DATABASE NORMALIZATION

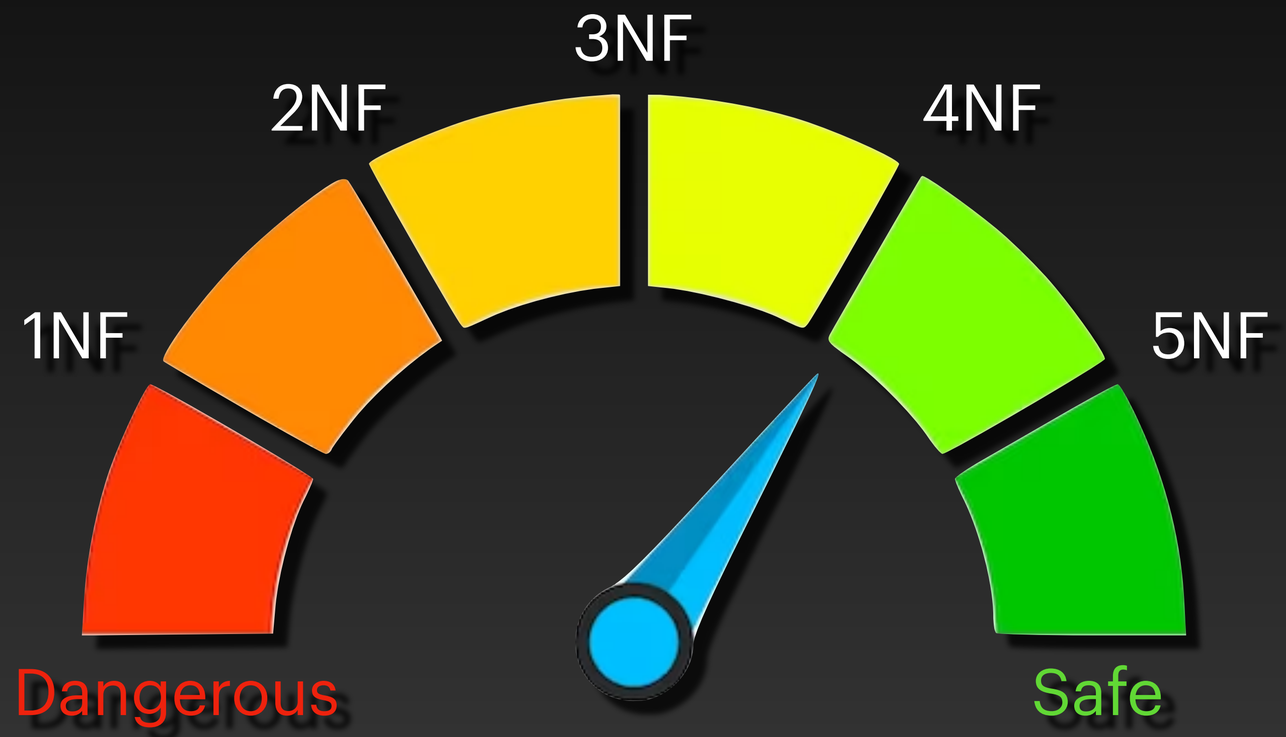
1NF, 2NF, 3NF, 4NF, 5NF

TECH VAULT

AMR ELHELW

customer_id	birth_date
123	17-APR-1992
123	05-OCT-1989

Normal Forms



First Normal Form (1NF)

team	members
Frontend	Tom, Anna, John
Backend	Jenny, Bob

Multi-valued cells violate 1NF

team	member
Frontend	Tom
Backend	Anna
Frontend	John
Backend	Jenny
Backend	Bob

```
SELECT member
FROM team_members
WHERE team='Backend'
```

member
Bob
Anna
Jenny

Most senior

Least senior

team	member_by_seniority
Frontend	John
Frontend	Tom
Backend	Jenny
Backend	Anna
Backend	Bob

```
SELECT member_by_seniority
FROM team_members
WHERE team='Backend'
```

member_by_seniority
Anna
Bob
Jenny

Using row order to convey information violates 1NF

team	member	join_date
Frontend	John	1998-08-28
Frontend	Tom	2006-09-08
Backend	Jenny	2021-01-28
Backend	Anna	2021-12-26
Backend	Bob	2022-09-21

```
SELECT member
FROM team_members
WHERE team='Backend'
ORDER BY join_date
```

team	member	join_date
Frontend	John	1998-08-28
Frontend	Tom	2006-09-08
Backend	Jenny	2021-01-28
Backend	Anna	2021-12-26
Backend	Bob	2022-09-21
Frontend	Tom	2006-09-08

Having duplicate rows violates 1NF

Primary Key (PK)

id	team	member	join_date
101	Frontend	John	1998-08-28
102	Frontend	Tom	2006-09-08
103	Backend	Jenny	2021-01-28
104	Backend	Anna	2021-12-26
105	Backend	Bob	2022-09-21

To guarantee there are no duplicate rows,
always define a primary key for each table

First Normal Form (1NF)

- No multi-valued attributes
- Cannot use row order to convey information
- Always define a primary key

Also

- No mixing data types within a column
- No multiple columns with the same name

Student_ID	Courses
S01	Math (95), Econ (80)
S02	Phys (86)
S03	Chem (91), Bio (76), Phys (74), Math (81)

student_course

student_id	course	score
S01	Math	95
S01	Econ	80
S02	Phys	86
S03	Chem	91
S03	Bio	76
S03	Phys	74
S03	Math	81

Second Normal Form (2NF)

student_course

	student_id	course	score	instructor
	S01	Math	95	Bob
	S01	Econ	80	Anna
	S02	Phys	86	Tom
	S03	Chem	91	Julia
	S03	Bio	76	Alice
	S03	Phys	74	Tom
	S03	Math	81	Bob

student_course

student_id	course	score	instructor
S01	Math	95	John
S01	Econ	80	Anna
S02	Phys	86	Tom
S03	Chem	91	Julia
S03	Bio	76	Alice
S03	Phys	74	Tom
S03	Math	81	Bob

Update Anomaly

Deletion Anomaly

New course: “English” taught by Nancy

Insertion Anomaly

student_course

Primary Key (PK)

Non-key attributes

student_id	course	score	instructor
S01	Math	95	Bob
S01	Econ	80	Anna
S02	Phys	86	Tom
S03	Chem	91	Julia
S03	Bio	76	Alice
S03	Phys	74	Tom
S03	Math	81	Bob

Second Normal Form (2NF)

- Must satisfy 1NF
- Each non-key attribute must depend on the entire PK

$\{student\_id, course\} \rightarrow \{score\}$

$\{course\} \rightarrow \{instructor\}$ ✗

course

course_id	instructor
Math	Bob
Econ	Anna
Phys	Tom
Chem	Julia
Bio	Alice

$\{course\_id\} \longrightarrow \{instructor\}$

student_course

student_id	course_id	score
S01	Math	95
S01	Econ	80
S02	Phys	86
S03	Chem	91
S03	Bio	76
S03	Phys	74
S03	Math	81

$\{student\_id, course\_id\} \longrightarrow \{score\}$

Third Normal Form (3NF)

course

course_id	instructor	phone
Math	Tom	123-4567
Econ	Anna	987-6543
Phys	Tom	345-6789
Chem	Julia	765-4321
Bio	Alice	456-7890

Third Normal Form (3NF)

- Must satisfy 2NF
- No non-key attribute is dependent on another non-key attribute

$\{course\_id\} \rightarrow \{instructor\}$

$\{course\_id\} \rightarrow \{instructor\} \rightarrow \{phone\}$ ✗

Transitive Dependency

course

course_id	instructor	phone
Math	Bob	123-4567
Econ	Anna	987-6543
Phys	Tom	345-6789
Chem	Julia	765-4321
Bio	Alice	456-7890

$\{course\_id\} \longrightarrow \{instructor\}$

instructor

instr_id	phone
Bob	123-4567
Anna	987-6543
Tom	345-6789
Julia	765-4321
Alice	456-7890

$\{instr\_id\} \longrightarrow \{phone\}$

Boyce-Codd Normal Form (BCNF)

Boyce-Codd Normal Form

BCNF - Upgraded version of 3NF (3.5 NF)

3NF

- Must satisfy 2NF
- No non-key attribute is dependent on another non-key attribute

BCNF

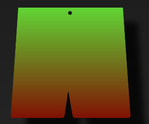
- Must satisfy 3NF
- No attribute is dependent on a non-key attribute

Fourth Normal Form (4NF)

Jeans

Colors: Blue, Black

Sizes: L, M



Shorts

Colors: Green, Grey

Sizes: S, XS

T-shirt

Colors: Green, Red, White

Sizes: L, S

Products

Item	Color	Size
Jeans	Blue	L
Jeans	Blue	M
Jeans	Black	L
Jeans	Black	M
Shorts	Green	S
Shorts	Green	XS
Shorts	Grey	S
Shorts	Grey	XS
T-Shirt	Green	L
T-Shirt	Green	S
T-Shirt	Red	L
T-Shirt	Red	S
T-Shirt	White	L
T-Shirt	White	S

Jeans

Colors: Blue, Black

Sizes: L, M

Shorts

Colors: Green, Grey

Sizes: S, XS

T-shirt

Colors: Green, Red, White

Sizes: L, S

Products

Item	Color	Size
Jeans	Blue	L
Jeans	Blue	M
Jeans	Black	L
Jeans	Black	M
Shorts	Green	S
Shorts	Green	XS
Shorts	Grey	S
Shorts	Grey	XS
T-Shirt	Green	L
T-Shirt	Green	S
T-Shirt	Red	L
T-Shirt	Red	S
T-Shirt	White	L

$$\{item\} \twoheadrightarrow \{color\}$$
$$\{item\} \twoheadrightarrow \{size\}$$

Fourth Normal Form (4NF)

- Must satisfy BCNF
- Every non-trivial multi-valued dependency must depend on the key

$X \twoheadrightarrow Y$ is trivial if Y is a subset of X, or X and Y together form the whole set of attributes of the table

Jeans

Colors: Blue, Black

Sizes: L, M

Shorts

Colors: Green, Grey

Sizes: S, XS

T-shirt

Colors: Green, Red, White

Sizes: L, S

Product_colors

Item	Color
Jeans	Blue
Jeans	Black
Shorts	Green
Shorts	Grey
T-Shirt	Green
T-Shirt	Red
T-Shirt	White

$\{item\} \twoheadrightarrow \{color\}$

Product_sizes

Item	Size
Jeans	L
Jeans	M
Shorts	S
Shorts	XS
T-Shirt	L
T-Shirt	S

$\{item\} \twoheadrightarrow \{size\}$

Fifth Normal Form (5NF)

Fifth Normal Form (5NF)

- Must satisfy 4NF
- There must be no join-dependencies.

- *Join-dependency: a table can be decomposed into a join of multiple tables without loss of information*

P1

Cheese
Veggie
Mushroom

P2

Cheese
Chicken

P3

Cheese
Mushroom
Chicken

Anna

Brands:
P2, P3
Flavors:
Chicken
Mushroom

Tom

Brands:
P1, P2, P3
Flavors:
Cheese
Veggie
Chicken

Preferences

Customer	Brand	Flavor
Anna	P2	Chicken
Anna	P3	Mushroom
Anna	P3	Chicken
Tom	P1	Cheese
Tom	P1	Veggie
Tom	P2	Cheese
Tom	P2	Chicken
Tom	P3	Cheese
Tom	P3	Chicken

P1

Cheese
Veggie
Mushroom

P2

Cheese
Chicken

P3

Cheese
Mushroom
Chicken

Anna

Brands:
P2, P3
Flavors:
Chicken
Mushroom

Tom

Brands:
P1, P2 , P3
Flavors:
Cheese
Veggie
Chicken

Brand_flavor

Brand	Flavor
P1	Cheese
P1	Veggie
P1	Mushroom
P2	Cheese
P2	Chicken
P3	Cheese
P3	Mushroom
P3	Chicken

Customer_brand

Customer	Brand
Anna	P2
Anna	P3
Tom	P1
Tom	P2
Tom	P3

Customer_flavor

Customer	Flavor
Anna	Chicken
Anna	Mushroom
Tom	Cheese
Tom	Veggie
Tom	Chicken

Summary

1NF

- No multi-valued attributes
- Cannot use row order to convey information
- Always define a primary key
- No mixing data types within a column
- No multiple columns with the same name

2NF

- Each non-key attribute must depend on the entire PK

3NF

- No non-key attribute is dependent on another non-key attribute

BCNF

- No attribute is dependent on a non-key attribute

4NF

- Every non-trivial multi-valued dependency must depend on the key

5NF

- There must be no join-dependencies