

# CS 327E Lecture 5

Shirley Cohen

September 14, 2016

# Plan for Today

- Finish Normalization
- Reading Quiz (based on Chapter 2 of our SQL book)
- Lab 1 Requirements
- Git and Github Demo
- Mini Setup Session for Lab 1

# Functional Dependencies

Definition:

If two records agree on the attributes

$$A_1, A_2, \dots, A_n$$

then they must also agree on the attributes

$$B_1, B_2, \dots, B_n$$

Formally:

$$A_1, A_2, \dots, A_n \rightarrow B_1, B_2, \dots, B_n$$

# 1NF

## *Drug*

<u>drug_nbr</u>	<u>drug_name</u>	<u>drug_qty</u>	<u>start_date</u>	<u>end_date</u>	<u>drug_price</u>
48	Amoxicillin	500	01/01/13	03/31/15	0.30
48	Amoxicillin	500	04/01/15	01/15/16	3.00
48	Amoxicillin	500	01/16/16		3.50
50	Lipitor	150	10/01/12	03/31/14	0.75
50	Lipitor	150	04/01/14		1.00
72	Singulair	250	01/01/15	05/31/15	0.20
72	Singulair	250	06/01/15	07/31/15	0.80
72	Singulair	250	08/01/15		0.20

### **FDs:**

1. drug\_nbr → drug\_name, drug\_qty
2. drug\_nbr, start\_date → end\_date, drug\_price

# 1NF to 2NF

## Drug

<u>drug_nbr</u>	drug_name	drug_qty	<u>start_date</u>	end_date	drug_price
48	Amoxicillin	500	01/01/13	03/31/15	0.30
48	Amoxicillin	500	04/01/15	01/15/16	3.00
48	Amoxicillin	500	01/16/16		3.50
...	...	...	...	...	...

1NF

**FDs:**

drug\_nbr → drug\_name, drug\_qty

drug\_nbr, start\_date → end\_date, drug\_price

## Drug'

<u>drug_nbr</u>	drug_name	drug_qty
48	Amoxicillin	500
50	Lipitor	150
72	Singulair	250

2NF

## Price

<u>drug_nbr</u>	<u>start_date</u>	end_date	drug_price
48	01/01/13	03/31/15	0.30
48	04/01/15	01/15/16	3.00
48	01/16/16		3.50

2NF

# 1NF to 2NF

Rule: A database schema is in 2NF *iff* it is in 1NF and there exists no partial FDs on the primary key (i.e. all non-key attributes must be dependent on the entire PK)

**Student Semester**

<u>EID</u>	<u>Semester</u>	<u>Course</u>	<u>Grade</u>	<u>GPA</u>
alice1	Fall15	Stats	A	3.9
alice1	Fall15	DB	A	3.9
alice1	Fall15	Alg	A-	3.9
bob20	Fall15	DB	A	3.7
bob20	Fall15	Alg	B+	3.7
carol30	Fall15	Stats	A-	3.5
carol30	Fall15	Alg	B+	3.5

1NF

**Student Semester Grade**

<u>EID</u>	<u>Semester</u>	<u>Course</u>	<u>Grade</u>
alice1	Fall15	Stats	A
alice1	Fall15	DB	A
alice1	Fall15	Alg	A-
bob20	Fall15	DB	A
bob20	Fall15	Alg	B+
carol30	Fall15	Stats	A-
carol30	Fall15	Alg	B+

2NF

**FDs:**

1. EID, Semester, Course  $\rightarrow$  Grade
2. EID, Semester  $\rightarrow$  GPA

**Student Semester GPA**

<u>EID</u>	<u>Semester</u>	<u>GPA</u>
alice1	Fall15	3.9
bob20	Fall15	3.7
carol30	Fall15	3.5

2NF

# 2NF to 3NF

Rule: A database schema is in 3NF *iff* it is in 2NF and there exists no non-key attributes that depend on other non-key attributes.

## Student\_Major

<u>EID</u>	Name	Major	College
alice1	Alice	Math	Natural Sciences
bob20	Bob	CS	Natural Sciences
carol30	Carol	Math	Natural Sciences

FDs:

$EID \rightarrow Name, Major, College$   
 $Major \rightarrow College$

2NF

## Student\_Major'

<u>EID</u>	Name	Major
alice1	Alice	Math
bob20	Bob	CS
carol30	Carol	Math

3NF

## Major\_College

<u>Major</u>	College
Math	Natural Sciences
CS	Natural Sciences

3NF

# Concept Question 1

Suppose we add a *drug\_description* field to the *Drug* table as shown below and then discover that *drug\_name* → *drug\_description*. Does this change the normal form for this table? If so, which normal form is the table in now?

## Drug

<u>drug_nbr</u>	drug_name	drug_description	drug_qty
48	Amoxicillin	Treats bacterial infections	500
50	Lipitor	Reduces cholesterol levels	150
72	Singulair	Prevents asthma symptoms	250

### FDs:

*drug\_nbr* → *drug\_name*, *drug\_description*, *drug\_qty*

*drug\_name* → *drug\_description*

- A. 1NF
- B. 2NF
- C. 3NF
- D. None of the above



# Solution CQ 1

Drug

<u>drug_nbr</u>	drug_name	drug_description	drug_qty
48	Amoxicillin	Treats bacterial infections	500
50	Lipitor	Reduces cholesterol levels	150
72	Singulair	Prevents asthma symptoms	250

2NF

**FDs:**

drug\_nbr → drug\_name, drug\_description, drug\_qty  
drug\_name → drug\_description

Drug'

<u>drug_nbr</u>	drug_name	drug_qty
48	Amoxicillin	500
50	Lipitor	150
72	Singulair	250

3NF

Drug\_Description

<u>drug_name</u>	drug_description
Amoxicillin	Treats bacterial infections
Lipitor	Reduces cholesterol levels
Singulair	Prevents asthma symptoms

3NF

# Quiz Question 1

Although the text is system-agnostic, what relational database system is used in the examples of Beaulieu's *Learning SQL*?

- A. PostgreSQL
- B. MySQL
- C. Microsoft SQL Server
- D. Oracle Database

## Quiz Question 2

What MySQL data type is used to store fixed-length strings?

- A. CHAR
- B. VARCHAR
- C. STRING
- D. STR

## Quiz Question 3

Why would you choose a `TIMESTAMP` over a `DATE` type?

- A. `TIMESTAMP` is more precise than a `DATE`
- B. Only for representing the time
- C. `TIMESTAMP` is for representing a date and time (year, month, day, hour, minute, second) while `DATE` is for representing a date (year, month, day)
- D. Never—`DATE` should always be used instead of `TIMESTAMP`

# Quiz Question 4

What SQL statement would you use to create a new row in a table?

- A. APPEND
- B. NEW
- C. INSERT
- D. ALTER

# Quiz Question 5

Below is the output from executing a MySQL command:

```
mysql> ??????????????????
```

Field	Type	Null	Key	Default	Extra
branch_id	smallint(5) unsigned	NO	PRI	NULL	auto_increment
name	varchar(20)	NO		NULL	
address	varchar(30)	YES		NULL	
city	varchar(20)	YES		NULL	
state	varchar(2)	YES		NULL	
zip	varchar(12)	YES		NULL	

What is the command that was executed?

- A. SHOW CUSTOMER;
- B. DESCRIBE CUSTOMER;
- C. SELECT \* FROM CUSTOMER;
- D. UPDATE CUSTOMER;

# Lab 1 Requirements

<http://www.cs.utexas.edu/~scohen/labs/lab1.pdf>

# Next Steps

- Complete set up for Lab 1 this weekend
- Use Chapter 6 from our Wrangling text to find datasets
- Work on Lab 1 during class time next week (and outside of class time)
- Use sign-up sheet during class time next week if your team needs support
- Use Piazza or come to office hours if your team needs help outside of class time
- **Submit Lab 1 by 11:59pm next Friday**