

# Normalization

# Normalization Introduction

- It is a systematic way of ensuring that a database structure is :
  - suitable for general-purpose querying
  - free of certain undesirable characteristics—insertion, update, and deletion anomalies
- Any of the above mentioned problem may lead to loss of data integrity
- E.F. Codd and other theorists has desinged these normalization forms (rules)

# Normalization Introduction

- Till now 6 Normalization forms (also known as 6NF) has been launched till date.
- Latest Normal Form 6NF is introduced in the year 2002 by Chris Date, Hugh Darwen and Nikoz Lorentzos
- Informally a table is said to be normalized if it is in 3NF
- Moving towards extra normalization may cause performance issues due to increased granularity in tables

**1<sup>st</sup> NF**



# First Normalization Form

Defined by : E.F.Codd (1970)

In order to follow the 1<sup>st</sup> NF table should satisfies following five conditions

1. There's no top-to-bottom ordering to the rows.
2. There are no duplicate rows.
3. Every row-and-column intersection contains exactly one value from the applicable domain (and nothing else).
4. All columns are regular [i.e. rows have no hidden components such as row IDs, object IDs, or hidden timestamps].

# First Normalization Form : Repeating Groups

Customer ID	First Name	Surname	Telephone Number
123	Robert	Ingram	555-861-2025
456	Jane	Wright	555-403-1659
456	Jane	Wright	555-776-4100
789	Maria	Fernandez	555-808-9633

# First Normalization Form : Repeating Groups

A design that complies 1NF

Customer ID	First Name	Surname
123	Robert	Ingram
456	Jane	Wright
789	Maria	Fernandez

Customer ID	Telephone Number
123	555-861-2025
456	555-403-1659
456	555-776-4100
789	555-808-9633

**2<sup>nd</sup> NF**





# Second Normalization Form

Defined by : E.F.Codd (1971)

In order to follow the 2<sup>nd</sup> NF table should satisfies following conditions

1. Table should be in 1NF
2. 1NF table is in 2NF if and only if **none** of its **non-prime attributes** are **functionally dependent** on a part (proper subset) of a candidate key

Note : When a 1NF table has no composite candidate keys (candidate keys consisting of more than one attribute), the table is automatically in 2NF.

# Second Normalization Form

Defined by : E.F.Codd (1971)

<u>Employee</u>	<u>Skill</u>	Current Work Location
Jones	Typing	114 Main Street
Jones	Shorthand	114 Main Street
Jones	Whittling	114 Main Street
Bravo	Light Cleaning	73 Industrial Way
Ellis	Alchemy	73 Industrial Way
Ellis	Flying	73 Industrial Way
Harrison	Light Cleaning	73 Industrial Way

# Second Normalization Form

Defined by : E.F.Codd (1971)

<u>Employee</u>	Current Work Location
Jones	114 Main Street
Bravo	73 Industrial Way
Ellis	73 Industrial Way
Harrison	73 Industrial Way

<u>Employee</u>	<u>Skill</u>
Jones	Typing
Jones	Shorthand
Jones	Whittling
Bravo	Light Cleaning
Ellis	Alchemy
Ellis	Flying
Harrison	Light Cleaning

# Second Normalization Form

Defined by : E.F.Codd (1971)

## Electric Toothbrush Models

<b>Manufacturer</b>	<b>Model</b>	<b><u>Model Full Name</u></b>	<b>Manufacturer Country</b>
Forte	X-Prime	Forte X-Prime	Italy
Forte	Ultraclean	Forte Ultraclean	Italy
Dent-o-Fresh	EZbrush	Dent-o-Fresh EZBrush	USA
Kobayashi	ST-60	Kobayashi ST-60	Japan
Hoch	Toothmaster	Hoch Toothmaster	Germany
Hoch	Contender	Hoch Contender	Germany

# Second Normalization Form

Defined by : E.F.Codd (1971)

## Electric Toothbrush Manufacturers

<u>Manufacturer</u>	Manufacturer Country
Forte	Italy
Dent-o-Fresh	USA
Kobayashi	Japan
Hoch	Germany

## Electric Toothbrush Models

<u>Manufacturer</u>	<u>Model</u>	Model Full Name
Forte	X-Prime	Forte X-Prime
Forte	Ultraclean	Forte Ultraclean
Dent-o-Fresh	EZbrush	Dent-o-Fresh EZBrush
Kobayashi	ST-60	Kobayashi ST-60
Hoch	Toothmaster	Hoch Toothmaster
Hoch	Contender	Hoch Contender

**3<sup>rd</sup> NF**



# Third Normalization Form

Defined by : E.F.Codd (1971)

In order to follow the 3<sup>rd</sup> NF table should satisfies following conditions

1. Table should be in 2NF
2. Every **non-prime** attribute of R is **non-transitively** dependent (i.e. directly dependent) on **every key of R**
  - A **non-prime attribute** of R is an attribute that does not belong to any candidate key of R
3. **Bill Kent**: every non-key attribute "**must**" provide a fact about the key, the whole key, and nothing but the key

# Third Normalization Form

Defined by : E.F.Codd (1971)

## Tournament Winners

<u>Tournament</u>	<u>Year</u>	Winner	Winner Date of Birth
Indiana Invitational	1998	Al Fredrickson	21 July 1975
Cleveland Open	1999	Bob Albertson	28 September 1968
Des Moines Masters	1999	Al Fredrickson	21 July 1975
Indiana Invitational	1999	Chip Masterson	14 March 1977



# Third Normalization Form

Defined by : E.F.Codd (1971)

**Tournament Winners**

<u>Tournament</u>	<u>Year</u>	Winner
Indiana Invitational	1998	Al Fredrickson
Cleveland Open	1999	Bob Albertson
Des Moines Masters	1999	Al Fredrickson
Indiana Invitational	1999	Chip Masterson

**Player Dates of Birth**

<u>Player</u>	Date of Birth
Chip Masterson	14 March 1977
Al Fredrickson	21 July 1975
Bob Albertson	28 September 1968

## 3.5 NF (BCNF)



# Boyce-Codd Normalization Form

Defined by : E.F.Codd and Raymond Boyce(1974)

It is a slightly stronger version of the third normal form (3NF)

Also known as concept of Trivial Functional dependency

1. Table should be in 2NF
2. Every **non-prime** attribute of R is **non-transitively** dependent (i.e. directly dependent) on **every key of R**
  - A **non-prime attribute** of R is an attribute that does not belong to any candidate key of R
3. **Bill Kent**: every non-key attribute "**must**" provide a fact about the key, the whole key, and nothing but the key

If a non-key attribute X is dependent on an attribute A then X should be the candidate key / superset of candidate key

# Boyce-Codd Normalization Form

Defined by : E.F.Codd and Raymond Boyce(1974)

Today's Court Bookings			
Court	Start Time	End Time	Rate Type
1	09:30	10:30	SAVER
1	11:00	12:00	SAVER
1	14:00	15:30	STANDARD
2	10:00	11:30	PREMIUM-B
2	11:30	13:30	PREMIUM-B
2	15:00	16:30	PREMIUM-A

# Boyce-Codd Normalization Form

Defined by : E.F.Codd and Raymond Boyce(1974)

Rate Type	Court
SAVER	1
STANDARD	1
PREMIUM-A	2
PREMIUM-B	2

Start Time	End Time	Rate Type
09:30	10:30	Saver
11:00	12:00	Saver
14:00	15:30	Standard
10:00	11:30	Premium-B
11:30	13:30	Premium-B
15:00	16:30	Premium-A

**4<sup>th</sup> NF**



# Fourth Normalization Form

Defined by : Ronald Fagin(1977)

4<sup>th</sup> normalization form is concerned with multivalued dependency

1. A table is in 4NF if and only if, for every one of its non-trivial multivalued dependencies  $X \twoheadrightarrow Y$ 
  - X is a superkey : That means X is either a candidate key or a superset thereof

# Fourth Normalization Form

Defined by : Ronald Fagin(1977)

## Pizza Delivery Permutations

<u>Restaurant</u>	<u>Pizza Variety</u>	<u>Delivery Area</u>
A1 Pizza	Thick Crust	Springfield
A1 Pizza	Thick Crust	Shelbyville
A1 Pizza	Thick Crust	Capital City
A1 Pizza	Stuffed Crust	Springfield
A1 Pizza	Stuffed Crust	Shelbyville
A1 Pizza	Stuffed Crust	Capital City
Elite Pizza	Thin Crust	Capital City
Elite Pizza	Stuffed Crust	Capital City
Vincenzo's Pizza	Thick Crust	Springfield
Vincenzo's Pizza	Thick Crust	Shelbyville
Vincenzo's Pizza	Thin Crust	Springfield
Vincenzo's Pizza	Thin Crust	Shelbyville



# Fourth Normalization Form

Defined by : Ronald Fagin(1977)

Varieties By Restaurant	
<u>Restaurant</u>	<u>Pizza Variety</u>
A1 Pizza	Thick Crust
A1 Pizza	Stuffed Crust
Elite Pizza	Thin Crust
Elite Pizza	Stuffed Crust
Vincenzo's Pizza	Thick Crust
Vincenzo's Pizza	Thin Crust

Delivery Areas By Restaurant	
<u>Restaurant</u>	<u>Delivery Area</u>
A1 Pizza	Springfield
A1 Pizza	Shelbyville
A1 Pizza	Capital City
Elite Pizza	Capital City
Vincenzo's Pizza	Springfield
Vincenzo's Pizza	Shelbyville

**5<sup>th</sup> NF**



# Fifth Normalization Form

Also known as **Project-join normal form (PJ/NF)**

1. designed to reduce redundancy in relational databases recording multi-valued facts by isolating semantically related multiple relationships

A join dependency  $*\{A, B, \dots Z\}$  on  $R$  is implied by the candidate key(s) of  $R$  if and only if each of  $A, B, \dots, Z$  is a superkey for  $R$

# Fifth Normalization Form

Travelling Salesman	Brand	Product Type
Jack Schneider	Acme	Vacuum Cleaner
Jack Schneider	Acme	Breadbox
Willy Loman	Robusto	Pruning Shears
Willy Loman	Robusto	Vacuum Cleaner
Willy Loman	Robusto	Breadbox
Willy Loman	Robusto	Umbrella Stand
Louis Ferguson	Robusto	Vacuum Cleaner
Louis Ferguson	Robusto	Telescope
Louis Ferguson	Acme	Vacuum Cleaner
Louis Ferguson	Acme	Lava Lamp
Louis Ferguson	Nimbus	Tie Rack

# Fifth Normalization Form

Product Types By Travelling Salesman	
Travelling Salesman	Product Type
Jack Schneider	Vacuum Cleaner
Jack Schneider	Breadbox
Willy Loman	Pruning Shears
Willy Loman	Vacuum Cleaner
Willy Loman	Breadbox
Willy Loman	Umbrella Stand
Louis Ferguson	Telescope
Louis Ferguson	Vacuum Cleaner
Louis Ferguson	Lava Lamp
Louis Ferguson	Tie Rack

Product Types By Brand	
Brand	Product Type
Acme	Vacuum Cleaner
Acme	Breadbox
Acme	Lava Lamp
Robusto	Pruning Shears
Robusto	Vacuum Cleaner
Robusto	Breadbox
Robusto	Umbrella Stand
Robusto	Telescope
Nimbus	Tie Rack

Brands By Travelling Salesman	
Travelling Salesman	Brand
Jack Schneider	Acme
Willy Loman	Robusto
Louis Ferguson	Robusto
Louis Ferguson	Acme
Louis Ferguson	Nimbus

**Example**



# Normalization Example

## denormalized table

Student#	Advisor	Adv-Room	Class1	Class2	Class3
1022	Jones	412	101-07	143-01	159-02
4123	Smith	216	201-01	211-02	214-01

# Normalization Example

## 1NF

Student#	Advisor	Adv-Room
1022	Jones	412
4123	Smith	216

Student#	Class#
1022	101-07
1022	143-01
1022	159-02
4123	201-01
4123	211-02
4123	214-01



# Normalization Example

## 2NF

Student#	Advisor	Adv-Room
1022	Jones	412
4123	Smith	216

Student#	Class#
1022	101-07
1022	143-01
1022	159-02
4123	201-01
4123	211-02
4123	214-01

# Normalization Example

## 3NF

Student#	Advisor	Adv-Room
1022	Jones	412
4123	Smith	216



Student#	Advisor
1022	Jones
4123	Smith

Name	Room	Dept
Jones	412	42
Smith	216	42

Student#	Class#
1022	101-07
1022	143-01
1022	159-02
4123	201-01
4123	211-02
4123	214-01