAR(20), Address CHAR(20), State CHAR(20), City CHAR(15), PostalCode CHAR(15), Email CHAR(20), Username CHAR(20), Password CHAR(20), PRIMARY KEY (CustomerID))
```

```
Create TABLE ProductLine (LineID CHAR(30), Lname CHAR(30), PRIMARY KEY (LineID))
```

```
Create TABLE Order (OrderID CHAR(20), orderDate DATE, CustomerID CHAR(20), PRIMARY KEY (OrderID), FOREIGN KEY (CustomerID) REFERENCES Customer)
```

```
Create TABLE Product (ProducedID CHAR(20), Pname CHAR(20), Finish CHAR(20), IMAGE IMAGE, LineID CHAR(20), PRIMARY KEY (ProducedID), FOREIGN KEY (LineID) REFERENCES ProductLine)
```

```
Create TABLE OrderLine (Quantity INTEGER, SalesPrice FLOAT, ProductID CHAR(20), OrderID CHAR(20), PRIMARY KEY (ProductID, OrderID), FOREIGN KEY (ProductID) REFERENCES Product, FOREIGN KEY (OrderID) REFERENCES Order)
```

## 6.2 Normalization

**Normalization** eliminates redundancy by decomposing a table into two or more tables in higher normal form. Ex: A table in first normal form might be replaced by two tables in third normal form. In principle, normalization decomposes tables to any higher normal form. Fourth and fifth normal form are complex, however, and have limited practical value. As a practical matter, database designers usually normalize tables to Boyce-Codd normal form.

Column A **depends on** column B when each B value is related to at most one A value. A and B may be simple or composite columns. In a **Boyce-Codd normal form** table, if column A depends on column B, then B must be unique. Normalizing a table to Boyce-Codd normal form involves three steps:

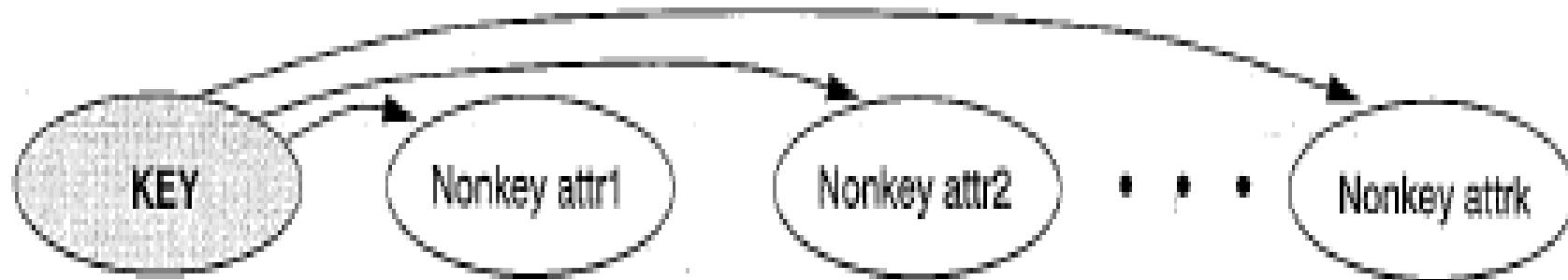


Figure 19.5 FDs in a BCNF Relation

Normalization eliminates redundancy by removing A from the original table. Since the data relating A and B is recorded in a new table, no information is lost.

# (CA 6.2.1 – Step 3 – Decomposition)

CHALLENGE  
ACTIVITY

6.2.1: Normal form.

	A	B	C	D
--	---	---	---	---

B A B C D

**AD** → A B C D

Table Old has:

- columns A, B, C, D

B a KEY

**AD** a KEY

BCNF

3

Table Old already BCNF; **?????**  
**No decomposition needed**

# TA time (Alvaro) – 4 minutes

## (CA 6.2.1 – Step 3 – Decomposition)

### CHALLENGE ACTIVITY

6.2.1: Normal form.



Table Old has:

- columns A, B, C, D
- candidate keys B and (C, D)
- dependency  $C \rightarrow A$  (in addition to dependencies on candidate keys)

Revise table Old and create table New to achieve Boyce-Codd normal form.

What columns are in the revised table Old, table New, and the table New primary key?

A	B	C	D	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	table Old
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	table New
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	table New primary key



ta

hrke

0

## Denormalization

Boyce-Codd normal form is ideal for tables with frequent inserts, updates, and deletes. In a database used primarily for reporting, changes are infrequent and redundancy is acceptable. In fact, redundancy can be desirable in reporting databases, as processing is faster and queries are simpler. Therefore, reporting databases may contain tables that, by design, are not in third normal form.

**Denormalization** means intentionally introducing redundancy by merging tables. Denormalization eliminates join queries and therefore improves query performance. Denormalization results in first and second normal form tables and should be applied selectively and cautiously.

In the figure below, the Booking, Passenger, and Fare tables are denormalized into a single Booking table. The red highlight indicates redundancy in the denormalized table.

Figure 6.2.1: Denormalization example.

The diagram illustrates the denormalization process. It starts with three separate tables: Booking, Passenger, and Fare. An arrow points down to a single merged Booking table, where certain fields from the original tables are highlighted in red to indicate redundancy.

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

Passenger	
● PassengerNumber	PassengerName
222	Elvira Yin
333	Deepak Chopra
444	Mary Hutcher

● FareClass	BoardingZoneNumber
First	1
Economy	3
Business	2

↓

Booking				
● PassengerNumber	PassengerName	● FlightCode	FareClass	BoardingZoneNumber
222	Elvira Yin	AZ312	First	1
222	Elvira Yin	BF999	Economy	3
222	Elvira Yin	GC848	Business	2
333	Deepak Chopra	GC848	First	1
444	Mary Hutcher	GC848	First	1

## Denormalization



Boyce-Codd normal form is ideal for tables with frequent inserts, updates, and deletes. In a database used primarily for reporting, changes are infrequent and redundancy is acceptable. In fact, redundancy can be desirable in reporting databases, as processing is faster and queries are simpler. Therefore, reporting databases may contain tables that, by design, are not in third normal form.

**Denormalization** means intentionally introducing redundancy by merging tables. Denormalization eliminates join queries and therefore improves query performance. Denormalization results in first and second normal form tables and should be applied selectively and cautiously.

In the figure below, the Booking, Passenger, and Fare tables are denormalized into a single Booking table. The red highlight indicates redundancy in the denormalized table.

### PARTICIPATION ACTIVITY

#### 6.2.3: Denormalization.

- 1) Denormalization never results in second-normal-form tables.

- True
- False

Correct

Denormalization can introduce a dependency of a column on another non-key column, resulting in a second-normal-form table that is no longer in third normal form.

?????

# From 5:00 to 5:05 PM – 5 minutes

01.31.2024 | ZyBook SET 1 - 5 | Set 1  
(5 - We) | LECTURE 5 NORMALIZATION

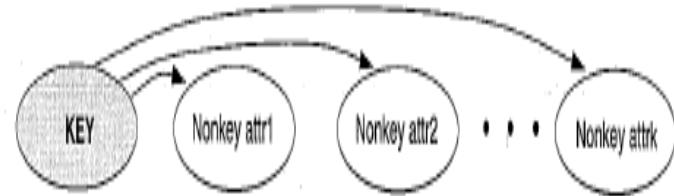


Figure 19.5 FDs in a BCNF Relation

- 1. SET 1 Empty ▾
- 6. SET 1 - 5:NORMALIZATION 0% 0% ▾

Please work on  
SET 1 – 6: Normalization

# SET 1 Lecture 5

## NORMALIZATION

### 6. SET 1 - 5:NORMALIZATION



6.1 First, second, and third normal form New Hidden

6.2 Boyce-Codd normal form New Hidden

6.3 Applying normal form Hidden

6.4 LAB - Implement strong entity (Sakila) Hidden Lab

6.5 LAB - Implement supertype and subtype entities (Sakila) Hidden Lab

# Next

02.05.2024

(6 · Mo)

EXAM 1 Practice

(PART of 20 points)

## Activity for entire class

Entire zyBook



Entire class

## Time spent (hours:minutes:seconds)

04:00:00

03:20:00

02:40:00

02:00:00

01:20:00

00:40:00

00:00:00



## AT HOME LOCKDOWN BROWSER & MONITORING TESTING - Requires Respondus LockDown Browser + Webcam



All Sections

After installing the UH Lockdown Browser, please take it opened till Jan. 31, 2024 NOON.

...

Posted on:

Jan 29, 2024, 10:42 PM

## We will use UH LOCKDOWN BROWSER for EXAM REVIEWS and EXAMS



All Sections

University of Houston Respondus Student Guide: Student GuideLinks to an external site. Univer...

...

Posted on:

Jan 29, 2024, 8:55 PM



COSC3380 23578 - Databa... > Quizzes > AT HOME LOCKDOWN BR...

H 2024 Spring 1

Home

Announcements

Assignments

Discussions

Grades

People

Pages

Files

Syllabus

Outcomes



Rubrics

Quizzes

Modules

Collaborations

New Analytics

Studio

LockDown  
Browser

Course Reserves

Attendance

Top Hat

BryteWave Course  
Materials

...

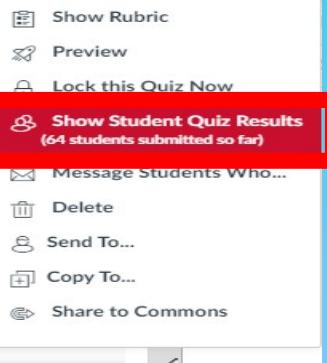
Published

Preview

Edit

...

Rele...



## AT HOME LOCKDOWN BROWSER & MONITORING TESTING - Requires Respondus LockDown Browser Webcam

### AT HOME LOCKDOWN BROWSER & MONITORING TESTING

#### LockDown Browser Settings

- Don't require Respondus LockDown Browser for this exam
- Require Respondus LockDown Browser for this exam

#### >Password Settings

Access code for this exam (optional)

Quiz Type Graded Quiz

Points 100

Assignment Group CLASS PARTICIPATION [20 points]

Shuffle Answers Yes

Time Limit 10 Minutes

Multiple Attempts No

View Responses Always

Show Correct Answers From Jan 29 at 10:30pm to Jan 31 at 12pm

One Question at a Time Yes

Require Respondus LockDown  
Browser

Required to View Quiz Results Yes

Webcam Required Yes

Lock Questions After Answering Yes

02.07.2024

(7 - We)

**TA Download**

ZyBook SET 1 Sections

**(4 PM)**

**(PART of 30 points)**

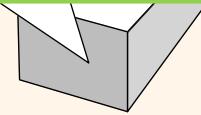
**EXAM 1 Review**

**(PART of 20 points)**

EXAM 1 REVIEW (CANVAS) (100 points) - Requires Respondus LockDown Browser + Webcam

Uses UH Lockdown Browser;  
Web Camera required

# From 5:05 to 5:15 PM – 5 minutes.



01.31.2024  
(5 - We)

## ZyBook SET 1 - 5

### Set 1

#### LECTURE 5 NORMALIZATION



CLASS PARTICIPATION 20 points

20% of Total + :

## CLASS 5

### Class 5 END PARTICIPATION

Not available until Jan 31 at 5:05pm | Due Jan 31 at 5:15pm | 100 pts

VH, publish



This is a synchronous online class.

Attendance is required.

Recording or distribution of class materials is prohibited.

1. At the beginning of selected classes there is an assessment in the first 10 minutes. (beige BOX in the Detailed Syllabus)

2. At the end of selected classes there is an assessment in the last 10 minutes. (blue BOX in the Detailed Syllabus)

3. ZyBook sections will be downloaded and used for 30% of Total Score on the dates specified in the Detailed Syllabus.

4. EXAMS are in CANVAS. No late EXAMS.

5. I have to be present in TEAMS in order to take any graded assignment assigned during that class.

**At 5:15 PM.**

---

**End Class 5**

**VH, Download Attendance Report  
Rename it:  
1.31.2024 Attendance Report FINAL**

**VH, upload **Class 5** to CANVAS.**