# 7+8+9: Functional Dependencies and Normalization

how "good" is our data model design?

what do we know about the quality of the logical schema?



how do we know the database design won't cause any problems?

> grouping of attributes so far has been intuitive - how do we validate the design?

#### issue of quality of design an illustration - figure 7.1b



STOCK (Store, Product, Price, Quantity, Location, Discount, Sq\_ft, Manager)

#### are there data redundancies?

#### for price? location? quantity? discount?

#### STOCK

| Store | Product        | Price | Quantity | Location | Discount | Sq_ft | Manager |
|-------|----------------|-------|----------|----------|----------|-------|---------|
|       |                |       |          |          |          |       |         |
| 15    | Refrigerator   | 1850  | 120      | Houston  | 5%       | 2300  | Metzger |
| 15    | Dishwasher     | 600   | 150      | Houston  | 5%       | 2300  | Metzger |
| 13    | Dishwasher     | 600   | 180      | Tulsa    | 10%      | 1700  | Metzger |
| 14    | Refrigerator   | 1850  | 150      | Tulsa    | 5%       | 1900  | Schott  |
| 14    | Television     | 1400  | 280      | Tulsa    | 10%      | 1900  | Schott  |
| 14    | Humidifier     | 55    | 30       | Tulsa    |          | 1900  | Schott  |
| 17    | Television     | 1400  | 10       | Memphis  |          | 2300  | Creech  |
| 17    | Vacuum Cleaner | 300   | 150      | Memphis  | 5%       | 2300  | Creech  |
| 17    | Dishwasher     | 600   | 150      | Memphis  | 5%       | 2300  | Creech  |
| 11    | Computer       |       | 180      | Houston  | 10%      | 2300  | Creech  |
| 11    | Refrigerator   | 1850  | 120      | Houston  | 5%       | 2300  | Creech  |
| 11    | Lawn Mower     | 300   |          | Houston  |          | 2300  | Creech  |

Figure 7.1c An instance of the relation schema, STOCK

repeated appearances of a data value ≠ data redundancy

# what is data redundancy?

unneeded repetition that does not add new meaning = data redundancy

> data redundancy → modification anomalies

#### are there data redundancies?

#### yes - for price, location, and discount

#### STOCK

| Store | Product        | Price | Quantity | Location | Discount | Sq_ft | Manager |
|-------|----------------|-------|----------|----------|----------|-------|---------|
|       |                |       |          |          |          |       |         |
| 15    | Refrigerator   | 1850  | 120      | Houston  | 5%       | 2300  | Metzger |
| 15    | Dishwasher     | 600   | 150      | Houston  | 5%       | 2300  | Metzger |
| 13    | Dishwasher     | 600   | 180      | Tulsa    | 10%      | 1700  | Metzger |
| 14    | Refrigerator   | 1850  | 150      | Tulsa    | 5%       | 1900  | Schott  |
| 14    | Television     | 1400  | 280      | Tulsa    | 10%      | 1900  | Schott  |
| 14    | Humidifier     | 55    | 30       | Tulsa    |          | 1900  | Schott  |
| 17    | Television     | 1400  | 10       | Memphis  |          | 2300  | Creech  |
| 17    | Vacuum Cleaner | 300   | 150      | Memphis  | 5%       | 2300  | Creech  |
| 17    | Dishwasher     | 600   | 150      | Memphis  | 5%       | 2300  | Creech  |
| 11    | Computer       |       | 180      | Houston  | 10%      | 2300  | Creech  |
| 11    | Refrigerator   | 1850  | 120      | Houston  | 5%       | 2300  | Creech  |
| 11    | Lawn Mower     | 300   |          | Houston  |          | 2300  | Creech  |

Figure 7.1c An instance of the relation schema, STOCK

#### insertion anomaly - adding a washing machine with a price

(cannot add without store number)

#### update anomaly - store 11 moves to Cincinnati

(requires updating multiple rows)

#### deletion anomaly - store 17 is closed

(pricing data about vacuum cleaner is lost)

#### STOCK

| Store | Product        | Price | Quantity | Location | Discount | Sq_ft | Manager |
|-------|----------------|-------|----------|----------|----------|-------|---------|
|       |                |       |          |          |          |       |         |
| 15    | Refrigerator   | 1850  | 120      | Houston  | 5%       | 2300  | Metzger |
| 15    | Dishwasher     | 600   | 150      | Houston  | 5%       | 2300  | Metzger |
| 13    | Dishwasher     | 600   | 180      | Tulsa    | 10%      | 1700  | Metzger |
| 14    | Refrigerator   | 1850  | 150      | Tulsa    | 5%       | 1900  | Schott  |
| 14    | Television     | 1400  | 280      | Tulsa    | 10%      | 1900  | Schott  |
| 14    | Humidifier     | 55    | 30       | Tulsa    |          | 1900  | Schott  |
| 17    | Television     | 1400  | 10       | Memphis  |          | 2300  | Creech  |
| 17    | Vacuum Cleaner | 300   | 150      | Memphis  | 5%       | 2300  | Creech  |
| 17    | Dishwasher     | 600   | 150      | Memphis  | 5%       | 2300  | Creech  |
| 11    | Computer       |       | 180      | Houston  | 10%      | 2300  | Creech  |
| 11    | Refrigerator   | 1850  | 120      | Houston  | 5%       | 2300  | Creech  |
| 11    | Lawn Mower     | 300   |          | Houston  |          | 2300  | Creech  |

#### Figure 7.1c An instance of the relation schema, STOCK

# decomposition of the STOCK instance

| C1 | ΓN | D١ | _ |
|----|----|----|---|
| 21 | U  | N  | _ |

| Store | Location | Sq_ft | Manager |
|-------|----------|-------|---------|
|       |          |       |         |
| 15    | Houston  | 2300  | Metzger |
| 13    | Tulsa    | 1700  | Metzger |
| 14    | Tulsa    | 1900  | Schott  |
| 17    | Memphis  | 2300  | Creech  |
| 11    | Houston  | 2300  | Creech  |

| Product         | Price |
|-----------------|-------|
| Refrigerator    | 1850  |
| Dishwasher      | 600   |
| Television      | 1400  |
| Humidifier      | 55    |
| Vacuum Cleaner  | 300   |
| Computer        |       |
| Lawn Mower      | 300   |
| Washing Machine | 750   |

INVENTORY

| Store | Product        | Quantity | Discount |
|-------|----------------|----------|----------|
| 15    | Refrigerator   | 120      | 5%       |
| 15    | Dishwasher     | 150      | 5%       |
| 13    | Dishwasher     | 180      | 10%      |
| 14    | Refrigerator   | 150      | 5%       |
| 14    | Television     | 280      | 10%      |
| 14    | Humidifier     | 30       |          |
| 17    | Television     | 10       |          |
| 17    | Vacuum Cleaner | 150      | 5%       |
| 17    | Dishwasher     | 150      | 5%       |
| 11    | Computer       | 180      | 10%      |
| 11    | Refrigerator   | 120      | 5%       |
| 11    | Lawn Mower     |          |          |

no data redundancies/ modification anomalies in STORE or PRODUCT

data redundancies/ modification anomalies persist in INVENTORY

Figure 7.2 A decomposition of the STOCK instance in Figure 7.1c

# normalized

#### free from data redundancies/modification anomalies

| STORE | ST | ORE |  |
|-------|----|-----|--|
|-------|----|-----|--|

| JIOKE |          |       |         |
|-------|----------|-------|---------|
| Store | Location | Sq_ft | Manager |
| 15    | Houston  | 2300  | Metzger |
| 13    | Tulsa    | 1700  | Metzger |
| 14    | Tulsa    | 1900  | Schott  |
| 17    | Memphis  | 2300  | Creech  |
| 11    | Houston  | 2300  | Creech  |

| PRO | DU | CT |
|-----|----|----|
|     |    | -  |

| Product         | Price |
|-----------------|-------|
|                 |       |
| Refrigerator    | 1850  |
| Dishwasher      | 600   |
| Television      | 1400  |
| Humidifier      | 55    |
| Vacuum Cleaner  | 300   |
| Computer        |       |
| Lawn Mower      | 300   |
| Washing Machine | 750   |

| INVENTO | RY |
|---------|----|
|---------|----|

| Store | Product        | Quantity |
|-------|----------------|----------|
| 15    | Refrigerator   | 120      |
| 15    | Dishwasher     | 150      |
| 13    | Dishwasher     | 180      |
| 14    | Refrigerator   | 150      |
| 14    | Television     | 280      |
| 14    | Humidifier     | 30       |
| 17    | Television     | 10       |
| 17    | Vacuum Cleaner | 150      |
| 17    | Dishwasher     | 150      |
| 11    | Computer       | 180      |
| 11    | Refrigerator   | 120      |
| 11    | Lawn Mower     |          |

#### DISC\_STRUCTURE

| Quantity | Discount |
|----------|----------|
| 120      | 5%       |
| 150      | 5%       |
| 180      | 10%      |
| 280      | 10%      |
| 30       |          |
| 10       |          |



# relational schema reverse-engineered

from tables



Figure 7.4a A reverse-engineered logical schema for the set of tables in Figure 7.3

# design-specific ER diagram reverse-engineered

from relational schema



Figure 7.4b Design-specific ER diagram reverse-engineered from the logical schema in Figure 7.4a

how do we systematically identify data redundancies?

how do we know how to decompose the base relation schema under investigation?

how we know that the data decomposition is correct and complete?

# the seeds of data redundancy are undesirable functional dependencies.

## functional dependencies



relationship between attributes that if we are given the value of one of the attributes we can look up the value of the other

the building block of normalization principles

# functional dependencies

an attribute A (atomic or composite) in a relation schema R functionally determines another attribute B (atomic or composite) in R if:

for a given value  $a_1$  of A there is a single, specific value  $b_1$  of B in every relation state  $r_i$  of R.



## examples of functional dependency

Store  $\rightarrow$  Location Store  $\rightarrow$  Sq\_ft Store  $\rightarrow$  Manager Product  $\rightarrow$  Price {Store, Product}  $\rightarrow$  Quantity Quantity  $\rightarrow$  Discount

#### STOCK

| Store | Product        | Price | Quantity | Location | Discount | Sq_ft | Manager |
|-------|----------------|-------|----------|----------|----------|-------|---------|
|       |                |       |          |          |          |       |         |
| 15    | Refrigerator   | 1850  | 120      | Houston  | 5%       | 2300  | Metzger |
| 15    | Dishwasher     | 600   | 150      | Houston  | 5%       | 2300  | Metzger |
| 13    | Dishwasher     | 600   | 180      | Tulsa    | 10%      | 1700  | Metzger |
| 14    | Refrigerator   | 1850  | 150      | Tulsa    | 5%       | 1900  | Schott  |
| 14    | Television     | 1400  | 280      | Tulsa    | 10%      | 1900  | Schott  |
| 14    | Humidifier     | 55    | 30       | Tulsa    |          | 1900  | Schott  |
| 17    | Television     | 1400  | 10       | Memphis  |          | 2300  | Creech  |
| 17    | Vacuum Cleaner | 300   | 150      | Memphis  | 5%       | 2300  | Creech  |
| 17    | Dishwasher     | 600   | 150      | Memphis  | 5%       | 2300  | Creech  |
| 11    | Computer       |       | 180      | Houston  | 10%      | 2300  | Creech  |
| 11    | Refrigerator   | 1850  | 120      | Houston  | 5%       | 2300  | Creech  |
| 11    | Lawn Mower     | 300   |          | Houston  |          | 2300  | Creech  |

Figure 7.1c An instance of the relation schema, STOCK

Store  $\rightarrow$  LocationProduct  $\rightarrow$  PriceStore  $\rightarrow$  Sq\_ft{Store, Product}  $\rightarrow$  QuantityStore  $\rightarrow$  ManagerQuantity  $\rightarrow$  Discount

all tuples have the same price value for any given product value (not necessarily the other way around)

# a FD in R is undesirable when the determinant is not a candidate key of R.

- a candidate key is a superkey with no proper subset that uniquely identifies a tuple of a relation.
- {uniqueness property + irreducibility}

# normalization

systematic validation of participation of attributes in a relation schema from a data redundancy perspective

#### normal forms (NFs) stepwise progression towards attaining the goal of a fully normalized relation schema

# first normal form



no multi-valued attributes or composite attributes

ensuring that if the base relation were stored as a file it does not contain records with a variable number of fields

# first normal form



# by definition, a relation schema is in 1NF

normal forms associated with functional dependencies are second normal form (2NF), third normal form (3NF), and boyce-codd normal form (BCNF)

# 1NF violation



#### ALBUM

| Album_no | Artist_nm                            | Price | Stock |
|----------|--------------------------------------|-------|-------|
| DC122    | Britnov Spears                       | 17.05 | 1000  |
| B2123    | Britney spears                       | 17.95 | 1000  |
| JT111    | Justin Timberlake                    | 17.95 | 1200  |
| BTL007   | {John Lennon, Paul McCartney, George | 23.95 |       |
|          | Harrison, Ringo Star}                |       |       |
| MJ100    | Michael Jackson                      | 17.95 |       |
| JM456    | John Mayer                           | 16.95 | 1000  |
| JM151    | John Mayer                           | 16.95 | 1000  |
| MX789    | Madonna                              | 11.95 | 500   |
| DJM237   | {John Denver, Michael Jackson,       | 11.95 | 2000  |
|          | Madonna}                             |       |       |
| DR711    | Diana Ross                           | 12.95 | 1000  |
| PM137    | Paul McCartney                       | 19.95 |       |



#### NEW\_ALBUM

| <u>Album_no</u> | <u>Artist_nm</u>  | Price | Stock |
|-----------------|-------------------|-------|-------|
| BS123           | Britney Spears    | 17.95 | 1000  |
| JT111           | Justin Timberlake | 17.95 | 1200  |
| BTL007          | John Lennon       | 23.95 |       |
| BTL007          | Paul McCartney    | 23.95 |       |
| BTL007          | George Harrison   | 23.95 |       |
| BTL007          | Ringo Star        | 23.95 |       |
| MJ100           | Michael Jackson   | 17.95 |       |
| JM456           | John Mayer        | 16.95 | 1000  |
| JM151           | John Mayer        | 16.95 | 1000  |
| MX789           | Madonna           | 11.95 | 500   |
| DJM237          | John Denver       | 11.95 | 2000  |
| DJM237          | Michael Jackson   | 11.95 | 2000  |
| DJM237          | Madonna           | 11.95 | 2000  |
| DR711           | Diana Ross        | 12.95 | 1000  |
| PM137           | Paul McCartney    | 19.95 |       |

## does the below relation satisfy 1NF?

#### EMPLOYEE

| Name     | Age | Sex | Emp#   |
|----------|-----|-----|--------|
| Anderson | 21  | F   | 010110 |
| Decker   | 22  | Μ   | 010100 |
| Glover   | 22  | Μ   | 101000 |
| Jackson  | 21  | F   | 201100 |
| Moore    | 19  | Μ   | 111100 |
| Nakata   | 20  | F   | 111101 |

# full functional dependency defined

a functional dependency of the form  $Z \rightarrow A$  is a 'full functional dependency' if and only if no proper subset of Z functionally determines A.

if  $Z \rightarrow A$  and  $X \rightarrow A$ , and X is a proper subset of Z, then Z does not fully functionally determine A, i.e.,  $Z \rightarrow A$  is not a full functional dependency; it is a <u>partial dependency</u>.

# second normal form



at least one of the following conditions applies:

- primary key consists of a single attribute
- no non-key attributes
- every non-key attribute depends on all of the primary key (fully functionally dependent)

# 2NF violation

| NEW_ALBUM |
|-----------|
|-----------|

| Album_no | Artist_nm         | Price | Stock |
|----------|-------------------|-------|-------|
| BS123    | Britney Spears    | 17.95 | 1000  |
| JT111    | Justin Timberlake | 17.95 | 1200  |
| BTL007   | John Lennon       | 23.95 |       |
| BTL007   | Paul McCartney    | 23.95 |       |
| BTL007   | George Harrison   | 23.95 |       |
| BTL007   | Ringo Star        | 23.95 |       |
| MJ100    | Michael Jackson   | 17.95 |       |
| JM456    | John Mayer        | 16.95 | 1000  |
| JM151    | John Mayer        | 16.95 | 1000  |
| MX789    | Madonna           | 11.95 | 500   |
| DJM237   | John Denver       | 11.95 | 2000  |
| DJM237   | Michael Jackson   | 11.95 | 2000  |
| DJM237   | Madonna           | 11.95 | 2000  |
| DR711    | Diana Ross        | 12.95 | 1000  |
| PM137    | Paul McCartney    | 19.95 |       |

candidate key: {Album\_no, Artist\_nm} Album\_no  $\rightarrow$  Price Album\_no  $\rightarrow$  Stock

# modification anomalies

#### change the value of price or stock of Album\_no BTL007 in NEW\_ALBUM

multiple tuples require update and failure to update some can cause an update anomaly

#### add a new tuple (Album\_no: XY11, Price: 17.95 and Stock:100) to NEW\_ALBUM

cannot insert without artist name, which is an insertion anomaly

#### delete Album\_no BTL007 from NEW\_ALBUM

requires deletion of multiple tuples and failure to delete some can cause a deletion anomaly

# resolution of 2NF violation



pull out the undesirable FD(s) from the target relation schema as a separate relation schema(s)

keep the determinant (left side of the FD equation) of the pulled-out relation schema as an attribute(s) in the leftover target relation schema

# resolution of 2NF violation

#### ALBUM\_INFO

| <u>Album_no</u> | Price | Stock |
|-----------------|-------|-------|
| BS123           | 17.95 | 1000  |
| JT111           | 17.95 | 1200  |
| BTL007          | 23.95 |       |
| MJ100           | 17.95 |       |
| JM456           | 16.95 | 1000  |
| JM151           | 16.95 | 1000  |
| MX789           | 11.95 | 500   |
| DJM237          | 11.95 | 2000  |
| DR711           | 12.95 | 1000  |
| PM137           | 19.95 |       |

#### ALBUM\_ARTIST

| Album_no | <u>Artist_nm</u>  |
|----------|-------------------|
| BS123    | Britney Spears    |
| JT111    | Justin Timberlake |
| BTL007   | John Lennon       |
| BTL007   | Paul McCartney    |
| BTL007   | George Harrison   |
| BTL007   | Ringo Star        |
| MJ100    | Michael Jackson   |
| JM456    | John Mayer        |
| JM151    | John Mayer        |
| MX789    | Madonna           |
| DJM237   | John Denver       |
| DJM237   | Michael Jackson   |
| DJM237   | Madonna           |
| DR711    | Diana Ross        |
| PM137    | Paul McCartney    |

## does the below relation satisfy 2NF?

#### EMPLOYEE

| Name     | Age | Sex | Emp#   |
|----------|-----|-----|--------|
| Anderson | 21  | F   | 010110 |
| Decker   | 22  | Μ   | 010100 |
| Glover   | 22  | Μ   | 101000 |
| Jackson  | 21  | F   | 201100 |
| Moore    | 19  | Μ   | 111100 |
| Nakata   | 20  | F   | 111101 |

yes, because the primary key is one attribute

somewhat arbitrary

# choosing a primary key

least number of attributes may be a good choice

numeric attributes and/or of small sizes may be easy to work with for a developer

# MAJOR(<u>SNUM</u>, MAJOR-DEPT, DEPT-HEAD)

| SNUM | MAJOR-DEPT | DEPT-HEAD |
|------|------------|-----------|
| 1000 | ANTHRO     | KHALIFA   |
| 2000 | STAT       | WAYNE     |
| 3000 | MUSIC      | DRAKE     |
| 4000 | STAT       | WAYNE     |
| 5000 | STAT       | WAYNE     |
| 6000 | MUSIC      | DRAKE     |

what information is lost if student number 1000 is deleted?

if the dept. head of MUSIC changes, what is required to update the relation?

what happens if we want to insert DRE as the department head of ZOOLOGY? what is required to accomplish this update?

# transitive dependency: existence of a functional dependence between two non-key attributes

 $MAJOR-DEPT \rightarrow DEPT-HEAD$  $DEPT-HEAD \rightarrow MAJOR-DEPT$ 

## transitive dependency illustrated



### transitive dependency illustrated



# third normal form





- it is in second normal form
- it has no transitive dependencies

| attribute | non-key attribute |           |
|-----------|-------------------|-----------|
| SNUM      | MAJOR-DEPT        | DEPT-HEAD |
| 1000      | ANTHRO            | KHALIFA   |
| 2000      | STAT              | WAYNE     |
| 3000      | MUSIC             | DRAKE     |
| 4000      | STAT              | WAYNE     |
| 5000      | STAT              | WAYNE     |
| 6000      | MUSIC             | DRAKE     |

SNUM → MAJOR-DEPT DEPT-HEAD → MAJOR-DEPT X SNUM → DEPT-HEAD MAJOR-DEPT -/-> SNUM X MAJOR-DEPT → DEPT-HEAD DEPT-HEAD -/-> SNUM

## STU-MAJOR (<u>SNUM</u>, MAJOR-DEPT) HEAD (<u>MAJOR-DEPT</u>, DEPT-HEAD)

| SNUM | MAJOR-DEPT |  |           |
|------|------------|--|-----------|
| 1000 | ANTHRO     | resolution: move<br>transitive dependency to<br>new relation |           |
| 2000 | STAT       |  |           |
| 3000 | MUSIC      |  |           |
| 4000 | STAT       |  |           |
| 5000 | STAT       |  |           |
| 6000 | MUSIC      |  |           |
|      |            | MAJOR-DEPT   | DEPT-HEAD |
|      |            | ANTHRO   | KHALIFA   |
|      |            | STAT   | WAYNE     |
|      |            | MUSIC  | DRAKE     |

# **3NF** violation

#### FLIGHT (Flight #, Origin, Destination, Mileage)

FLIGHT

| Flight# | Origin      | Destination | Mileage |
|---------|-------------|-------------|---------|
| DL507   | Seattle     | Denver      | 1537    |
| DL123   | Chicago     | Dallas      | 1058    |
| DL723   | Boston      | St. Louis   | 1214    |
| DL577   | Denver      | Los Angeles | 1100    |
| DL5219  | Minneapolis | St. Louis   | 580     |
| DL357   | Chicago     | Dallas      | 1058    |
| DL555   | Denver      | Houston     | 1100    |
| DL5237  | Cleveland   | St. Louis   | 580     |
| DL5271  | Chicago     | Cleveland   | 300     |

Flight # → Origin Flight # → Destination {Origin, Destination} → Mileage

# **3NF** violation

#### FLIGHT (Flight #, Origin, Destination, Mileage)

FLIGHT

| Flight# | Origin      | Destination | Mileage |
|---------|-------------|-------------|---------|
| DL507   | Seattle     | Denver      | 1537    |
| DL123   | Chicago     | Dallas      | 1058    |
| DL723   | Boston      | St. Louis   | 1214    |
| DL577   | Denver      | Los Angeles | 1100    |
| DL5219  | Minneapolis | St. Louis   | 580     |
| DL357   | Chicago     | Dallas      | 1058    |
| DL555   | Denver      | Houston     | 1100    |
| DL5237  | Cleveland   | St. Louis   | 580     |
| DL5271  | Chicago     | Cleveland   | 300     |

add a flight between Cincinnati to Houston that is 1100 miles (need flight number) remove flight # DL507 (lost mileage and origin/destination information)

# resolution of 3NF violation

#### FLIGHT

| Flight# | Origin      | Destination |
|---------|-------------|-------------|
| DL507   | Seattle     | Denver      |
| DL123   | Chicago     | Dallas      |
| DL723   | Boston      | St. Louis   |
| DL577   | Denver      | Los Angeles |
| DL5219  | Minneapolis | St. Louis   |
| DL357   | Chicago     | Dallas      |
| DL555   | Denver      | Houston     |
| DL5237  | Cleveland   | St. Louis   |
| DL5271  | Chicago     | Cleveland   |

#### DISTANCE

| <u>Origin</u> | Destination | Mileage |
|---------------|-------------|---------|
| Seattle       | Denver      | 1537    |
| Chicago       | Dallas      | 1058    |
| Boston        | St. Louis   | 1214    |
| Denver        | Los Angeles | 1100    |
| Minneapolis   | St. Louis   | 580     |
| Denver        | Houston     | 1100    |
| Cleveland     | St. Louis   | 580     |
| Chicago       | Cleveland   | 300     |

| Normal Form | Requirements   | Decomposition Rules   |
|-------------|--|---|
| First       | No multi-valued attributes   | Form new relations for each<br>multivalued attribute or repeating<br>group  |
| Second      | Satisfy at least one of the following<br>three conditions:<br>Primary key consists of a single<br>attribute<br>No non-key attributes<br>No non-key attribute should be<br>functionally dependent on part of<br>the primary key (every non-key<br>attribute should be fully functionally<br>dependent on the primary key) | Decompose and setup a new<br>relation for each partial key with its<br>dependent attribute(s). Make sure to<br>keep a relation with the original<br>primary key and any attributes that<br>are fully functionally dependent on it |
| Third       | No transitive dependencies.<br>Relation should be in second normal<br>form and should not have a non-key<br>attribute functionally determined by<br>another non-key attribute (or a set<br>of non-key attributes)  | Decompose and set up a new<br>relation that includes the nonkey<br>attribute(s) that functionally<br>determine(s) the other nonkey<br>attributes  |

# "The data depends on the key [1NF], the whole key [2NF] and nothing but the key [3NF] (so help me Codd<sup>1</sup>)."

<sup>1</sup> Source: http://en.wikipedia.org/wiki/Ted%5FCodd The examples in the next four slides are from: http://stackoverflow.com/questions/723998/

| Course ID | Semester | # Places | Course Name |
|-----------|----------|----------|-------------|
| IT101     | 2009-1   | 100      | Programming |
| IT101     | 2009-2   | 120      | Programming |
| IT102     | 2009-1   | 200      | Databases   |
| IT102     | 2010-1   | 150      | Databases   |
| IT103     | 2009-2   | 120      | Web Design  |

**X** Course ID  $\rightarrow$  Course Name

#### COURSE-DESC (Course ID, Course Name)

| Course ID | Course Name |
|-----------|-------------|
| IT101     | Programming |
| IT102     | Databases   |
| IT103     | Web Design  |

#### CLASS (Course ID, Semester, #Places)

| Course ID | Semester | # Places |
|-----------|----------|----------|
| IT101     | 2009-1   | 100      |
| IT101     | 2009-2   | 120      |
| IT102     | 2009-1   | 200      |
| IT102     | 2010-1   | 150      |
| IT103     | 2009-2   | 120      |

| Course ID | Semester | # Places | Teacher ID | Teacher<br>Name |
|-----------|----------|----------|------------|-----------------|
| IT101     | 2009-1   | 100      | 332        | Mr. Jones       |
| IT101     | 2009-2   | 120      | 332        | Mr. Jones       |
| IT102     | 2009-1   | 200      | 495        | Mr. Bentley     |
| IT102     | 2010-1   | 150      | 332        | Mr. Jones       |
| IT103     | 2009-2   | 120      | 242        | Mrs. Smith      |

#### **X** Teacher ID $\rightarrow$ Teacher Name

#### COURSE (Course ID, Semester, # Places, Teacher ID)

| Course ID | Semester | # Places | Teacher ID |
|-----------|----------|----------|------------|
| IT101     | 2009-1   | 100      | 332        |
| IT101     | 2009-2   | 120      | 332        |
| IT102     | 2009-1   | 200      | 495        |
| IT102     | 2010-1   | 150      | 332        |
| IT103     | 2009-2   | 120      | 242        |

#### TEACHER (Teacher ID, Teacher Name)

| Teacher ID | Teacher Name |
|------------|--------------|
| 332        | Mr. Jones    |
| 495        | Mr. Bentley  |
| 242        | Mrs. Smith   |

| SID | Activity    | Fee |
|-----|-------------|-----|
| 100 | Basketball  | 100 |
| 120 | Jousting    | 75  |
| 150 | Racquetball | 75  |
| 180 | Jousting    | 75  |

 $SID \rightarrow Activity$  $SID \rightarrow Fee$ 

$$\bigstar \quad \text{Activity} \rightarrow \text{Fee}$$

### STU-ACT (SID, Activity)

| SID | Activity    |
|-----|-------------|
| 100 | Basketball  |
| 120 | Jousting    |
| 150 | Racquetball |
| 180 | Jousting    |

### ACTIVITY-FEE (Activity, Fee)

| Activity    | Fee |
|-------------|-----|
| Basketball  | 100 |
| Jousting    | 75  |
| Racquetball | 75  |

## STU-CLASS (<u>Snum</u>, Sname, Major, <u>Cname</u>, Time Room)

| Snum | Sname     | Major        | Cname | Time | Room  |
|------|-----------|--------------|-------|------|-------|
| 0110 | KHUMAWALA | ACCOUNTING   | BA482 | MW3  | C-150 |
| 0110 | KHUMAWALA | ACCOUNTING   | BD445 | TR2  | C-213 |
| 0110 | KHUMAWALA | ACCOUNTING   | BA491 | TR3  | C-141 |
| 1000 | STEDRY    | ANTHROPOLOGY | AP150 | MWF9 | D-412 |
| 1000 | STEDRY    | ANTHROPOLOGY | BD445 | TR2  | C-213 |
| 2000 | KHUMAWALA | STATISTICS   | BA491 | TR3  | C-141 |
| 2000 | KHUMAWALA | STATISTICS   | BD445 | TR2  | C-213 |
| 3000 | GAMBLE    | ACCOUNTING   | BA482 | MW3  | C-150 |
| 3000 | GAMBLE    | ACCOUNTING   | BP490 | MW4  | C-150 |

what if the time for a class has to be updated? what if we need to add a new student to the database? what happens if we delete the student STEDRY?

## STU-CLASS (<u>Snum</u>, Sname, Major, <u>Cname</u>, Time Room)

| Snum | Sname     | Major        | Cname | Time | Room  |
|------|-----------|--------------|-------|------|-------|
| 0110 | KHUMAWALA | ACCOUNTING   | BA482 | MW3  | C-150 |
| 0110 | KHUMAWALA | ACCOUNTING   | BD445 | TR2  | C-213 |
| 0110 | KHUMAWALA | ACCOUNTING   | BA491 | TR3  | C-141 |
| 1000 | STEDRY    | ANTHROPOLOGY | AP150 | MWF9 | D-412 |
| 1000 | STEDRY    | ANTHROPOLOGY | BD445 | TR2  | C-213 |
| 2000 | KHUMAWALA | STATISTICS   | BA491 | TR3  | C-141 |
| 2000 | KHUMAWALA | STATISTICS   | BD445 | TR2  | C-213 |
| 3000 | GAMBLE    | ACCOUNTING   | BA482 | MW3  | C-150 |
| 3000 | GAMBLE    | ACCOUNTING   | BP490 | MW4  | C-150 |

#### identify the 6 undesirable functional dependencies

## STU-CLASS (Snum, Sname, Major, Cname, Time Room)

| Snum | Sname     | Major        | Cname                | Time | Room  |
|------|-----------|--------------|----------------------|------|-------|
| 0110 | KHUMAWALA | ACCOUNTING   | BA482                | MW3  | C-150 |
| 0110 | KHUMAWALA | ACCOUNTING   | BD445                | TR2  | C-213 |
| 0110 | KHUMAWALA | ACCOUNTING   | BA491                | TR3  | C-141 |
| 1000 | STEDRY    | ANTHROPOLOGY | AP150                | MWF9 | D-412 |
| 1000 | STEDRY    | ANTHROPOLOGY | BD445                | TR2  | C-213 |
| 2000 | KHUMAWALA | STATISTICS   | BA491                | TR3  | C-141 |
| 2000 | KHUMAWALA | STATISTICS   | BD445                | TR2  | C-213 |
| 3000 | GAMBLE    | ACCOUNTING   | BA482                | MW3  | C-150 |
| 3000 | GAMBLE    | ACCOUNTING   | BP490                | MW4  | C-150 |
|      | Snum →    | Sname Cname  | $e \rightarrow Time$ |      |       |

Snum  $\rightarrow$  Major Time  $\rightarrow$  Cname Cname  $\rightarrow$  Room Time  $\rightarrow$  Room

## decomposing relations - getting to 3NF

some basic rules:

if at 1NF and a non-key attribute is functionally dependent on part of the primary key, those attributes should appear together in a smaller relation

if at 2NF and there is a functional dependence between two non-key attributes, those attributes should appear together in a smaller relation

#### STUDENT (SNUM, SNAME, MAJOR)

| SNUM | SNAME     | MAJOR        |
|------|-----------|--------------|
| 0110 | KHUMAWALA | ACCOUNTING   |
| 1000 | STEDRY    | ANTHROPOLOGY |
| 2000 | KHUMAWALA | STATISTICS   |
| 3000 | GAMBLE    | ACCOUNTING   |

#### ENROLL (SNUM, CNAME)

#### CLASS (CNAME, TIME, ROOM)

| CNAME | TIME | ROOM  |
|-------|------|-------|
| BA482 | MW3  | C-150 |
| BD445 | TR2  | C-213 |
| BA491 | TR3  | C-141 |
| BP490 | MW4  | C-150 |
| AP150 | MWF9 | D-412 |

| SNUM | CNAME |
|------|-------|
| 0110 | BA482 |
| 0110 | BD445 |
| 0110 | BA491 |
| 1000 | AP150 |
| 1000 | BD445 |
| 2000 | BA491 |
| 2000 | BD445 |
| 3000 | BA482 |
| 3000 | BP490 |

### CAR (Model, <u># Cylinders</u>, Origin, Tax, Fee)

| Model       | #cylinders | Origin | Тах | Fee |
|-------------|------------|--------|-----|-----|
| Camry       | 4          | Japan  | 15  | 30  |
| Mustang     | 6          | USA    | 0   | 45  |
| Fiat        | 4          | Italy  | 18  | 30  |
| Accord      | 4          | Japan  | 15  | 30  |
| Century     | 8          | USA    | 0   | 45  |
| Mustang     | 4          | Canada | 0   | 30  |
| Monte Carlo | 6          | Canada | 0   | 45  |
| Civie       | 4          | Japan  | 15  | 30  |
| Mustang     | 4          | Mexico | 15  | 30  |
| Mustang     | 6          | Mexico | 15  | 45  |
| Civie       | 4          | Korea  | 15  | 30  |
|             |            |        |     |     |

**X** Origin  $\rightarrow$  Tax **X** # Cylinders  $\rightarrow$  Fee

#### CAR\_DESC (Model, <u># Cylinders</u>, Origin)

CAR\_TAX (Origin, Tax)

| Model   | # | Origin |
|---------|---|--------|
| Camry   | 4 | Japan  |
| Mustang | 6 | USA    |
| Fiat    | 4 | Italy  |
| Accord  | 4 | Japan  |
| Century | 8 | USA    |
| Mustang | 4 | Canada |
| Monte   | 6 | Canada |
| Civic   | 4 | Japan  |
| Mustang | 4 | Mexico |
| Mustang | 6 | Mexico |
| Civic   | 4 | Korea  |

| Origin | Tax |
|--------|-----|
| Japan  | 15  |
| USA    | 0   |
| Italy  | 18  |
| Canada | 0   |
| Mexico | 15  |
| Korea  | 15  |

#### CAR\_FEE (<u># Cylinders</u>, Fee)

| # Cylinders | Fee |
|-------------|-----|
| 4           | 30  |
| 6           | 45  |
| 8           | 45  |

#### SPORT (Stu #, Sport, Coach)

| Stu# | Sport      | Coach    |
|------|------------|----------|
| 125  | Football   | Register |
| 140  | Basketball | Lambert  |
| 220  | Baseball   | Register |
| 246  | Basketball | Lambert  |

Stu #  $\rightarrow$  Sport  $\Rightarrow$  Sport  $\rightarrow$  Coach

#### STUDENT-SPORT (Stu #, Sport)

| Stu# | Sport      |
|------|------------|
| 125  | Football   |
| 140  | Basketball |
| 220  | Baseball   |
| 246  | Basketball |

#### SPORT-COACH (Sport, Coach)

| Sport      | Coach    |
|------------|----------|
| Football   | Register |
| Basketball | Lambert  |
| Baseball   | Register |

# boyce-codd normal form



a relation is in BCNF if every determinant is a candidate key

# **BCNF** violation

#### STU\_SUB

| <u>Stu#</u> | <u>Subject</u> | Teacher  | Ap_score |
|-------------|----------------|----------|----------|
| IH123       | Chemistry      | Raturi   | 4        |
| IH123       | English        | Stephan  | 4        |
| IH235       | History        | Walker   | 5        |
| IH357       | English        | Campbell | 4        |
| IH571       | Chemistry      | Raturi   | 3        |
| IH235       | English        | Campbell | 4        |

{Stu #, Subject} → Teacher
{Stu #, Subject} → Ap\_score
X Teacher → Subject

Candidate keys: {Stu #, Subject}, {Stu #, Teacher}

# resolution of BCNF violation

| TEACH_SUB      |           |
|----------------|-----------|
| <u>Teacher</u> | Subject   |
|                |           |
| Raturi         | Chemistry |
| Stephan        | English   |
| Walker         | History   |
| Campbell       | English   |

| STU_A | Ρ |
|-------|---|
|-------|---|

| —           |                |          |
|-------------|----------------|----------|
| <u>Stu#</u> | <u>Teacher</u> | Ap_score |
| IH123       | Raturi         | 4        |
| IH123       | Stephan        | 4        |
| IH235       | Walker         | 5        |
| IH357       | Campbell       | 4        |
| IH571       | Raturi         | 3        |
| IH235       | Campbell       | 4        |

Teacher → Subject

{Stu #, Teacher}  $\rightarrow$  Ap\_Score

# BCNF violation - page 410

| Patient | Hospital   | Doctor    |
|---------|------------|-----------|
| Smith   | Methodist  | D. Cooley |
| Lee     | St. Luke's | Z. Zhang  |
| Marks   | Methodist  | D. Cooley |
| Marks   | St. Luke's | W. Lowe   |
| Lou     | Hermann    | R. Duke   |

Candidate keys: {Patient , Hospital}, {Patient, Doctor}

{Patient, Hospital} → Doctor
★ Doctor → Hospital

# resolution of BCNF violation

#### PAT-DOC (Patient, Doctor)

| Patient | Doctor    |
|---------|-----------|
| Smith   | D. Cooley |
| Lee     | Z. Zhang  |
| Marks   | D. Cooley |
| Marks   | W. Lowe   |
| Lou     | R. Duke   |

#### DOC-HOS (Doctor, Hospital)

| Doctor    | Hospital   |
|-----------|------------|
| D. Cooley | Methodist  |
| Z. Zhang  | St. Luke's |
| W. Lowe   | St. Luke's |
| R. Duke   | Hermann    |

| Normal Form | Requirements   | Decomposition Rules   |
|-------------|--|---|
| First       | No multi-valued attributes   | Form new relations for each multivalued attribute or repeating group  |
| Second      | Satisfy at least one of the following three<br>conditions:<br>Primary key consists of a single<br>attribute<br>No non-key attributes<br>No non-key attribute should be<br>functionally dependent on part of the<br>primary key (every non-key attribute<br>should be fully functionally dependent<br>on the primary key) | Decompose and setup a new relation for<br>each partial key with its dependent<br>attribute(s). Make sure to keep a relation<br>with the original primary key and any<br>attributes that are fully functionally<br>dependent on it |
| Third       | No transitive dependencies. Relation<br>should be in second normal form and<br>should not have a non-key attribute<br>functionally determined by another non-<br>key attribute (or a set of non-key<br>attributes)   | Decompose and set up a new relation<br>that includes the nonkey attribute(s) that<br>functionally determine(s) the other<br>nonkey attributes   |
| BCNF        | Every determinant is a candidate key   | Decompose and set up a new relation<br>that includes the non-candidate key<br>attribute(s) that functionally determine(s)<br>the other nonkey attributes.   |