6. Normalisation

Normalization - **makes a big difference as to how efficient our databases are.**

The process of separating data out in different tables. The following describe the level of normalization:

- **First Normal Form** - atomic, no repeated groups, unique rows
- **Second Normal Form** - all about the PRIMARY KEY (non-key attributes must be dependent on the primary key)
- **Third Normal Form** - all about the NON KEYS (there are no transitive dependencies between the non key & primary key)

Pros

- Reduces data redundancy
    - Eliminates repeated information
- Increases ease of database maintenance
    - One source of truth for each entry
    - **Easier Inserts**
        - Can insert data about single entity
    - **Easier updates**
        - One location for each update
    - **Easier Deletes**
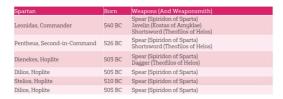   - One location of each delete

Cons

- Increases database depth
    - More tables
- Increases difficulty of querying
    - Multiple joins required
        - More effort for analysis/devs
        - More computational effort

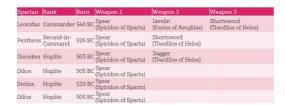If you need to priorities speed, a **non-normalized** database is more suitable.

First Normal Form (1NF)

- Data must be atomic (as small as it can be)
- There are no repeated groups
- Each row must be unique

Example: Does NOT fit criteria for **1NF**, multiple entries per row

Spartan	Born	Weapons (And Weaponsmith)
Leonidas, Commander	540 BC	Spear (Spiridon of Sparta) Javelin (Kostas of Amyklae) Shortsword (Theofilos of Helos)
Pentheus, Second-in-Command	526 BC	Spear (Spiridon of Sparta) Shortsword (Theofilos of Helos)
Dienekes, Hoplite	503 BC	Spear (Spiridon of Sparta) Dagger (Theofilos of Helos)
Dilios, Hoplite	505 BC	Spear (Spiridon of Sparta)
Stelios, Hoplite	510 BC	Spear (Spiridon of Sparta)
Dilios, Hoplite	505 BC	Spear (Spiridon of Sparta)

So: We separate the entries using **multiple columns, BUT** we have repeated groups.

Spartan	Rank	Born	Weapon 1	Weapon 2	Weapon 3
Leonidas	Commander	540 BC	Spear (Spiridon of Sparta)	Javelin (Kostas of Amyklae)	Shortsword (Theofilos of Helos)
Pentheus	Second-in-Command	526 BC	Spear (Spiridon of Sparta)	Shortsword (Theofilos of Helos)	
Dienekes	Hoplite	503 BC	Spear (Spiridon of Sparta)	Dagger (Theofilos of Helos)	
Dilios	Hoplite	505 BC	Spear (Spiridon of Sparta)		
Stelios	Hoplite	510 BC	Spear (Spiridon of Sparta)		
Dilios	Hoplite	505 BC	Spear (Spiridon of Sparta)		

So: We need to separate the weapon information into a separate table.**BUT** we have multiple entries per row again.

Spartan Name	Role	Born
Leonidas	Commander	540 BC
Pentheus	Second-in-Command	526 BC
Dienekes	Hoplite	503 BC
Dilios	Hoplite	505 BC
Stelios	Hoplite	510 BC
Dilios	Hoplite	505 BC

Spartan Name	Weapon (And Weaponsmith)
Leonidas	Spear (Spiridon of Sparta)
Leonidas	Javelin (Kostas of Amyklae)
Leonidas	Shortsword (Theofilos of Helos)
Pentheus	Spear (Spiridon of Sparta)
Pentheus	Shortsword (Theofilos of Helos)
Dienekes	Spear (Spiridon of Sparta)
Dienekes	Dagger (Theofilos of Helos)
Dilios	Spear (Spiridon of Sparta)
Stelios	Spear (Spiridon of Sparta)
Dilios	Spear (Spiridon of Sparta)

So, we need to split it further. **BUT** each row must be unique .(Dilios)

Spartan Name	Role	Born
Leonidas	Commander	540 BC
Pentheus	Second-in-Command	526 BC
Dienekes	Hoplite	503 BC
Dilios	Hoplite	505 BC
Stelios	Hoplite	510 BC
Dilios	Hoplite	505 BC

Spartan Name	Weapon	Weaponsmith	Smithy Location
Leonidas	Spear	Spiridon	Sparta
Leonidas	Javelin	Kostas	Amyklae
Leonidas	Shortsword	Theofilos	Helos
Pentheus	Spear	Spiridon	Sparta
Pentheus	Shortsword	Theofilos	Helos
Dienekes	Spear	Spiridon	Sparta
Dienekes	Dagger	Theofilos	Helos
Dilios	Spear	Spiridon	Sparta
Stelios	Spear	Spiridon	Sparta
Dilios	Spear	Spiridon	Sparta

So we assign each row a unique Primary KEY:

ID	Name	Role	Born
1	Leonidas	Commander	540 BC
2	Pentheus	Second-in-Command	526 BC
3	Dienekes	Hoplite	503 BC
4	Dilios	Hoplite	505 BC
5	Stelios	Hoplite	510 BC
6	Dilios	Hoplite	505 BC

Spartan ID	Weapon	Weaponsmith	Smithy Location
1	Spear	Spiridon	Sparta
1	Javelin	Kostas	Amyklae
1	Shortsword	Theofilos	Helos
2	Spear	Spiridon	Sparta
2	Shortsword	Theofilos	Helos
3	Spear	Spiridon	Sparta
3	Dagger	Theofilos	Helos
4	Spear	Spiridon	Sparta
5	Spear	Spiridon	Sparta
6	Spear	Spiridon	Sparta

ALL OUR DATA IS NOW ATOMIC. SO IT'S IN 1ST NORMAL FORM.

Second Normal Form (2NF)

- Already in 1st Normal Form
- All non-key attributes must functionally depend upon the full PRIMARY KEY

Meaning DEPENDENCE: Value of 1 Attribute(Column) is determined by another attribute(Column).

For example:

So the Number of Legs is functionally dependent on the type of animal

Example:

At the moment, the Primary Key is composite

Spartan ID	Weapon	Weaponsmith	Smithy Location
1	Spear	Spiridon	Sparta
1	Javelin	Kostas	Amyklae
1	Shortsword	Theofilos	Helos
2	Spear	Spiridon	Sparta
2	Shortsword	Theofilos	Helos
3	Spear	Spiridon	Sparta
3	Dagger	Theofilos	Helos
4	Spear	Spiridon	Sparta
5	Spear	Spiridon	Sparta
6	Spear	Spiridon	Sparta

PRIMARY KEY - SpartanID, Weapon. both Weaponsmith and Smithy location must be functionally dependent upon the FULL PRIMARY KEY. which in our case Weaponsmith data is not.

Video Player	Weapon	Weaponsmith	Smithy Location
1	Spear	Spiridon	Sparta
1	Javelin	Kostas	Amyklae
1	Shortsword	Theofilos	Helos
2	Spear	Spiridon	Sparta
2	Shortsword	Theofilos	Helos
3	Spear	Spiridon	Sparta
3	Dagger	Theofilos	Helos
4	Spear	Spiridon	Sparta
5	Spear	Spiridon	Sparta
6	Spear	Spiridon	Sparta

So we split these two tables

Spartan ID	Weapon
1	Spear
1	Javelin
1	Shortsword
2	Spear
2	Shortsword
3	Spear
3	Dagger
4	Spear
5	Spear
6	Spear

Weapon	Weaponsmith	Smithy Location
Spear	Spiridon	Sparta
Javelin	Kostas	Amyklae
Shortsword	Theofilos	Helos
Spear	Spiridon	Sparta
Shortsword	Theofilos	Helos
Spear	Spiridon	Sparta
Dagger	Theofilos	Helos
Spear	Spiridon	Sparta
Spear	Spiridon	Sparta
Spear	Spiridon	Sparta

And remove duplicates

Spartan ID	Weapon
1	Spear
1	Javelin
1	Shortsword
2	Spear
2	Shortsword
3	Spear
3	Dagger
4	Spear
5	Spear
6	Spear

Weapon	Weaponsmith	Smithy Location
Spear	Spiridon	Sparta
Javelin	Kostas	Amyklae
Shortsword	Theofilos	Helos
Dagger	Theofilos	Helos

Video Player

Spartan ID	Weapon ID
1	1
1	2
1	3
2	1
2	3
3	1
3	4
4	1
5	1
6	1

ID	Name	Role	Born
1	Leonidas	Commander	540 BC
2	Pentheus	Second-in-Command	526 BC
3	Dienekes	Hoplite	503 BC
4	Dilios	Hoplite	505 BC
5	Stelios	Hoplite	510 BC
6	Dilios	Hoplite	505 BC

Weapon ID	Weapon	Weaponsmith	Smithy Location
1	Spear	Spiridon	Sparta
2	Javelin	Kostas	Amyklae
3	Shortsword	Theofilos	Helos
4	Dagger	Theofilos	Helos

ALL OUR DATA IS NOW IN 1NF AND NON-KEY ATTRIBUTES ARE FUNCTIONALLY DEPENDENT ON THE PRIMARY KEY.

THIRD NORMAL FORM (3NF)

- Already in 2nd Normal Form

- There are no transitive dependencies

For example: the presenter and the TV shows does depend on Person and person's TV preferences but it's a transitive dependencies. the presenter is determined by what the show is which in turn is determined by person's identity. There is no inherent relationship between present and person. It's determined by IDENTITY of that TV Programme

Video Player

Person ← Favourite TV Show ← Presenter

ID PRIMARY KEY, Name is dependent on the identity of the Spartan, so is the role and so is when they were born => THIRD NORMAL FORM, no transitive dependency

Spartan ID	Weapon ID
1	1
1	2
1	3
2	1
2	3
3	1
3	4
4	1
5	1
6	1

ID	Name	Role	Born
1	Leonidas	Commander	540 BC
2	Pentheus	Second-in-Command	526 BC
3	Dienekes	Hoplite	503 BC
4	Dilios	Hoplite	505 BC
5	Stelios	Hoplite	510 BC
6	Dilios	Hoplite	505 BC

Weapon ID	Weapon	Weaponsmith	Smithy Location
1	Spear	Spiridon	Sparta
2	Javelin	Kostas	Amyklae
3	Shortsword	Theofilos	Helos
4	Dagger	Theofilos	Helos

This table already in 3RD NORMAL FORM: PRIMARY KEY is COMPOSITE KEY of BOTH COLUMNS

Spartan ID	Weapon ID
1	1
1	2
1	3
2	1
2	3
3	1
3	4
4	1
5	1
6	1

ID	Name	Role	Born
1	Leonidas	Commander	540 BC
2	Pentheus	Second-in-Command	526 BC
3	Dienekes	Hoplite	503 BC
4	Dilios	Hoplite	505 BC
5	Stelios	Hoplite	510 BC
6	Dilios	Hoplite	505 BC

Weapon ID	Weapon	Weaponsmith	Smithy Location
1	Spear	Spiridon	Sparta
2	Javelin	Kostas	Amyklae
3	Shortsword	Theofilos	Helos
4	Dagger	Theofilos	Helos

SMITHY LOCATION depends on Weaponsmith not the Weapon.

ID	Name	Role	Born
2	Pentheus	Second-in-Command	526 BC
3	Dienekes	Hoplite	503 BC
4	Dilios	Hoplite	505 BC
5	Stelios	Hoplite	510 BC
6	Dilios	Hoplite	505 BC

Video Player

Weapon ID	Weapon	Weaponsmith	Smithy Location
1	Spear	Spiridon	Sparta
2	Javelin	Kostas	Amyklae
3	Shortsword	Theofilos	Helos
4	Dagger	Theofilos	Helos

So we need to add to a different table:

Third Normal Form (3NF)

Video Player ady in Second Normal Form
- There are no transitive dependencies

ID	Name	Role	Born
1	Leonidas	Commander	540 BC
2	Pentheus	Second-in-Command	526 BC
3	Dienekes	Hoplite	503 BC
4	Dilios	Hoplite	505 BC
5	Stelios	Hoplite	510 BC
6	Dilios	Hoplite	505 BC

Spartan ID	Weapon ID
1	1
1	2
1	3
2	1
2	3
3	1
3	4
4	1
5	1
6	1

Weapon ID	Weapon	Smith ID
1	Spear	1
2	Javelin	2
3	Shortsword	3
4	Dagger	3

Smith ID	Weaponsmith	Smithy Location
1	Spiridon	Sparta
2	Kostas	Amyklae
3	Theofilos	Helos

ALL OUR DATA IS NOW IN 3NF AND NON-KEY ATTRIBUTES AND THERE ARE NO TRANSITIVE DEPENDENCIES

ASK