



# Almacenamiento y captura de datos

Claudio Aracena

Centro de Modelamiento Matemático  
Chatbot Chile



# Contenidos

- Captura de datos desde archivos
- **Base de datos**
- Captura y almacenamiento de datos en BD
- Captura de datos de la Web (Web scraping)
- Captura de datos de API (ej: Twitter)
- Captura y almacenamiento en arquitecturas Big data

Códigos y clase en:

<https://github.com/caracena/almacenamiento-captura-datos>

# Clase de hoy



## Base de datos

- Sistema Gestor de Base de datos
- Modelo entidad-relación
- SQL (Structured Query Language)
- Base de datos no relacionales





# Base de datos

Una Base de Datos es una colección de información perteneciente a un mismo contexto (o problema), que está almacenada de forma organizada en archivos.

Una base de datos está organizada mediante **tablas** que almacenan información concerniente a algún objeto o suceso. Estas tablas se relacionan formando vínculos o **relaciones** entre ellas, que ayudan a mantener la información de los diversos objetos de forma ordenada y coherente.

Cada una de estas tablas es una estructura que se parece a las hojas de cálculo, pues está dispuesta mediante filas y columnas (**registros** y **campos**).



# Sistema Gestor de Base de datos

Se define un Sistema Gestor de Base de Datos (SGBD o DBMS en inglés) como el conjunto de herramientas que facilitan la consulta, uso y actualización de una base de datos.

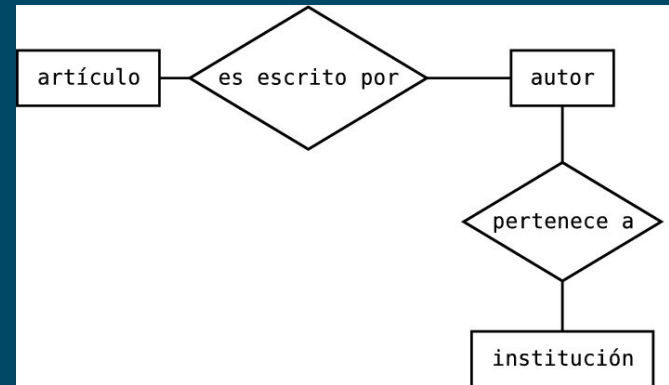




# Modelo entidad-relación

Este modelo consiste en plasmar el resultado del análisis del problema mediante diagramas entidad-relación.

- Entidad: Cualquier tipo de objeto o concepto sobre el que se recoge información; cosa, persona, concepto abstracto o suceso
- Relación: Una relación (o interrelación), es una correspondencia o asociación entre dos o más entidades.





# SQL (Structured Query Language)

La principal herramienta de un gestor de base de datos es la interfaz de programación con el usuario. Este interfaz consiste en un lenguaje muy sencillo llamado SQL, Structured Query Language. SQL está estandarizado por la ISO, es decir, todas las bases de datos que soporten SQL deben tener la misma sintaxis a la hora de aplicar el lenguaje.

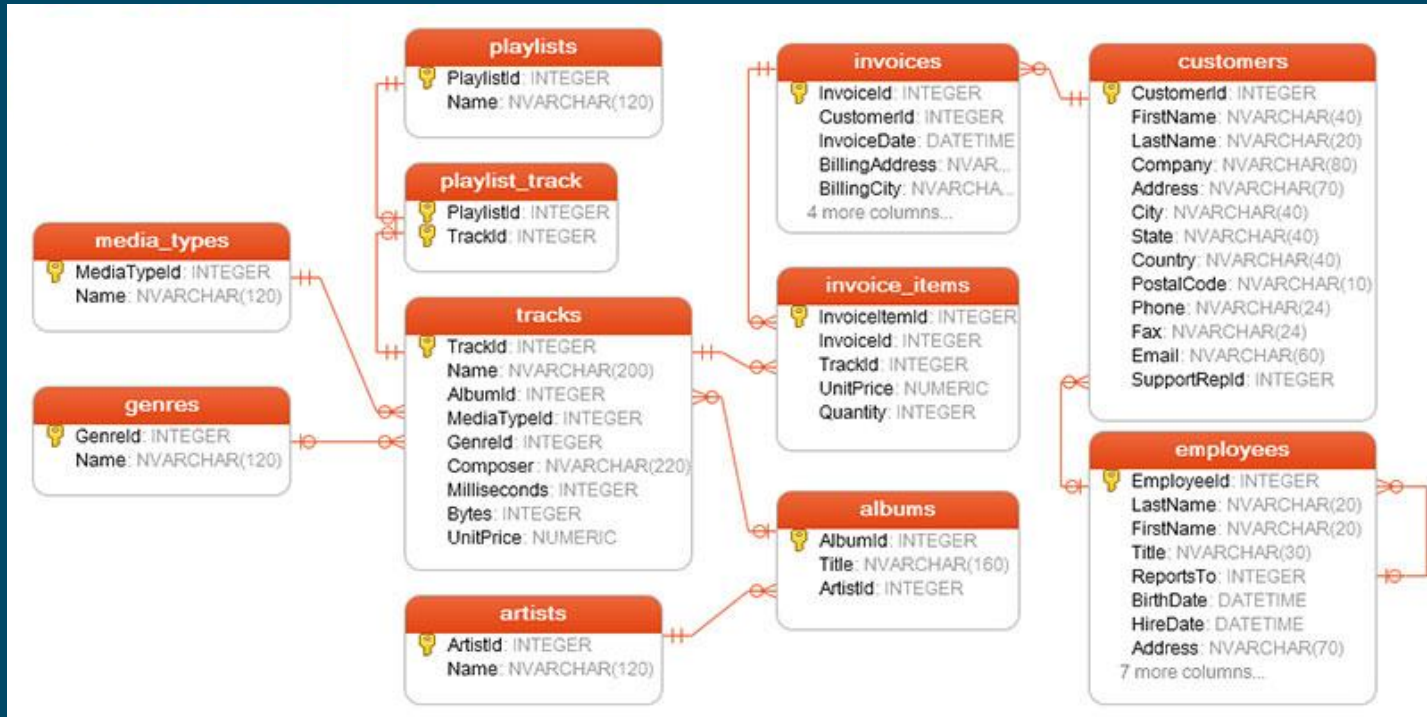
Comandos más comunes:

- SELECT, UPDATE, INSERT, DELETE, CREATE, DROP

# SQL (Structured Query Language)



- Abrir <https://sqliteonline.com/#fiddle-5c0a7bb908f70asdjpe3e8dk>







# Base de datos no relacionales (NoSQL)

NoSQL es el término genérico usado para referirse a almacenamiento de datos que no sigue el modelo tradicional de base de datos relacionales.

Específicamente, la data no sigue el modelo entidad-relación y no utiliza SQL como lenguaje de consulta.

Ejemplos de estas bases de datos son

- MongoDB (document-oriented)
- Cassandra, Hbase (column-oriented)
- Redis (key-value)
- Neo4j (graph-oriented)



# MongoDB



## Comparación entre MongoDB y base relacional

| ID      | Nombre | Email         | Edad |
|---------|--------|---------------|------|
| 11111-k | Juan   | juan@udd.cl   | 30   |
| 2222-2  | Carlos | carlos@udd.cl | 35   |

```
{_id: "11111-k", nombre: "Juan", email: "juan@udd.cl", edad: 30}  
{_id: "2222-2", nombre: "Carlos", email: "carlos@udd.cl", edad: 35}
```

Consola online:

[https://www.tutorialspoint.com/mongodb\\_terminal\\_online.php](https://www.tutorialspoint.com/mongodb_terminal_online.php)