

Tutorial: Manejo de Bases de Datos en R

Yeison Yovany Ocampo Naranjo
Científico de Datos en Nequi

R Day 2023

Contenido

- Introducción
- Aspectos importantes
- Lenguajes de consulta
- Definición
- Ejemplo en R



¿Cuál es la cantidad de datos más grande que ha usado?



Introducción

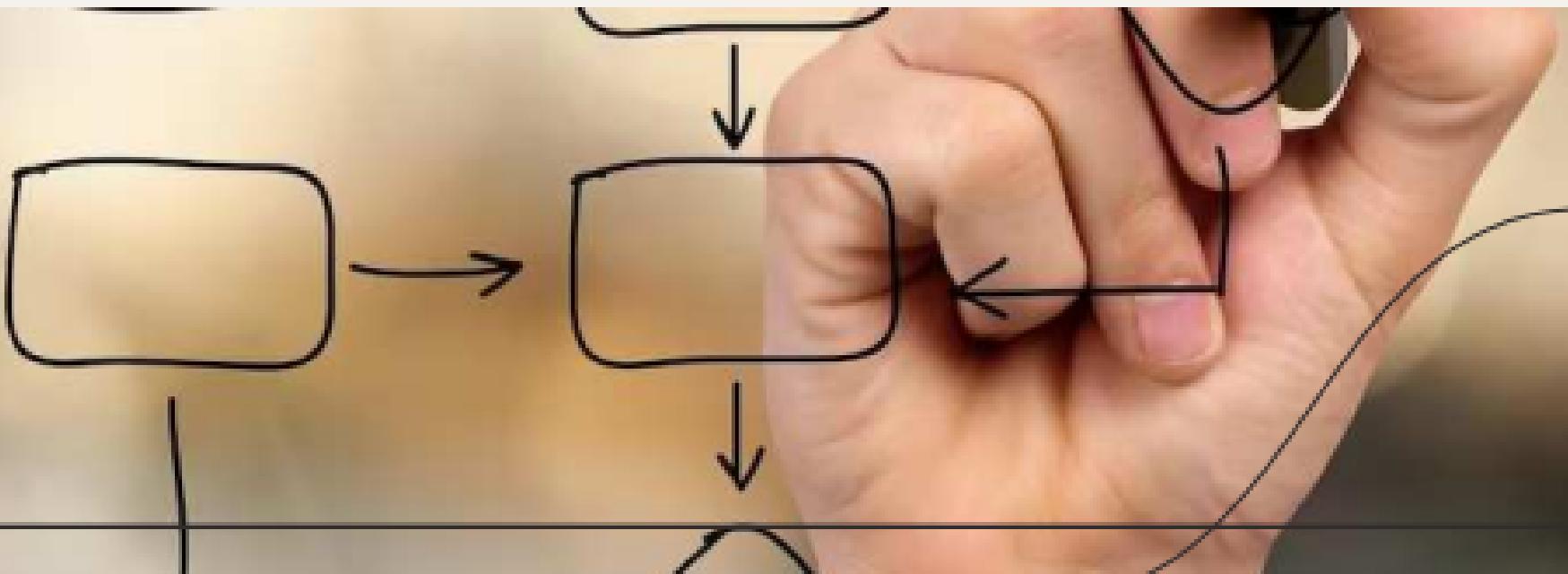
En esta presentación se explorarán los fundamentos de las bases de datos, su importancia en el mundo profesional y su aplicación en diferentes campos. Se abordarán conceptos clave como modelado de datos, lenguajes de consulta y estructuras de almacenamiento. Así como su integración en el lenguaje R



Las bases de datos son fundamentales en el entorno profesional, ya que permiten la gestión eficiente de la información. Su correcto diseño y administración son esenciales para la toma de decisiones informadas y el desarrollo de aplicaciones empresariales.



El modelado de datos es el proceso de definir la estructura y relaciones de la información en una base de datos. Se utilizan diagramas de entidad-relación y esquemas conceptuales para representar de manera clara y precisa la organización de los datos.



Lenguajes de Consulta

Los lenguajes de consulta permiten interactuar con las bases de datos para recuperar, actualizar o eliminar información. Ejemplos comunes son SQL (Structured Query Language) y sus variantes, que son fundamentales para el manejo eficiente de datos.



Sistemas de Gestión de Bases de Datos



Los sistemas de gestión de bases de datos (SGBD) son herramientas que permiten administrar de manera eficiente el almacenamiento, recuperación y seguridad de la información. Ejemplos incluyen MySQL, Oracle y MongoDB.

La seguridad y privacidad de los datos son aspectos críticos en el manejo de bases de datos profesionales. Se deben implementar medidas de protección, como cifrado, control de acceso y auditorías, para garantizar la integridad y confidencialidad de la información.



Tendencias Actuales en Bases de Datos

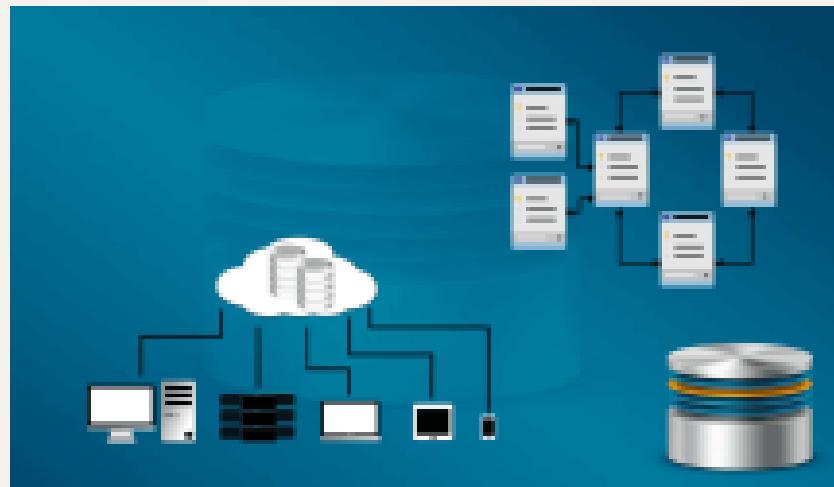
Las **tendencias actuales** en bases de datos incluyen el uso de tecnologías como bases de datos NoSQL, bases de datos en la nube y enfoques de Big Data. Estas tendencias están transformando la manera en que se gestionan y analizan grandes volúmenes de información.



Definición de Base de Datos

Una base de datos es un sistema organizado para recopilar, almacenar y gestionar datos de manera estructurada.

Facilitar la recuperación y manipulación eficiente de la información.



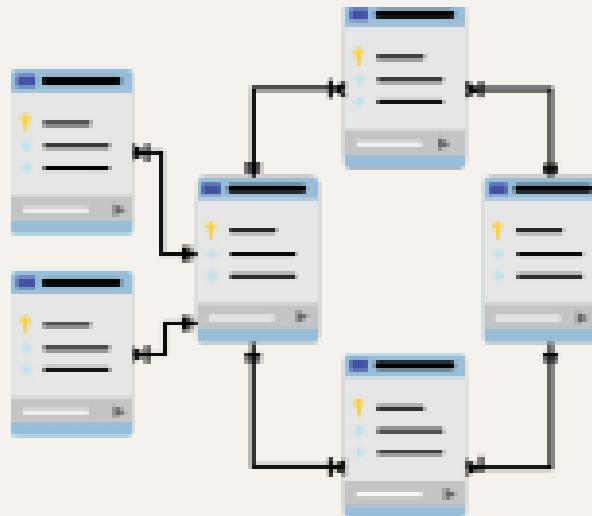
Datos, tablas, relaciones, y un sistema de gestión de bases de datos (DBMS).

Tipos de Base de Datos

- **Bases de Datos Relacionales:** Utilizan tablas y relaciones para organizar datos (Ejemplo: MySQL, PostgreSQL).
- **Bases de Datos NoSQL:** Diseñadas para manejar grandes volúmenes de datos no estructurados o semi-estructurados (Ejemplo: MongoDB, Cassandra).
- **Bases de Datos In-Memory:** Almacenan datos en la memoria principal para un acceso ultrarrápido (Ejemplo: Redis).
- **Bases de Datos Distribuidas:** Distribuyen datos en múltiples nodos para mejorar la escalabilidad y la disponibilidad (Ejemplo: Amazon DynamoDB).



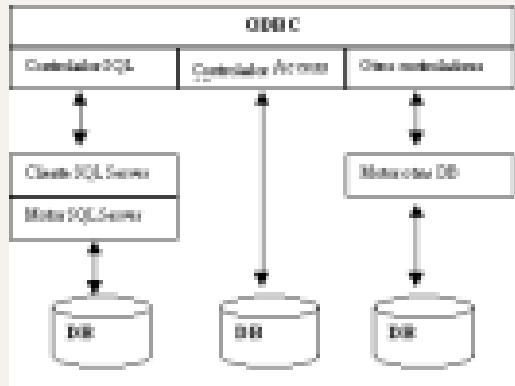
Bases de Datos Relacional (RDB)



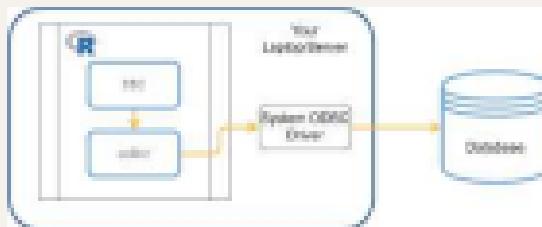
Una RDB (Relational Database) o BD Relacional es un tipo de base de datos que organiza la información en tablas bidimensionales, donde los datos se almacenan de manera estructurada y se pueden acceder mediante relaciones entre las tablas.

En una RDB, la información se organiza en filas y columnas, y se establecen relaciones lógicas entre las tablas para representar la estructura y las conexiones de los datos.

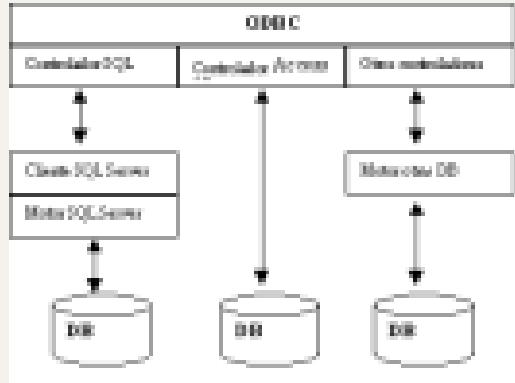
¿Qué es una ODBC?



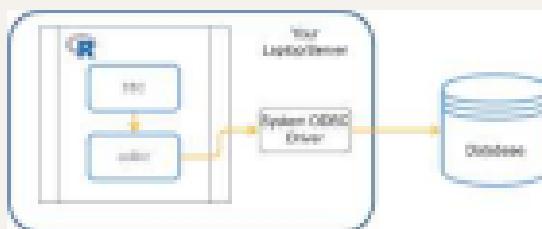
ODBC (Open Database Connectivity) es un estándar que proporciona una interfaz unificada para acceder a diversas bases de datos. La idea detrás de ODBC es permitir que las aplicaciones se conecten a cualquier sistema de gestión de bases de datos (DBMS) que admita ODBC, sin necesidad de cambiar el código de la aplicación.



¿Qué es una DSN?

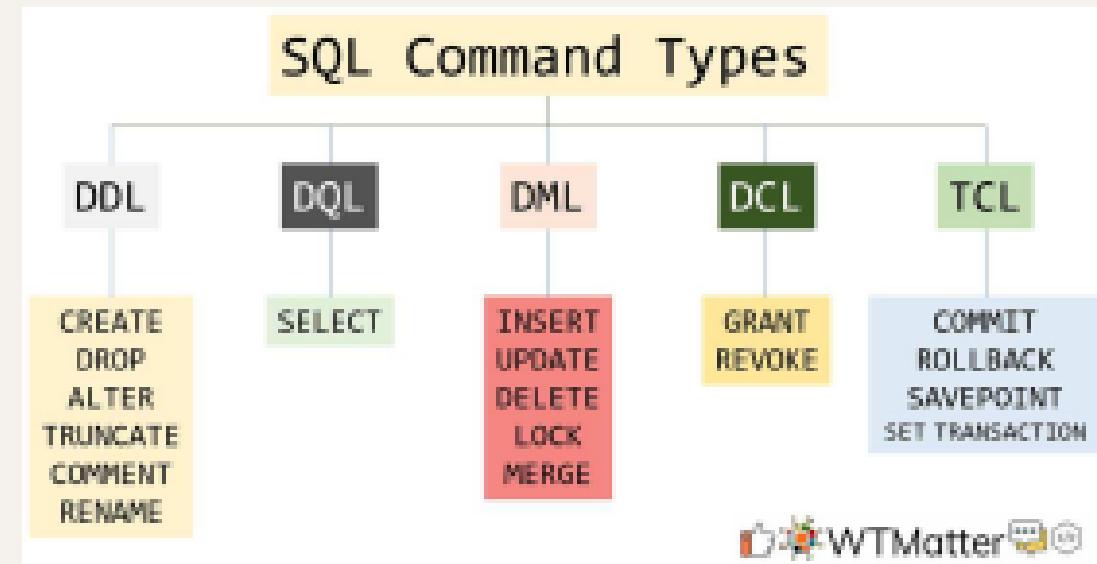


Data Source Name (Nombre de Origen de Datos): En el contexto de bases de datos y conexiones ODBC (Open Database Connectivity), un DSN es una cadena de conexión que especifica los detalles necesarios para conectarse a una base de datos. Contiene información como el tipo de controlador de base de datos, la dirección del servidor, el nombre de la base de datos, y las credenciales de acceso.



Principales comandos

- SQL SELECT
- SQL DISTINCT
- SQL WHERE
- SQL AND OR
- SQL IN
- SQL BETWEEN
- SQL LIKE
- SQL ORDER BY
- SQL GROUP BY
- SQL HAVING

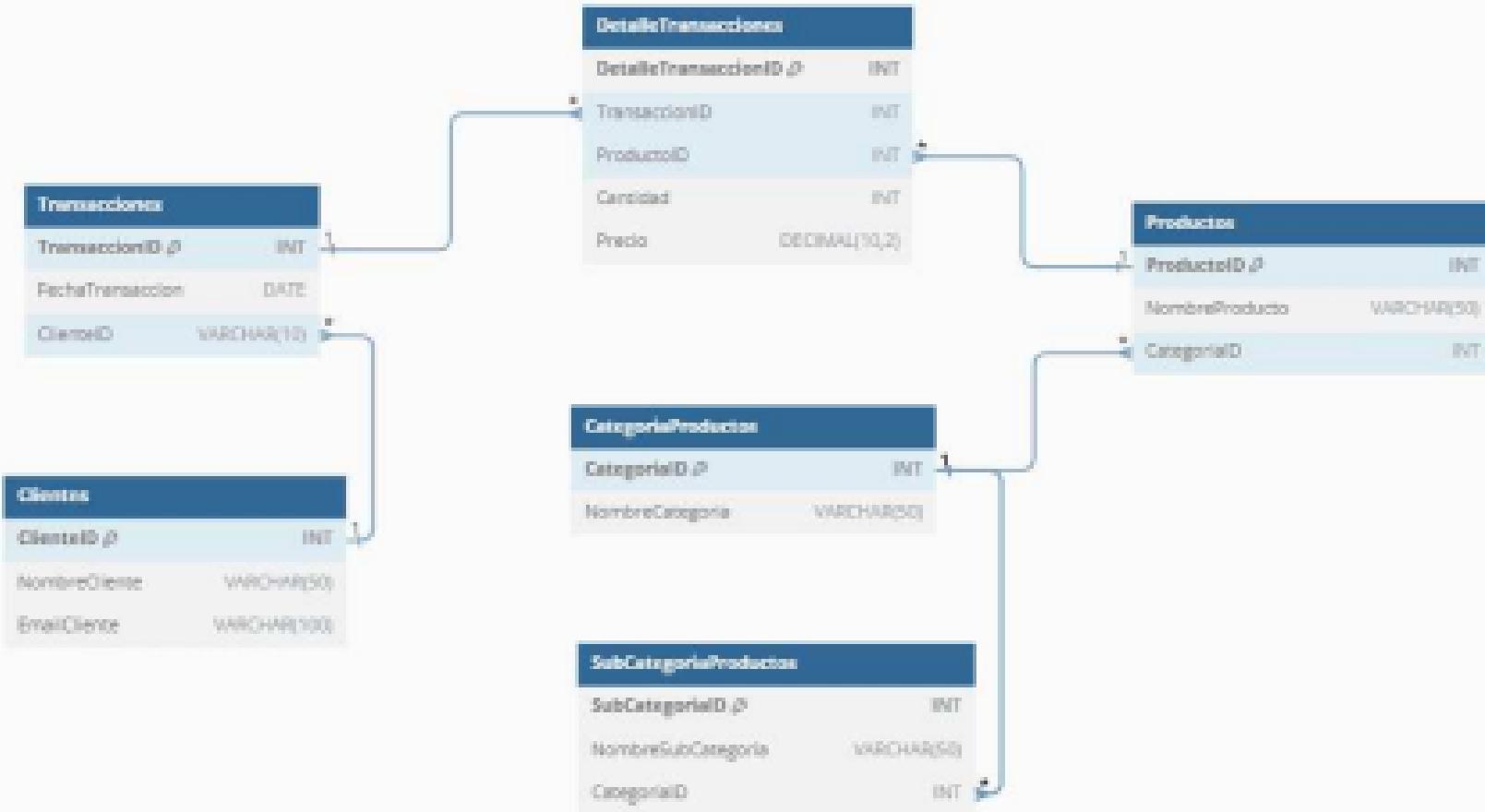


Ejemplo práctico de base de datos para analizar

Se realizará un análisis y reporte basados en la información de una base de datos relacional (RDB) de un supermercado donde se tiene información de:

- Clientes
- Compras
- Productos





¡Ahora si, vamos a R!

