Database normalization

Database normalization
- What and why?

- It’s a technique of structuring the database such that:
  - There is no redundant data
  - Columns and tables are organized such that data dependencies make sense.

- This way we avoid insertion, update, and deletion anomalies.

Database normalization
- How?

- By following a set of rules called normal forms.

- The most important normal forms are:
  - The first normal form;
  - The second normal form;
  - The third normal form;

Example

task_id	title	Assigned_user	email	phone
1	Become billionaire	Scrooge	scrooge@gmail.com	53331
2	Add feature x to program	Donald Duck	donald@gmail.com	53337
3	Do node homework	Donald Duck	donald@gmail.com	53337
4	Discuss DB structure	Donald Duck	donald@gmail.com	53337

First normal form

- Rule 1:
  - Each column on the table should not contain lists of values, it should contain only single values. E.g, assigned_user should contain only one user and not a list of users.

- Rule 2:
  - Each column should store only values of the same type, e.g. if you have a column created_date you should store only dates there and not names or something else.

First normal form

- Rule 3:
  - In a given table, each column should have a unique name, i.e. no repeated column names in the same table.
- Rule 4:
  - The order in which you store your data shouldn’t matter.

Second normal form

- Rule:
  - There should be no partial dependency.

- What is dependency:

user_id	name	email	phone
1	Donald Duck	donald@gmail.com	333-111-222
2	Pavel Ble	pavel@hotmail.com	312-246-346

  - If we select a user by user_id we get the name, email, and phone for that user_id
    - Name, email and phone are dependent on user_id

Second normal form

- Partial dependency can occur when you have a primary composite key (a primary key composed by two columns) and a given column depends only on part of the composite primary key.

Second normal form

user_id	name	email
1	Donald Duck	donald@gmail.com
2	Pavel Ble	pavel@hotmail.com

task_id	title
1	Add feature x to program
2	Do node homework
3	Discuss DB structure

Role depends only on the user and should be part of the user table

user_id	task_id	status	role
1	1	Done	Developer
2	2	Started	Student
1	3	Not started	Developer

Composite primary key

Third normal form

- Rule:
  - There should be no transitive dependency.

student_id	subject_id	marks	task_type	task_weight_on_performance_review
1	1	10	Weekly assignment	0.1
1	2	8	Project	0.4
2	1	9	Exam	0.6

Depends only on task_type and not on (student_id, subject_id)

- Solution: create an exam table.