OPTIMIZATION

REDUNDANCY / ANOMALIES / INCONSISTENCIES

Same piece of data is held in several different places
Caused by poor database construction
Entities and relationships not identified correctly
Avoided by normalization

NORMALIZATION

The process of efficiently organizing data in a database.
Creating well designed databases.
Without normalization, it becomes difficult to handle and update the database, without facing data loss.
Insert, update and delete anomalies are very frequent if a database is not normalized.
There are two goals of the normalization process:

Eliminate redundant data (Storing the same data in more than one table)
Ensure data dependencies make sense (Storing only related data in a table)

Reduce the amount of space a database consumes and ensure that data is logically stored.

1. First Normal Form
   - Look at data to be stored
   - Organize data into columns
   - Define what type of data each column contains
   - Put related columns into their own table
   - A table should have a primary key column that distinguishes a row uniquely
   - All columns must contain only a single value
   - No repeating groups of data

ID	NAME	AGE	ADDRESS	ORDERS
100	Sachin	36	Lower West Side	Cannon XL-200
100	Sachin	36	Lower West Side	Battery XL-200
100	Sachin	36	Lower West Side	Tripod Large

   - Single row instead of three and only orders change so they should be in a separate table
2. Second Normal Form
   - Database is in first normal form
   - All non-key attributes are fully functional dependent on the primary key

customerID	storeID	purchaseLocation
1	1	Los Angeles
1	3	San Francisco
2	1	Los Angeles
3	2	New York
4	3	San Francisco

   - Purchase location is only dependent on store id and should be in separate table
3. Third Normal Form
   - All of the columns are fully dependent upon the primary key
   - No transitive dependencies

customerID	customerName	city	zip
11	George	Roskilde	4000
12	Igor	Aalborg	9100
13	Benjamin	Roskilde	4000

   - Move cities to a new table since it is dependent on zip which is dependent on id