

Class of 2/18/2019

Terminology

- entity - an object or a thing (person, team, product, etc.)

Design Principles

- a table models a single entity, and an entity is modeled by a single table
- each field in a table is assigned a primitive data type (i.e a string, a number, a date, etc. No arrays or structs)
- Each field in a table is assigned a precise data type.
 - A field of type
- Every table contains a single Primary Key
 - NOTE: The primary key does not *necessarily* have to be restricted to a single field! It can be a combination of multiple fields, but it has to be *minimal*.
- Each child table contains a Foreign Key.
- Each m:n relationship is modeled with a junction table.

Data Anomalies

- anomaly - a flaw in the database that is caused by bad database design
 - insert anomaly - not being able to insert an entity into a table without inserting another entity.
 - *Ex:* For a Student table in a database containing only this table,

id (PK)	fname	lname	class_id	class_name
1	John	Appleseed	2	Intro to Economics
2	Greg	Dappleseed	1	Intro to Humanities
3	Megan	Ooo	2	Intro to Economics

Note we cannot add a new class without first adding a student.

- update anomaly - the same attributes are stored redundantly in another table
 - *Ex.* For the above table, note we cannot update the class name of any class without updating all classes with the same ID in the table
- delete anomaly - entity types are lost due to the deletion of other entity types
 - *Ex.* For the above table, note that we cannot remove classes without removing all students associated with that class

Normalization

- normalization - forms of a database that represent its organization
 - Only need to know up to **3NF!**
- First Normal Form (1NF) - a database is in 1NF when all tables in the database contain fields with only scalar values
 - No one cell in a field should contain multiple values, i.e this Employee table is *not* in 1NF:

id (PK)	name	managers
1	Jeff	
2	Bob	1
3	Rank	1, 2

Which field makes this table not in 1NF? *Hint: managers*

- Second Normal Form (2NF) - 1NF + all non-key attributes (non Primary Key attributes) must be functionally determined by the **entire** primary key.
 - This table is *not* in 2NF:

state (PK)	city (PK)	population	governor
Texas	Houston	2,313,000	Greg Abbott
Arkansas	Little Rock	198,606	Asa Hutchinson
Texas	Austin	950,715	Greg Abbott

Which field makes this table not in 2NF? *Hint: governor only depends on part of the primary key*

- Third Normal Form (3NF) - 2NF + all non-key attributes must be functionally determined by *only* the primary key
 - From the Student table above, note that class name does not depend on the primary key of the table.
- For $Y > X$, if a table is not in X normal form, then the table is not in Y normal form either.