

Almacenamiento y captura de datos

Claudio Aracena

GobLab - Universidad Adolfo Ibáñez
Chatbot Chile



Google Cloud Platform

Cloud SQL

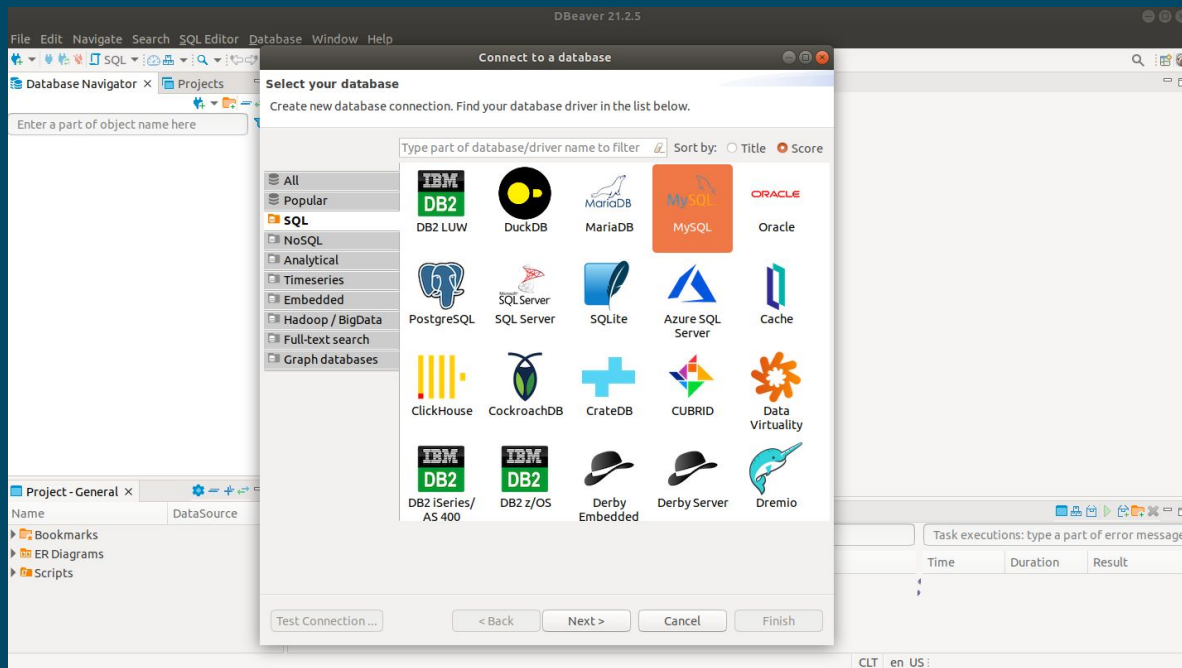
- Base de datos alojada en la nube de Google
- 3 opciones de motores (SGBD) de base de datos
 - MySQL
 - PostgreSQL
 - SQL Server
- Se paga por cpu, ram, espacio, tipo de espacio, ubicación de los servidores, entre otros

Dbeaver



- Cliente SQL

- <https://dbeaver.io/>





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	Juan	juan@udd.cl	30
2222	Carlos	carlos@udd.cl	35

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

MongoDb online: <https://www.mongodb.com/online>



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



Captura y almacenamiento de datos en BD

- Captura de datos desde BD
- Almacenamiento en BD
- Procesamiento de datos en Python





Conexión y consulta a base de datos

```
import sqlite3

conn = sqlite3.connect("data/chinook.db")

cur = conn.cursor()

cur.execute("SELECT * FROM albums")

rows = cur.fetchall()
```




Conexión y consulta a base de datos

```
import mysql.connector
```

```
conn = mysql.connector.connect(
```

```
    host="dirección servidor",
```

```
    user="usuario",
```

```
    password="contraseña",
```

```
    database="nombre_base"
```

```
)
```

```
cur = conn.cursor()
```

```
cur.execute("SELECT * FROM  
albums")
```

```
rows = cur.fetchall()
```



Consulta a base de datos con pandas

```
import pandas as pd
```

```
df = pd.read_sql("SELECT * FROM albums", conn)
```



Creación de base de datos

```
!pip install sqlalchemy
```

```
from sqlalchemy import create_engine
```

```
sql_engine = create_engine('sqlite:///data/test.db')
```

```
connection = sql_engine.raw_connection()
```

```
df.to_sql('artists_albums', connection, index=False)
```