

# Sistema de Gestión de Bases de Datos (SGBD) Enfoque en DuckDB

Jessica Lizeth Hernández Bracho - 1842553

PIA - BOR

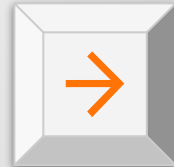




# SGBD



Sistema de Gestión de Bases de Datos



# ¿Qué es un SGBD?



Software que permite administrar una base de datos. Proporciona el método de organización necesario para el almacenamiento y recuperación flexible de grandes cantidades de datos.

- Permite crear, manipular y administrar bases de datos de forma eficiente, segura y estructurada.
- Ejemplos: MySQL, PostgreSQL, Microsoft SQL Server, Oracle Database y Microsoft Access.

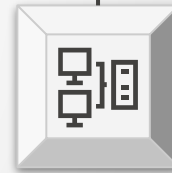
# Clasificación de SGBD

Modelo



## Relacionales

Organizan la información en tablas.  
(DuckDB, MySQL, Oracle)



## No relacionales

Incluyen modelos orientados a documentos, claves/valores, grafos.  
(MongoDB)

# Clasificación de SGBD

## Arquitectura

```
graph TD; A[Arquitectura] --> B[Centralizados]; A --> C[Distribuidos];
```



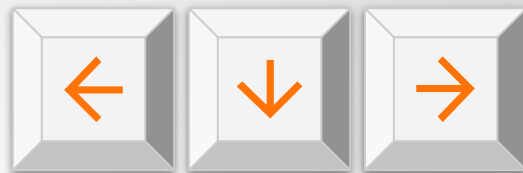
### Centralizados

La base de datos y el software que la gestiona residen en un único servidor.  
(Microsoft Access, SQLite)



### Distribuidos

Dividen los datos entre múltiples nodos o servidores.  
(Amazon DynamoDB)



DuckDB - SGBDR



## ¿Qué es DuckDB?

Sistema de Gestión de Bases de Datos Relacionales (SGBDR) de código abierto y orientado a columnas.

Está diseñado para admitir cargas de trabajo de consultas analíticas, también conocidas como procesamiento analítico en línea (OLAP).

# Características de DuckDB



## Simple

No requiere instalación  
compleja ni configuración.



## Rendimiento

Procesamiento por vectores  
en lugar de filas.  
Aumenta el rendimiento OLAP.



## Soporte SQL

Soporte completo de  
SQL moderno.



## Integración

Lectura directa de Parquet /  
CSV / JSON / Arrow.



## Portabilidad

Puede ejecutarse en  
navegadores web y apps  
móviles.



## MIT License

Código abierto, libre para  
uso personal, académico  
o comercial.



## ¿Cómo se compara DuckDB con otros SGBD?

SGBD	Tipo	Arquitectura de almac.	Servidor	Enfoque	Uso
DuckDB	Relacional (OLAP)	Columnar/ Vectorizado.	✗	Analítico	Ciencia de datos, notebooks.
SQLite	Relacional (OLTP)	Fila (row-based)	✗	Ligero	Móviles, Apps locales.
Postgre SQL	Relacional (Híbrido)	Fila	✓	Transaccional y analítico.	Web apps, sistemas de información.

# Ejemplo de uso

```
import duckdb
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('Sample - Superstore.csv', encoding='latin1')

df['Order Date'] = pd.to_datetime(df['Order Date'], errors='coerce')

con = duckdb.connect()
```

## 1- Cargar CSV

Carga el dataset, usando codificación para evitar errores de caracteres.

## 2- Conexión

Establece una conexión con DuckDB para ejecutar consultas.

```
# Ejecutar consultas usando df como tabla virtual
# 1. Ventas por categoría
category = con.execute("""
    SELECT "Category", SUM("Sales") AS total_ventas
    FROM df
    GROUP BY "Category"
""").df()
print("Total ventas por categoría:\n", category)

# 2. Ventas por año
anual_sales = con.execute("""
    SELECT STRFTIME("Order Date", '%Y') AS Year, SUM("Sales") AS total_sales
    FROM df
    GROUP BY Year
    ORDER BY Year
""").df()
print("Ventas por año:\n", anual_sales)
```



Consultas





### 1. Total ventas por categoría:

	Category	total_ventas
0	Furniture	741,999.7953
1	Technology	836,154.0330
2	Office Supplies	719,047.0320

### 2. Ventas por año:

	Year	total_sales
0	2014	484,247.4981
1	2015	470,532.5090
2	2016	609,205.5980
3	2017	733,215.2552

## Resultados

# Resultados

Total ventas por categoría:

	Category	total_ventas
0	Furniture	741999.7953
1	Technology	836154.0330
2	Office Supplies	719047.0320

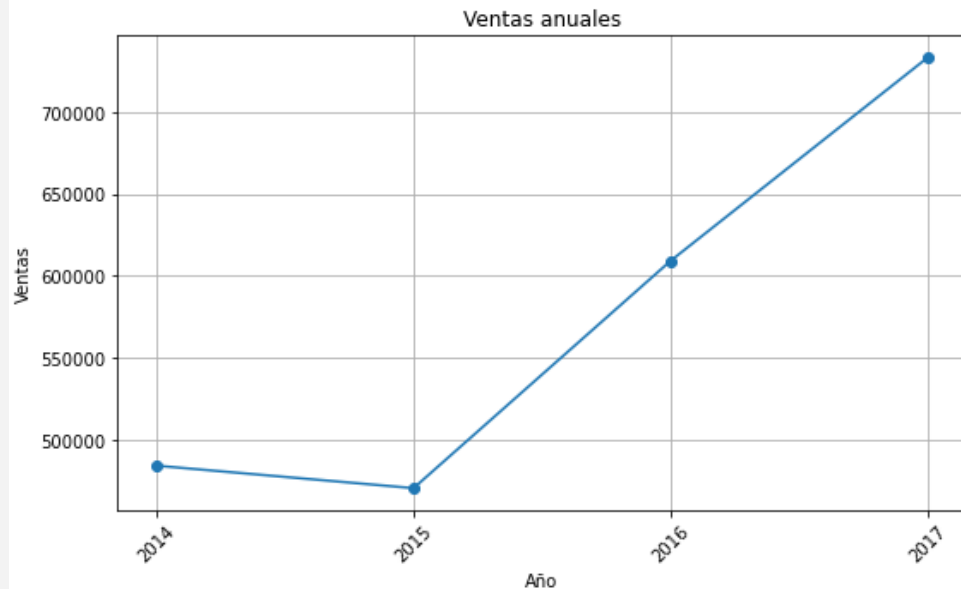
Ventas por año:

	Year	total_sales
0	2014	484247.4981
1	2015	470532.5090
2	2016	609205.5980
3	2017	733215.2552

# Resultados

```
# Graficar ventas anuales
plt.figure(figsize=(8,5))
plt.plot(anual_sales['Year'], anual_sales['total_sales'],
         marker='o', linestyle='--')
plt.title('Ventas anuales')
plt.xticks(rotation=45)
plt.xlabel('Año')
plt.ylabel('Ventas')
plt.grid(True)
plt.tight_layout()
plt.show()
```

# Resultados



## Gráfica

Matplotlib para graficar las ventas anuales, haciendo la información más comprensible visualmente.

```
import pandas as pd
import duckdb

# 1. Carga los datos nuevos desde CSV
df_nuevos = pd.read_csv('Cliente.csv')

# 2. Eliminar la primera columna (por posición)
df_nuevos.drop(df_nuevos.columns[0], axis=1, inplace=True)

# 2. Conexión a DuckDB (archivo tienda1.duckdb)
con = duckdb.connect('tienda1.duckdb')

# 3. Registra DataFrame
con.register('df_temp', df_nuevos)

# 4. Inserta datos nuevos en tabla Cliente
con.execute("""
INSERT INTO Cliente
SELECT * FROM df_temp
""")

# 5. Consulta para mostrar todos los datos de Cliente después de insertar
df_actualizado = con.execute("SELECT * FROM Cliente").fetchdf()

print("Datos actuales en la tabla Cliente:")
print(df_actualizado)
```



Ejemplo con base de datos





Datos actuales en la tabla Cliente:

	clienteID	c_nombre	segmento	pais	\
0	CG-12520	Claire Gute	Consumer	United States	
1	DV-13045	Darrin Van Huff	Corporate	United States	
2	SO-20335	Sean ODonnell	Home Office	United States	
3	AA-10480	Andrew Allen	Consumer	United States	
4	IM-15070	Irene Maddox	Consumer	United States	
5	HP-14815	Harold Pawlan	Home Office	United States	
6	RA-19885	Ruben Ausman	Corporate	United States	
7	ES-14080	Erin Smith	Corporate	United States	
8	ON-18715	Odella Nelson	Corporate	United States	
9	LH-16900	Lena Hernandez	Consumer	United States	
10	XT-2688	Hilary Lynds	Consumer	United States	
11	JJ-3659	Ursula Thies	Home Office	United States	
12	TA-1738	Nani Alekhov	Consumer	United States	
13	OZ-3445	Desdemona Tapner	Corporate	United States	
14	UE-7582	Reggy Bruck	Corporate	United States	
15	QH-4354	Hillel Huxton	Consumer	United States	
16	ZQ-6510	Mohammed Bumford	Consumer	United States	
17	VR-4151	Lemuel Featherstone	Home Office	United States	
18	CP-5108	Court Squeers	Corporate	United States	
19	HJ-9118	Care Stairs	Home Office	United States	
20	DJ-6938	Nelly Pitkaithly	Home Office	United States	
21	DT-5463	Sam Brerman	Corporate	United States	



Ejemplo con base de datos



# Operaciones relacionales

## 1. Proyección + Selección

Clientes que hicieron pedidos en California

- $\pi(\text{Nombre Cliente})(\sigma(\text{Estado} = \text{'California'})(\text{Pedidos}))$
- Lista de nombres de clientes en California con pedidos.

```
df_clientes_ca = con.execute("""
SELECT DISTINCT c.c_nombre
FROM Pedido p
JOIN Cliente c ON p.clienteID = c.clienteID
WHERE c.estado = 'California'
""").fetchdf()

print("Clientes que hicieron pedidos en California:")
print(df_clientes_ca)
```



	c_nombre
0	Ruben Ausman
1	Hillel Huxton
2	Andros Randles
3	Mohammed Bumford

# Operaciones relacionales

## 2. Join + Selección + Proyección

Nombres de productos tecnológicos vendidos

- $\pi(\text{Nombre Producto})(\sigma(\text{Categoría} = \text{'Technology'}) (\text{Pedidos} \bowtie \text{Productos}))$
- Listado de productos tecnológicos vendidos.

```
df_tecnologia = con.execute("""
SELECT DISTINCT pr.p_nombre
FROM DetallePedido dp
JOIN Producto pr ON dp.productoID = pr.productoID
WHERE pr.categoria = 'Technology'
""").fetchdf()

print("Productos tecnológicos vendidos:")
print(df_tecnologia)
```



	p_nombre
0	Verbatim 25 GB 6x Blu-ray Single Layer Record...
1	LF Elite 3D Dazzle Designer Hard Case Cover, L...
2	Fellowes Super Stor/Drawer
3	Invicta Office Printer Cart
4	Logitech LS21 Speaker System - 2.1-CH Wired
5	iPhone Case
6	Aurora Paper Trimmer
7	Hon Deluxe Fabric Upholstered Stacking Chairs ...
8	Self-Adhesive Address Labels for Typewriters b...
9	Canon imageCLASS D1320
10	Eldon Light Bulb - Duo Pack
11	Tenex Box - Single Width
12	Logitech Wireless Mouse M510

# Conclusión

01

## DuckDB

SGBD relacional embebido, optimizado para análisis de datos en memoria.

03

## Servidor

No requiere servidor, lo que lo hace ligero, rápido y fácil de usar.

02

## Integración

Se integra perfectamente con Python y Pandas, ideal para ciencia de datos.

04

## Archivos

Soporta lectura directa de archivos CSV.

# Referencias

- Wikipedia. (s.f.). Sistema de gestión de bases de datos – Lenguajes de modelación. Wikipedia.  
[https://es.wikipedia.org/wiki/Sistema\\_de\\_gesti%C3%B3n\\_de\\_bases\\_de\\_datos#Lenguajes\\_de\\_modelaci%C3%B3n](https://es.wikipedia.org/wiki/Sistema_de_gesti%C3%B3n_de_bases_de_datos#Lenguajes_de_modelaci%C3%B3n)
- IONOS. (s.f.). ¿Qué es un sistema gestor de base de datos (SGBD)?. IONOS Digital Guide.  
<https://www.ionos.mx/digitalguide/hosting/cuestiones-tecnicas/sistema-gestor-de-base-de-datos-sgbd/#c221916>



## Sistema de Gestión de Bases de Datos (SGBD) Enfoque en DuckDB

Jessica Lizeth Hernández Bracho - 1842553



# Referencias

- Dongee. (s.f.). ¿Cómo se puede clasificar un SGBD?. Dongee. <https://www.dongee.com/tutoriales/como-se-puede-clasificar-un-sgbd/#clasificaci%C3%B3n-de-sgbd-seg%C3%BA-la-arquitectura-del-sistema>
- Wikipedia. (s.f.). DuckDB. Wikipedia. <https://en.wikipedia.org/wiki/DuckDB>
- DuckDB Developers. (s.f.). DuckDB: The SQLite for Analytics. DuckDB. <https://duckdb.org/>



## Sistema de Gestión de Bases de Datos (SGBD) Enfoque en DuckDB

Jessica Lizeth Hernández Bracho - 1842553

