Databases and SQL

Cameron Zabriskie

BYU Statistics

September 17, 2020

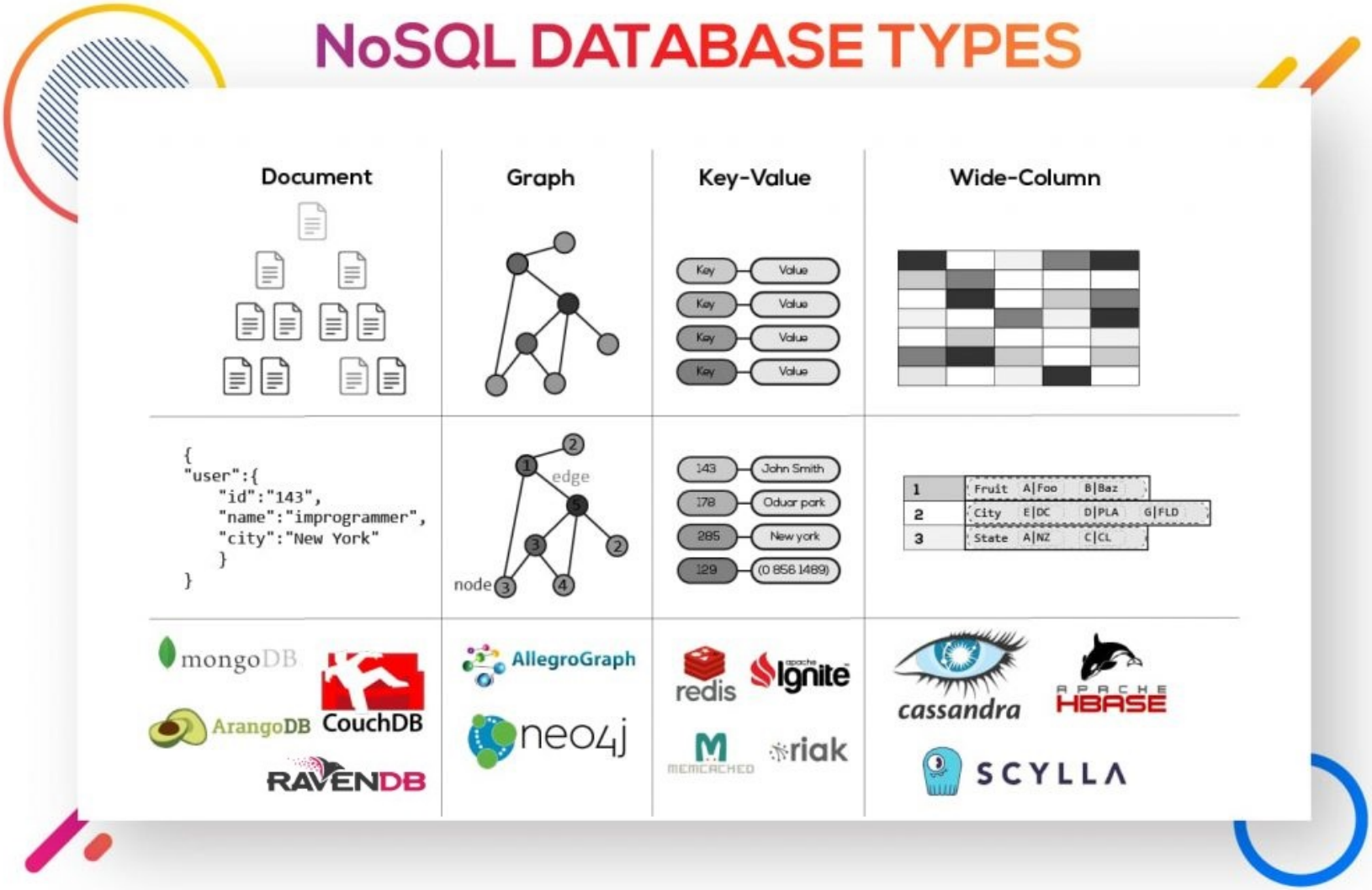
Database

A database is simply a structured set of data

- Non-relational Databases
- Relational Databases

Non-Relational Database

- Non-relational database (noSQL Database)- Data is stored in a way that does not involve a tabular format



Relational Database

- Relational database (SQL Database) - data is stored as columns and rows within tables. A row is a record and a column is a field with some data type in it.

Name	FName	City	Age	Salary
Smith	John	3	35	$280
Doe	Jane	1	28	$325
Brown	Scott	3	41	$265
Howard	Shemp	4	48	$359
Taylor	Tom	2	22	$250

Database Management System (DMS)

A database management system is software that allows the creation and modification of a database.
Some popular relational database management systems (RDMS) are:

- MySQL
- PostgreSQL
- Oracle DB
- SQL Server
- SQLite

Database Objects

A database can contain many different kinds of database objects

- Table
- View
- Index
- Function
- Procedure
- Parameter
- Trigger

Table

Every record in a table should be unique and be identified through a primary key

LastName	FirstName	Department	Supervisor Number	Tenure	SSN	Address	Primary Number	Sex	Race
Francis	Bill	HR	104	1221	146-86-5788	334 Diamond Street	570-495-5890	M	Black
McMahon	Geraldine	Sales	291	342	574-17-2592	2562 Blue Spruce Lane	570-407-1805		Asian
Hanks	Tom	HR	104	62	617-44-0493	3346 Matthews Street	623-582-2306	M	White
Smith	Alexis	CX	45	954	436-80-1670	1234 Layman Court	602-309-9992	F	White
Geuix	Penny	CX	45	423	287-58-2138	304 High Meadow Lane	775-574-8229	F	Hispanic

Table

Every record in a table should be unique and is identifiable through a "primary key"

Employee Number	LastName	FirstName	Department	Supervisor Number	Tenure	SSN	Address	Primary Number	Sex	Race
55	Francis	Bill	HR	104	1221	146-86-5788	334 Diamond Street	570-495-5890	M	Black
212	McMahon	Geraldine	Sales	291	342	574-17-2592	2562 Blue Spruce Lane	570-407-1805		Asian
421	Hanks	Tom	HR	104	62	617-44-0493	3346 Matthews Street	623-582-2306	M	White
86	Smith	Alexis	CX	45	954	436-80-1670	1234 Layman Court	602-309-9992	F	White
241	Geuix	Penny	CX	45	423	287-58-2138	304 High Meadow Lane	775-574-8229	F	Hispanic

Table

Tables may contain "foreign keys" to join to other tables

Employee Number	LastName	FirstName	Department	Supervisor Number	Tenure	SSN	Address	Primary Number	Sex	Race
55	Francis	Bill	HR	104	1221	146-86-5788	334 Diamond Street	570-495-5890	M	Black
212	McMahon	Geraldine	Sales	291	342	574-17-2592	2562 Blue Spruce Lane	570-407-1805		Asian
421	Hanks	Tom	HR	104	62	617-44-0493	3346 Matthews Street	623-582-2306	M	White
86	Smith	Alexis	CX	45	954	436-80-1670	1234 Layman Court	602-309-9992	F	White
241	Geuix	Penny	CX	45	423	287-58-2138	304 High Meadow Lane	775-574-8229	F	Hispanic

View

A view is a table defined by a query

LastName	FirstName	Department	Supervisor Number	Tenure	SSN	Address	Primary Number	Sex	Race
Francis	Bill	HR	104	1221	146-86-5788	334 Diamond Street	570-495-5890	M	Black
McMahon	Geraldine	Sales	291	342	574-17-2592	2562 Blue Spruce Lane	570-407-1805		Asian
Hanks	Tom	HR	104	62	617-44-0493	3346 Matthews Street	623-582-2306	M	White
Smith	Alexis	CX	45	954	436-80-1670	1234 Layman Court	602-309-9992	F	White
Geuix	Penny	CX	45	423	287-58-2138	304 High Meadow Lane	775-574-8229	F	Hispanic

SELECT FirstName, LastName, Department, Supervisor, Tenure
FROM employees;

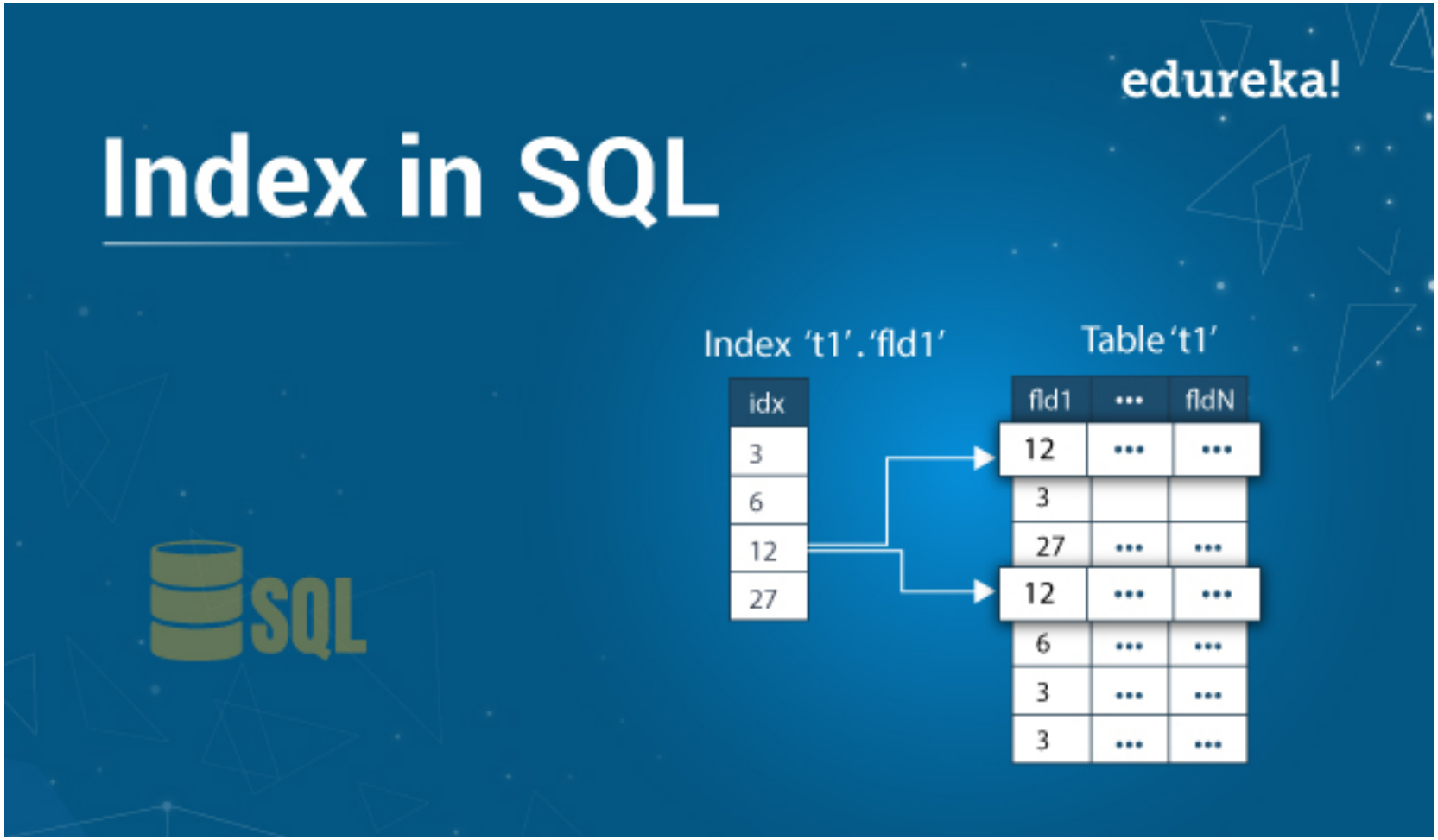
View

- allow users to see data without having access to an entire table
- hide complicated business logic that exists in a query

LastName	FirstName	Department	Supervisor Number	Tenure
Francis	Bill	HR	104	1221
McMahon	Geraldine	Sales	291	342
Hanks	Tom	HR	104	62
Smith	Alexis	CX	45	954
Geuix	Penny	CX	45	423

Index

An index is used to speed up the retrieval of data from a table. An index does this by caching searches on specific columns. This allows for the quick retrieval of a subset of interest.



Function

Functions can be stored and then used in your queries

Function: f(x) = x / 365.0

LastName	FirstName	Department	Supervisor Number	Tenure
Francis	Bill	HR	104	1221
McMahon	Geraldine	Sales	291	342
Hanks	Tom	HR	104	62
Smith	Alexis	CX	45	954
Geuix	Penny	CX	45	423

SELECT LastName, FirstName, Tenure, convert_months(Tenure) AS year_tenure
FROM employees

Function

Functions can be stored and then used in your queries

Function: $f(x) = x / 365.0$

LastName	FirstName	Department	Supervisor Number	Tenure
Francis	Bill	HR	104	3.34
McMahon	Geraldine	Sales	291	0.94
Hanks	Tom	HR	104	0.17
Smith	Alexis	CX	45	2.61
Geuix	Penny	CX	45	1.16

Procedure or Stored Procedure

Procedures are stored queries in the RDMS.

There are many reasons why you might want to store a query as a procedure.

- Extract, transform, load operation
- A query that you run often
- Used to allow easy data retrieval when using other software

Trigger

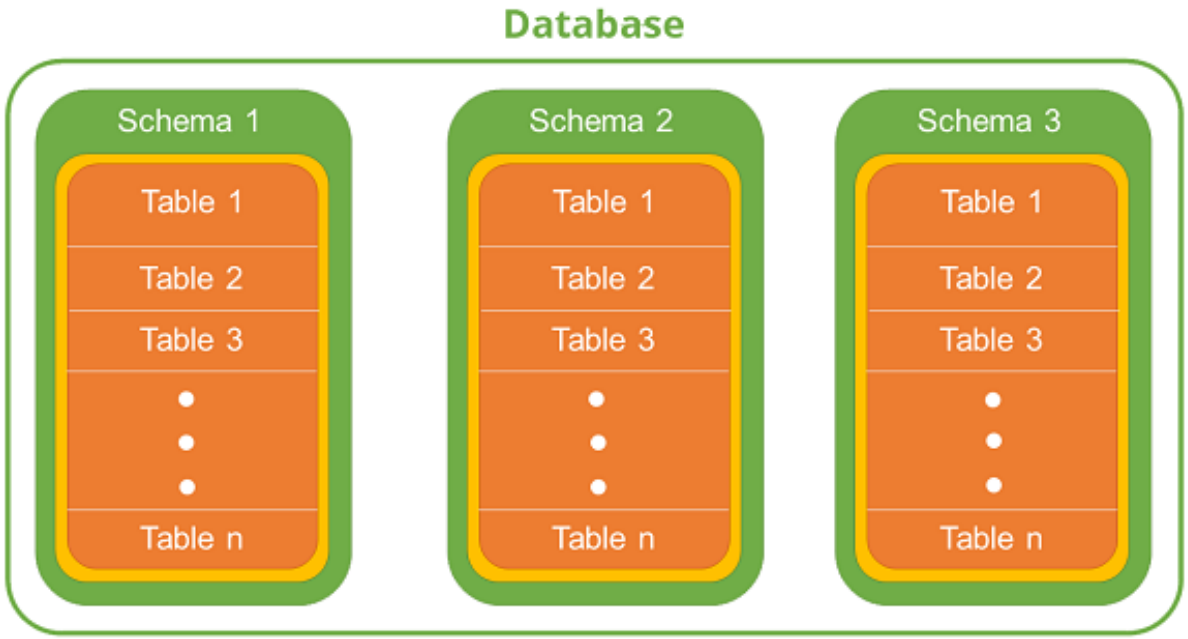
A trigger is a special stored procedure that executes when a condition is met.

- Send out an email when a table is updated
- Add to other tables when data is inserted into a table

When there are new records in the employee table, send records to the manager of the department.

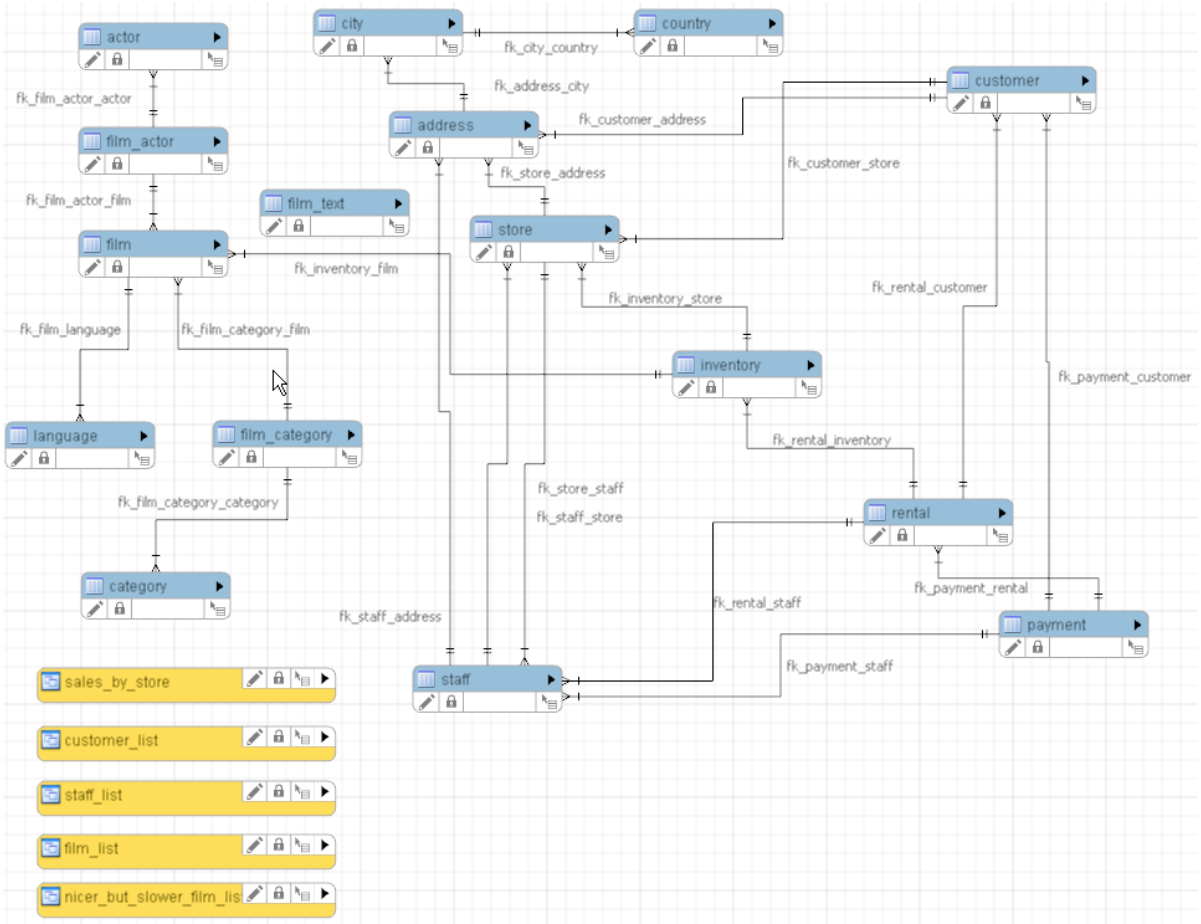
Schemas

Within a database you might have many schemas, which are just a collection of database objects that belong together

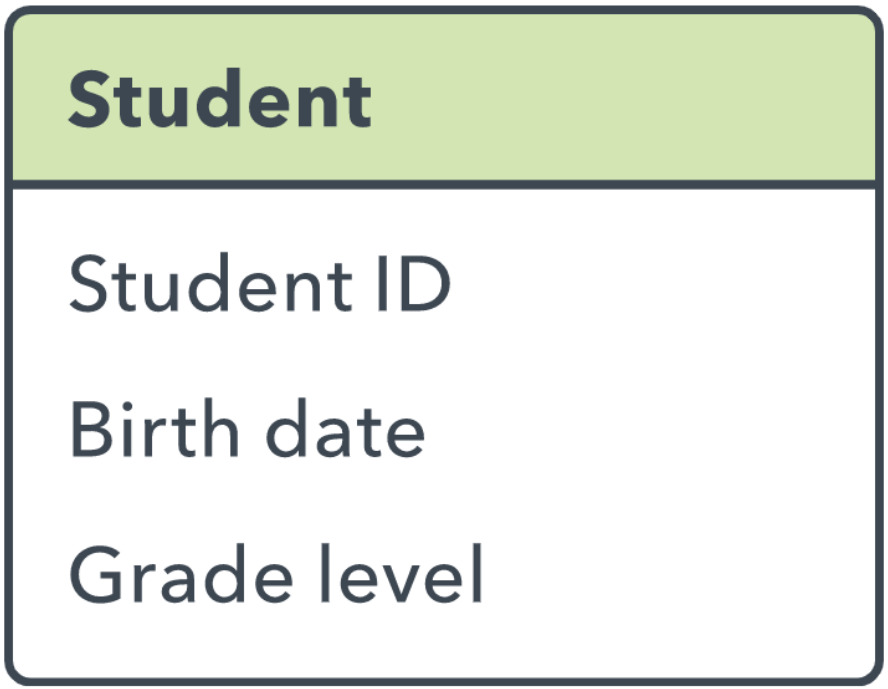


Reading an Entity Relationship Diagram (ERD)

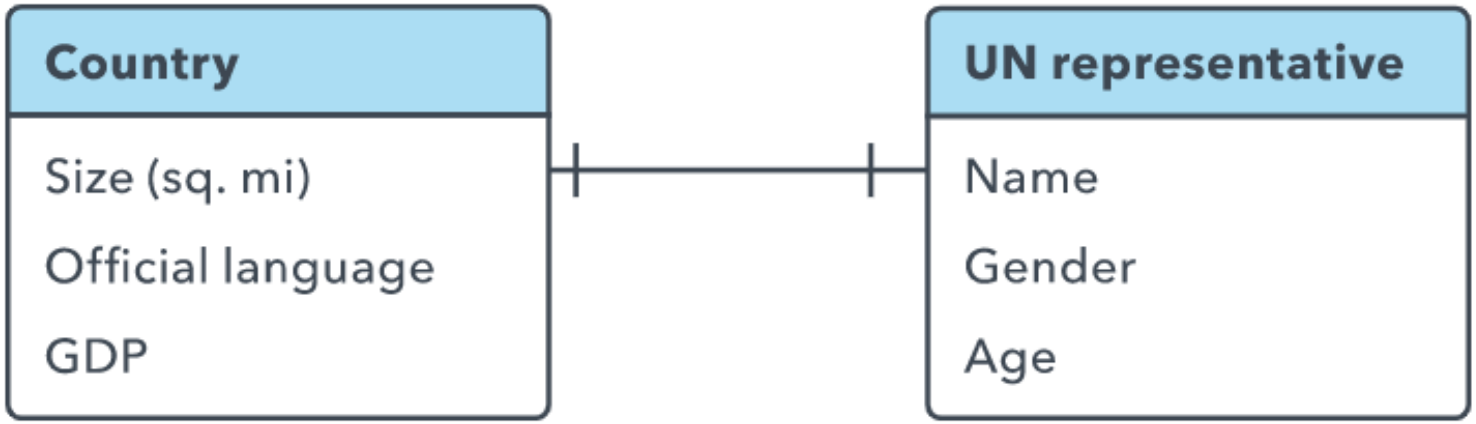
In some software, you are able to create a database diagram that shows the relationship of the objects in a database schema. They are pretty simple to read, you just need to know what the symbols mean.



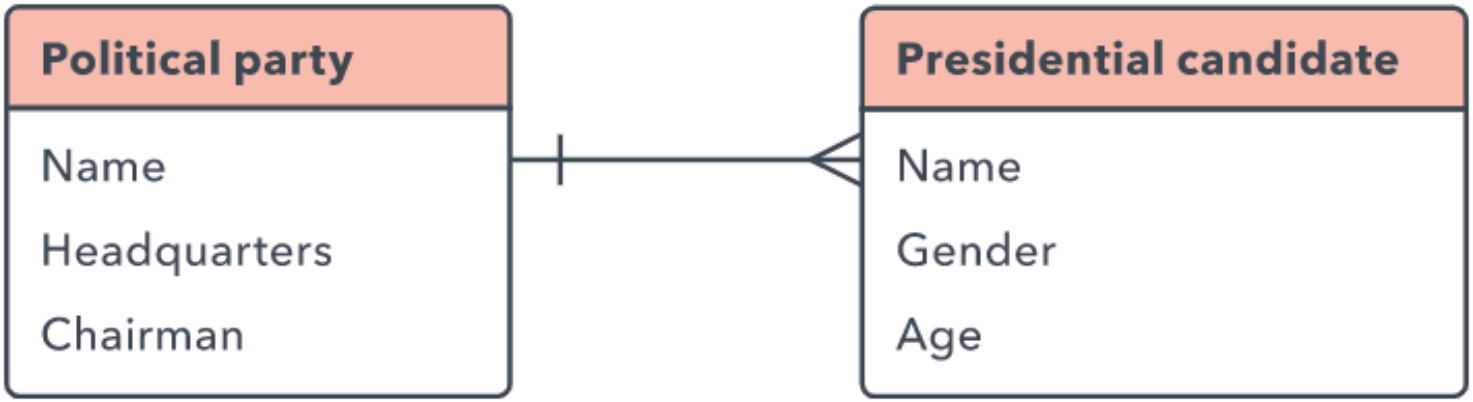
Table



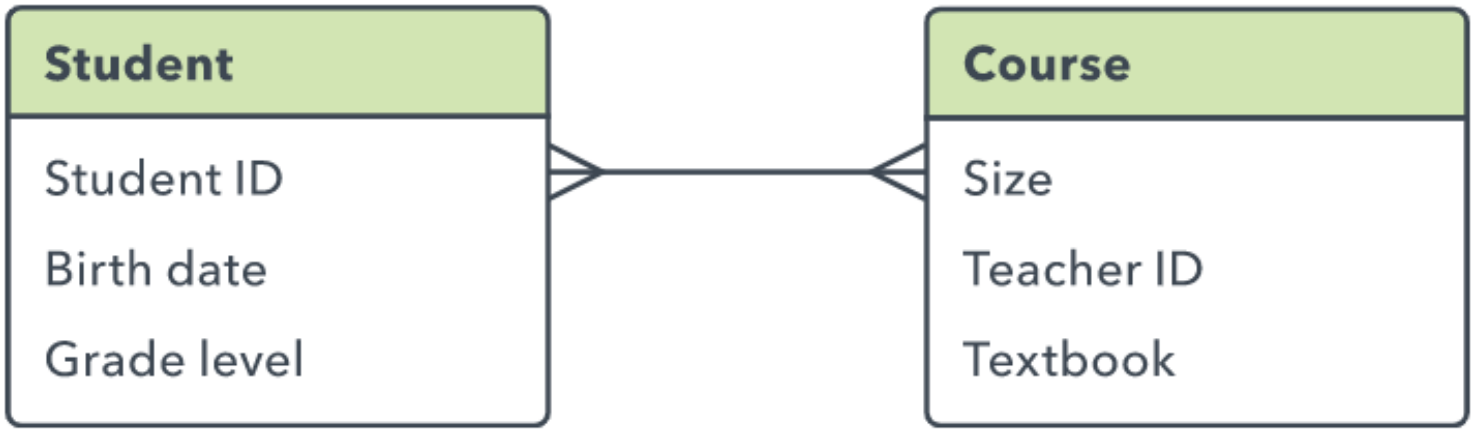
One-to-One Relationship



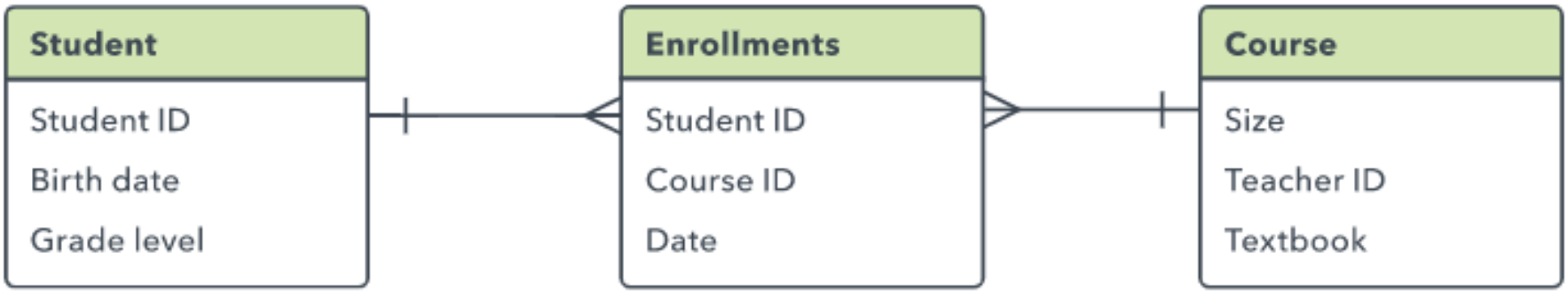
One-to-Many Relationship



Many-to-Many Relationship

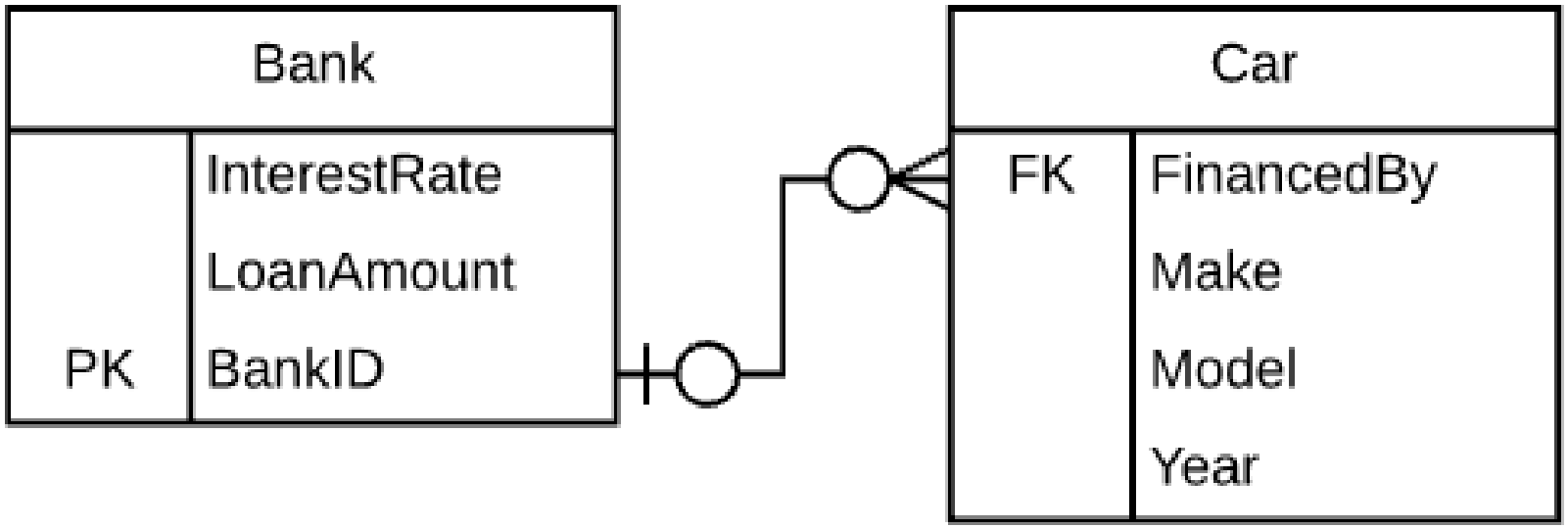


Many-to-Many Relationship



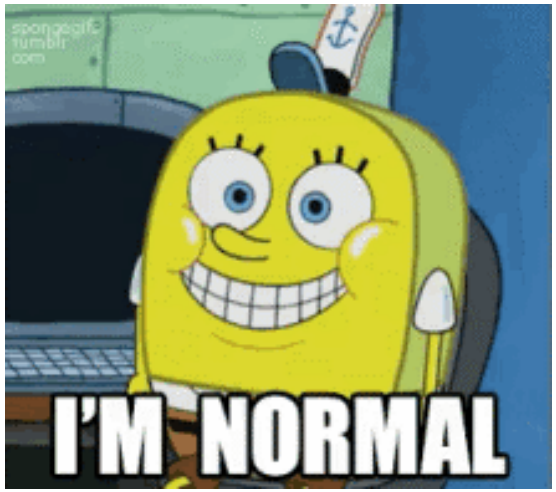
A Quick Example

Many ERDs specify the primary key, foreign key, and data types of the fields in each table



Database Normalization

A table can be in different levels of normalization based on if they follow certain rules. Although there might be a good case to not follow these rules, typically the more normalized the database the more organized and easier to access the table will be



First Normal Form 1NF

For a table to be in first normal form, the following conditions must be met.

- Each entry in a table must only contain one value
- tables cannot have columns that are identical in nature

First Normal Form 1NF

Is this table in 1NF?

ProductID	Color	Price
1	brown, yellow	$15
2	red, green	$13
3	blue, orange	$11

Second Normal Form 2NF

For a table to be in second normal form, the following conditions must be met.

- The table must be in 1NF
- All non-key columns must be dependent on the primary key

Second Normal Form 2NF

Is this table in 2NF?

Order Number (PK)	Product ID (PK)	Product Name	Product Type
1	7	Xbox One	Electronics
2	8	Custom T-Shirt	Apparel

Third Normal Form 3NF

"Every non-key attribute must provide a fact about the key, the whole key, and nothing but the key so help me Codd."

For a table to be in third normal form, the following conditions must be met.

- The table must be in 2NF
- Every non-key column must be independent of every other non-key column

Third Normal Form 3NF

Is this table in 3NF?

Order	Price	Tax
14325	$40.99	$2.05
14326	$13.73	$.69
14327	$24.15	$1.21

Other Normal Forms

Other forms of normalization forms exist, but the first three are the most common

- Boyce-Codd normal form
- 4NF, 5NF, and 6NF
- domain-key normal form