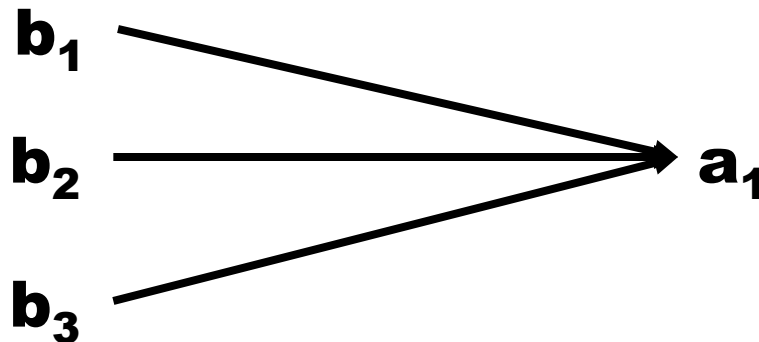


# **Database Design: Normalization**

A thick, horizontal yellow brushstroke underline that spans the width of the slide, positioned below the title text.

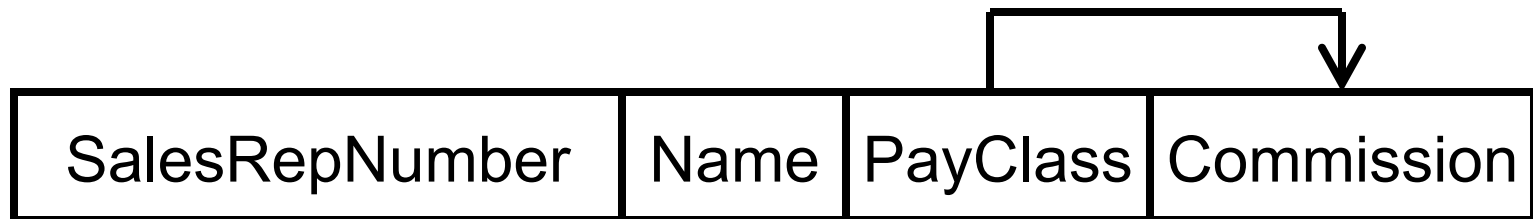
# Functional Dependence

*An attribute  $A$  is functionally dependent on attribute(s)  $B$  if: given a value  $b$  for  $B$  there is one and only one corresponding value  $a$  for  $A$  (at a time).*



# Example: functional dependence

*All sales representatives in a given pay class have the same commission rate.*



# Keys



Primary Key: a minimal set of attributes that form a candidate key

*Any attribute or collection of attributes that functionally determine all attributes in a record is a Candidate Key.*

*Note: since no two rows in a relational table can be duplicates, the entire record is always a candidate key.*

# Primary Key (C)



- Primary key determines all attributes
- No subset of the attributes in the primary key is a candidate key

*A key consisting of more than one attribute is called a "composite key."*

# Good Primary Keys



- Do not change over the life of the database
- Are not “intelligent keys”
- Are not too long
- Do not consist of too many attributes (3 or fewer is good)

# Foreign Keys

A value in the “child” table that matches with the related value in the “parent” table.

SalesRep(**SalesRepNumber**, Name)

[ 03 | Mary Jones ]



[ 124 | 03 ]

Customer(CustomerNumber, **SalesRepNumber**)

# What is normalization



- The process by which we efficiently organize data to achieve these goals:
  - Eliminating redundancy
  - Ensuring data is stored in the correct table
  - Eliminating need for restructuring database when data is added.
  - Five levels of normal form
  - In order to achieve one level of normal form, each previous level must be met
- **Third normal form is sufficient for most typical database applications.**



# Normal Forms



- A set of conditions on table structure that improves maintenance. Normalization removes processing anomalies:
  - Update
  - Inconsistent Data
  - Addition
  - Deletion

# Normal Forms



All attributes depend on the key, the whole key and nothing but the key.

1NF Keys and no repeating groups

2NF No partial dependencies

3NF All determinants are candidate keys

# 1st Normal Form



- Table has a primary key
- Table has no repeating groups
- There are no repeating or duplicate fields.

A multivalued attribute is an attribute that may have several values for one record

# Example



## ■ Multivalued attribute:

Orders(OrderNumber, OrderDate, {PartNumber})  
[ 12491 | 9/02/2001 | BT04, BZ66 ]

## Example

item	colors	price	tax
T-shirt	red, blue	12.00	0.60
polo	red, yellow	12.00	0.60
T-shirt	red, blue	12.00	0.60
sweatshirt	blue, black	25.00	1.25

Table is not in first normal form because:

- Multiple items in color field
- Duplicate records / no primary key

# Normalization: 1NF



- Every repeating group becomes a new table with the appropriate foreign key relationships preserved.



## Example

item	color	price	tax
T-shirt	red	12.00	0.60
T-shirt	blue	12.00	0.60
polo	red	12.00	0.60
polo	yellow	12.00	0.60
sweatshirt	blue	25.00	1.25
sweatshirt	black	25.00	1.25

Table is now in first normal form.

# 2nd Normal Form



- No partial dependencies

*No attribute depends on only some of the attributes of a composite key.*

- All non-key fields depend on all components of the primary key.

- **Guaranteed when primary key is a single field..**



## Example



item	color	price	tax
T-shirt	red	12.00	0.60
T-shirt	blue	12.00	0.60
polo	red	12.00	0.60
polo	yellow	12.00	0.60
sweatshirt	blue	25.00	1.25
sweatshirt	black	25.00	1.25

Table is not in second normal form because:

- **price** and **tax** depend on **item**, but not **color**

## Example



item	color
T-shirt	red
T-shirt	blue
polo	red
polo	yellow
sweatshirt	blue
sweatshirt	black

item	price	tax
T-shirt	12.00	0.60
polo	12.00	0.60
sweatshirt	25.00	1.25

Tables are now in second normal form.

# 3rd Normal Form



- No non-key field depends upon another.
- All non-key fields depend only on the primary key



## Example

item	color
T-shirt	red
T-shirt	blue
polo	red
polo	yellow
sweatshirt	blue
sweatshirt	black

item	price	tax
T-shirt	12.00	0.60
polo	12.00	0.60
sweatshirt	25.00	1.25

Tables are not in third normal form because:

- **tax** depends on **price**, not **item**



## Example

item	color
T-shirt	red
T-shirt	blue
polo	red
polo	yellow
sweatshirt	blue
sweatshirt	black

item	price
T-shirt	12.00
polo	12.00
sweatshirt	25.00

price	tax
12.00	0.60
25.00	1.25

Tables are now in third normal form.



## Another Example

Name	Assignment 1	Assignment 2
Jeff Smith	Article Summary	Poetry Analysis
Nancy Jones	Article Summary	Reaction Paper
Jane Scott	Article Summary	Poetry Analysis

Table is not in first normal form because:

- Assignment field repeating
- First and last name in one field
- No (guaranteed unique) primary key field

## Another Example



Assignment ID	Description
1	Article Summary
2	Poetry Analysis
3	Reaction Paper

Student ID	First Name	Last Name
1	Jeff	Smith
2	Nancy	Jones
3	Jane	Scott

Assignment ID	Student ID
1	1
1	2
1	3
2	1
2	3
3	2

Tables are in third normal form.

# Normalization



- Improves maintenance for database changes

*Tends to slow down retrieval*