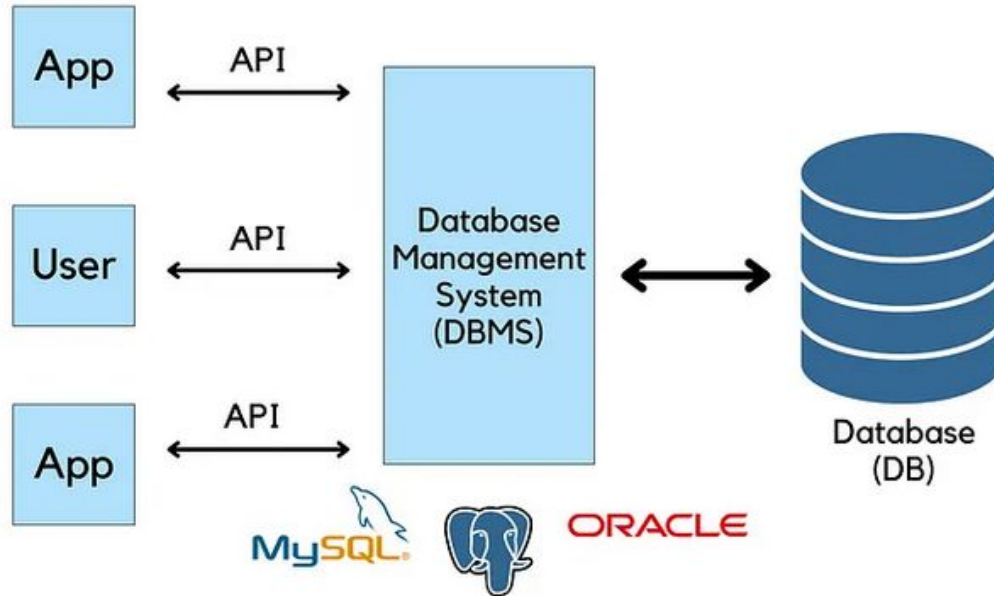# Databases

What is a database?

A database is an organized collection of structured information, or data, typically stored electronically in a computer system. It is usually controlled by a database management system (DBMS) (Source: Oracle)

Popular DBMS: Oracle, MySQL, SQL Server, and PostgreSQL.

Differences between an spreadsheet and a database

Databases allow complex data manipulation, while spreadsheets aren't meant for users who need much data manipulation.

Databases allow multiple users to quickly access and query the data, while spreadsheets were designed only for a single user or a small number of users.

Databases are designed to store larger collections of data, while spreadsheets have a limitation.

Types of Databases

- Relational databases (SQL Database)

- Non-Relational databases (NoSQL Database)

Non-Relational databases (NoSQL Database)

- Stores data without tables, rows, or keys. It stores data in a non-tabular form.

- Helps satisfy specific requirements of the type of data being stored.

- It is like a collection of documents.

- A document can contain a lot of detailed information about a customer.

- Each customer can have different types of information (stored in the same document.)

Non-Relational databases (NoSQL Database)

Key	Document

1001	<pre>{
	    "CustomerID": 99,
	    "OrderItems": [
	        { "ProductID": 2010,
	          "Quantity": 2,
	          "Cost": 520
	        },
	        { "ProductID": 4365,
	          "Quantity": 1,
	          "Cost": 18
	        }],
	          "OrderDate": "04/01/2017"
	}</pre>

1002	<pre>{
	    "CustomerID": 220,
	    "OrderItems": [
	        { "ProductID": 1285,
	          "Quantity": 1,
	          "Cost": 120
	        }],
	          "OrderDate": "05/08/2017"
	}</pre>

Relational databases - Tables

Relational databases use tables to store data.

A **table** is a collection of related data entries and contains rows and columns to store data.

A table represents real-world objects.

Relational databases - Records, Fields

- Tables have rows (**records**) which have cells and each cell has a value.

- Tables contain one or more columns (**attributes, fields**). All the cells in a column contain the same type of information.

Table: employees

id	first_name	last_name	email
1	Peter	Smith	peter@gmail.com
2	Jessica	Novillo	jessica@gmail.com
3	Sophie	Anderson	sophie@gmail.com

Relational databases - Row Properties

- There should not be two identical records.
- All records in a table must have the same format and the same number of entries.
- The order of the records is irrelevant. They are identified by their content, not by their position.

Relational databases - Column Properties

- Null values are permitted for the attributes.
- Default values can be specified for an attribute automatically inserted if no other value is specified for an attribute.
- Attributes could act as keys.

Data Integrity

Entity integrity: It specifies that there should be no duplicate rows in a table.

Domain integrity: It enforces valid entries for a given column by restricting the type, the format, or the range of values.

Referential integrity specifies that rows cannot be deleted, which are used by other records.
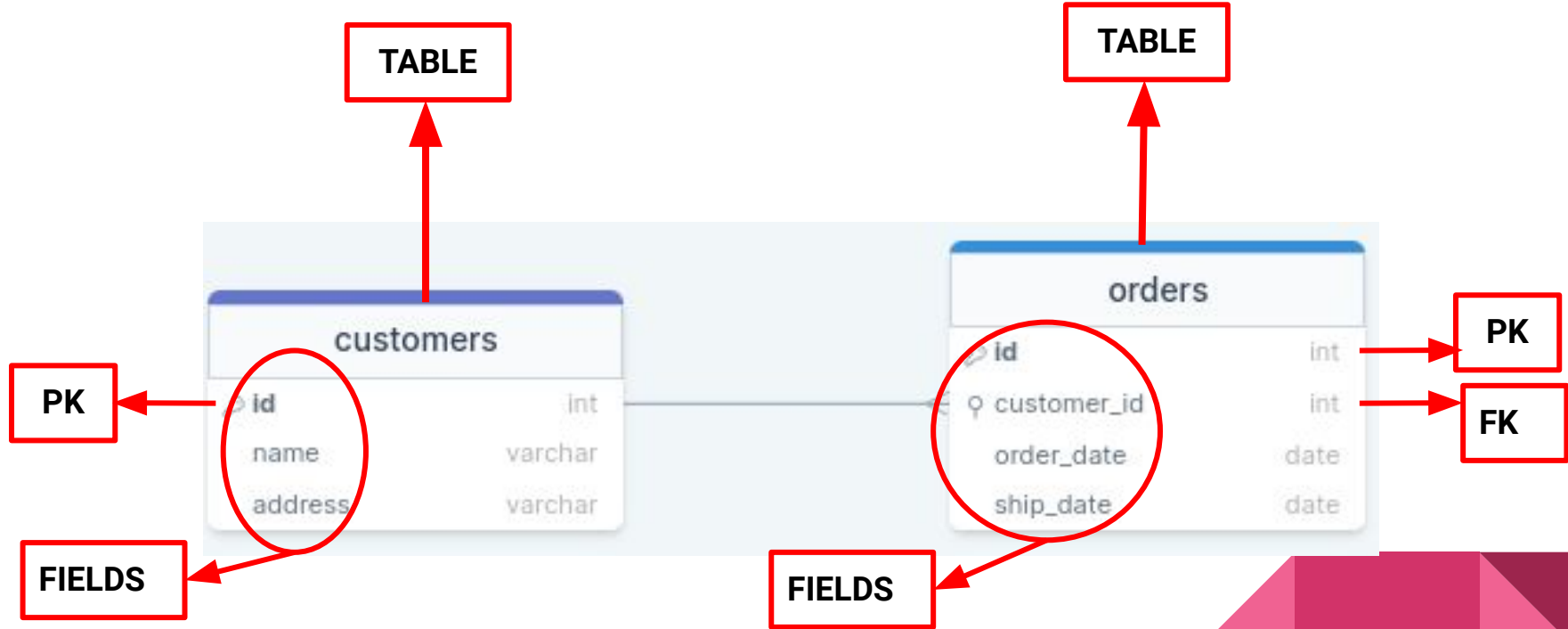
Keys

- The tables have keys which link information from different tables.
  - **Primary Key:** Unique value in a table.
  - **Foreign Key:** When a primary key is added to a record in another table it is know as a foreign key in the second table.

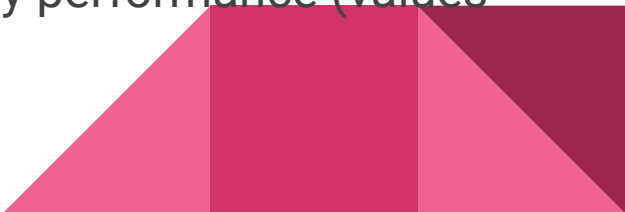# Relational databases (SQL Database)

EMPLOYEES

PRIMARY KEY →

employee_ID	first_name	last_name	department
1	Anna	Wrigley	Sales
2	Kristine	Lambeau	Design
3	Thea	Comiskey	Sales

SALES

customer_ID	order_total	salesperson_ID
001	$422.01	3
002	$899.76	1
003	$560.00	3

← FOREIGN KEY

Database Index

- Table keys are stored in indexes.
- Each time a new row with a unique key is added, the index is automatically updated.
- Indexes are a powerful tool used in the background of a database to speed up querying.
- Indexes contain all the necessary information needed to access items quickly and efficiently.
- Indexes serve as lookup tables to efficiently store data for quicker retrieval.
- There are non-unique indexes used to improve query performance (values that are used frequently).

Relational databases (SQL Database)

- The connection between primary and foreign keys creates a relationship between records within both tables.

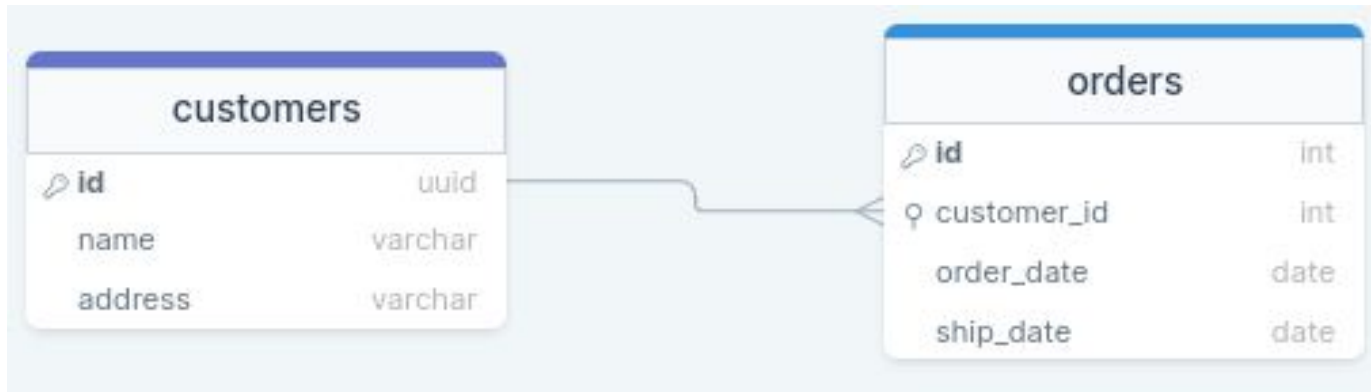- Some popular relational database management systems (RDBMS) are Oracle, MySQL, SQL Server, and PostgreSQL.

Database relationships

One -to-one: Both tables can have only one record on each side of the relationship.

Database relationships

One-to-many: The primary key table contains only one record that relates to none, one, or many records in the related table.

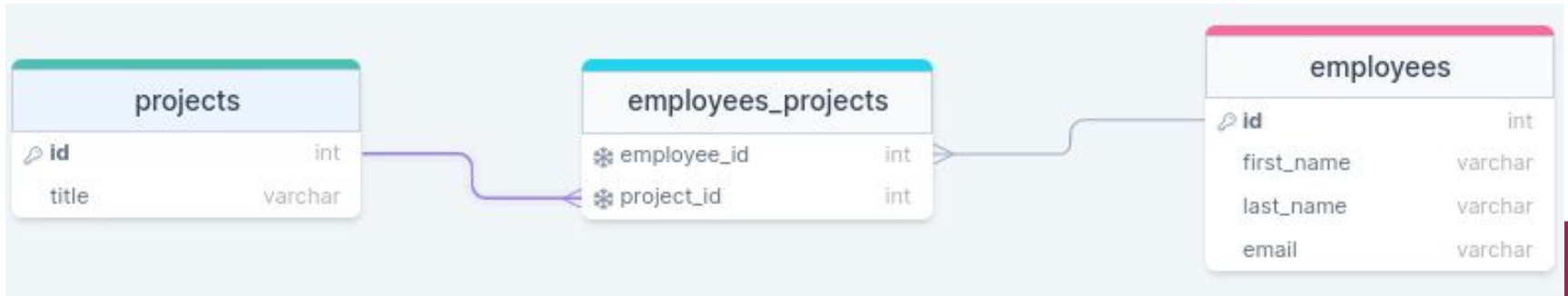# Database relationships

Many-to-many

How do we represent this relationship?

Database relationships

Many-to-many: Each record in both tables can relate to none or any number of records in the other table. These relationships require a third table, called an **associate or linking table,** because relational systems cannot directly accommodate the relationship.
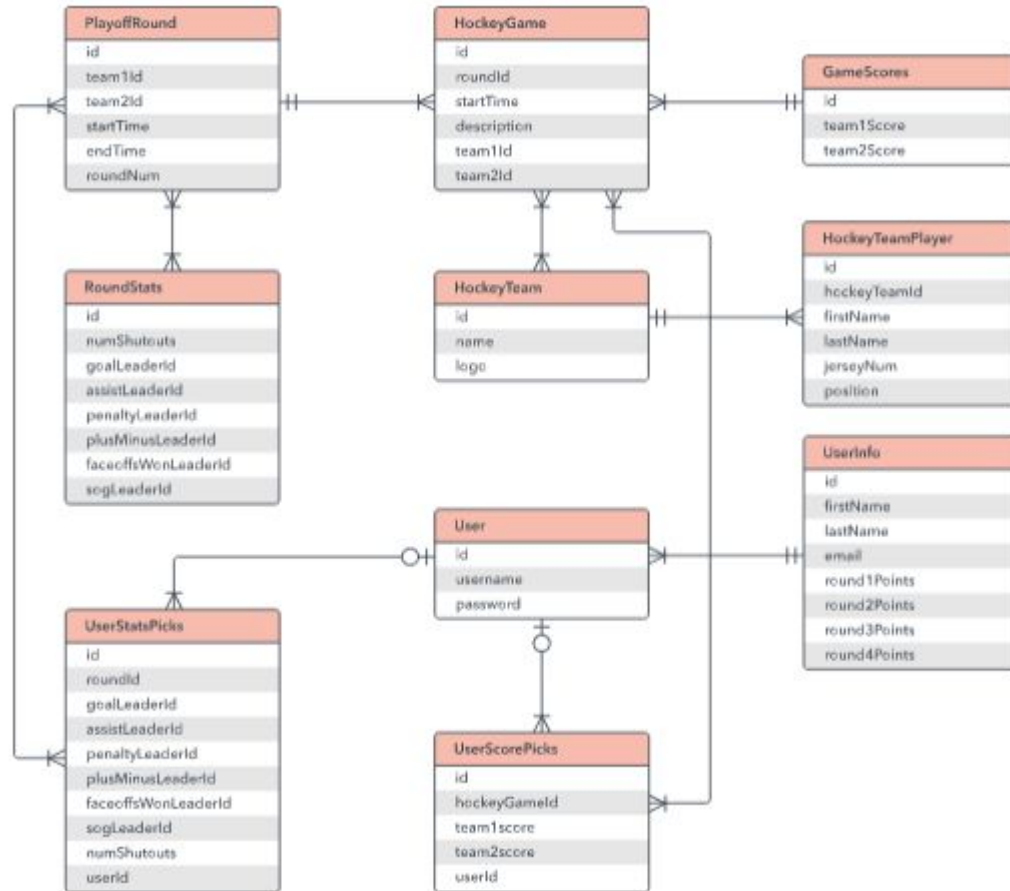
ER Diagram

Entity Relationship (ER) Diagram is a type of flowchart to represent how entities relate to each other.
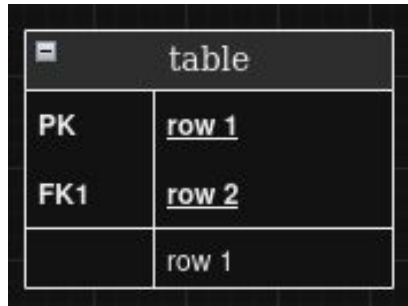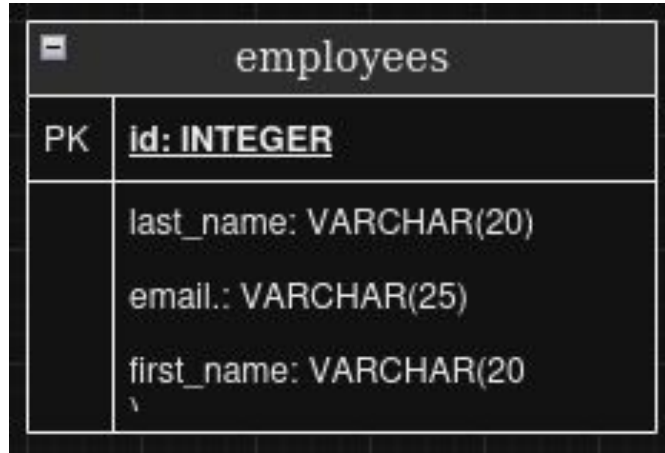
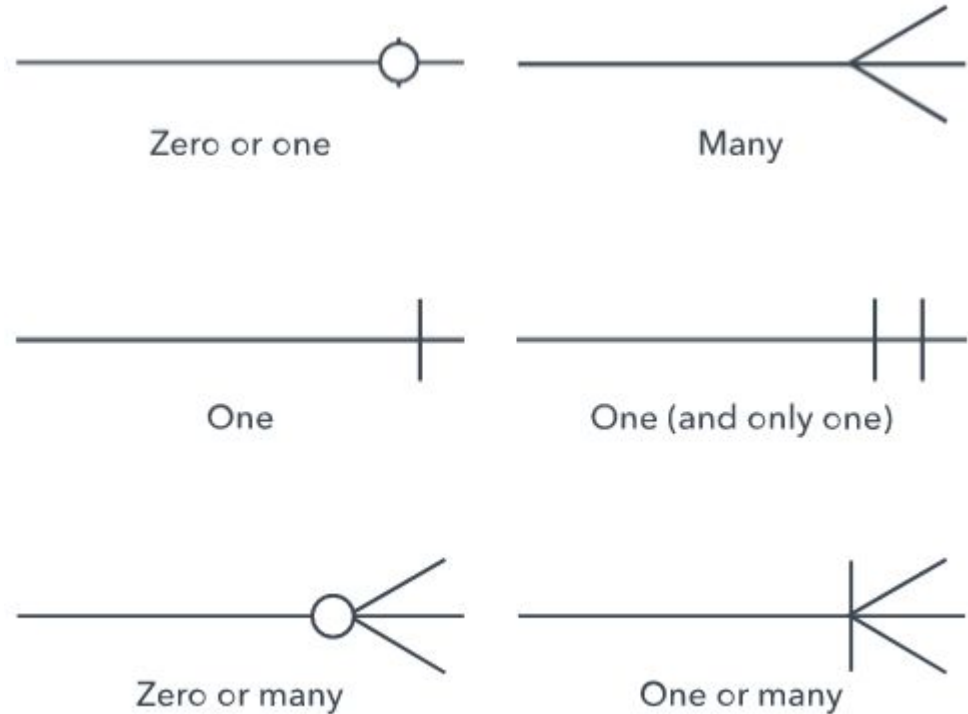We will use this diagram to design our databases.

ER Diagram

ER Diagram

Entity

employees	
PK	**id: INTEGER**
	last_name: VARCHAR(20)
	email.: VARCHAR(25)
	first_name: VARCHAR(20

table	
PK	row 1
FK1	row 2
	row 1

Cardinality

Zero or one

Many

One

One (and only one)

Zero or many

One or many

Data Types (MySQL)

VARCHAR(size): String (letters, numbers, special characters). The size specifies the maximum string length in characters from 0 to 65535

SMALLINT(size): Small integer. Range: -32768 to 32767. Unsigned range from 0 to 65535.

MEDIUMINT(size): Medium integer. Range: -8388608 to 8388607. Unsigned range from 0 to 16777215.

INTEGER(size): Range is from -2147483648 to 2147483647. Unsigned range is from 0 to 4294967295.

FLOAT(size, d): floating point number.The total number of digits is specified in size. The number of digits after the decimal point is specified in the d parameter.

BOOLEAN: Zero is considered as false, nonzero values are considered as true.

More info here: https://dev.mysql.com/doc/refman/8.2/en/data-types.html