




# DESARROLLO E IMPLEMENTACIÓN DE UN DATA WAREHOUSE PARA HISTORIAL DE SPOTIFY

FRANCISCO JAVIER DELGADO RODRIGUEZ Y ALBERTO PALENCIA QUIÑONES

BASES DE DATOS AVANZADAS  
ESCUELA SUPERIOR DE INFORMATICA



## Contenido

Introducción.....	1
Diseño de la Base de Datos .....	1
Atributos de las Tablas .....	1
Desarrollo de la Aplicación en Java .....	3
Ejecución del Proyecto.....	3
Ejecución de las Consultas .....	4
Conclusiones .....	5

## Introducción

Este proyecto tiene como objetivo la creación e implementación de un Data Warehouse (DWH) para almacenar y analizar el historial de reproducciones de Spotify. Para ello, se diseñó una base de datos relacional en MySQL y se desarrolló una aplicación en Java para la carga y consulta de datos.

## Diseño de la Base de Datos

La base de datos está organizada en un esquema dimensional para facilitar el análisis de los datos. Entre sus principales tablas se incluyen:

- **dim\_artist:** Contiene información sobre los artistas.
- **dim\_album:** Almacena los álbumes.
- **dim\_track:** Registra las canciones.
- **dim\_platform:** Incluye las plataformas de reproducción.
- **dim\_time:** Gestiona la información temporal.
- **fact\_streaming:** Tabla de hechos que almacena la información de reproducción.

El script DWH.sql define la estructura de la base de datos, incluyendo las claves primarias y foráneas necesarias para asegurar la integridad referencial.

## Atributos de las Tablas

- **dim\_artist**
  - **id\_artist:** Identificador único del artista (Clave primaria).
  - **artist\_name:** Nombre del artista.
- **dim\_album**
  - **id\_album:** Identificador único del álbum (Clave primaria).

- album\_name: Nombre del álbum.
- **dim\_track**
  - id\_track: Identificador único de la canción (Clave primaria).
  - track\_name: Nombre de la canción.
  - spotify\_track\_uri: Identificador único de la canción en Spotify.
  - id\_album: Referencia al álbum de la canción (Clave foránea).
  - id\_artist: Referencia al artista de la canción (Clave foránea).
- **dim\_platform**
  - id\_platform: Identificador único de la plataforma (Clave primaria).
  - platform\_name: Nombre de la plataforma de reproducción.
- **dim\_time**
  - id\_time: Identificador único del tiempo (Clave primaria).
  - ts: Marca temporal (timestamp) de la reproducción.
  - year: Año de la reproducción.
  - month: Mes de la reproducción.
  - day: Día de la reproducción.
- **fact\_streaming** (Tabla de hechos)
  - id\_streaming: Identificador único del evento de reproducción (Clave primaria).
  - spotify\_track\_uri: Referencia a la canción reproducida.
  - ts: Timestamp de la reproducción.
  - platform: Plataforma utilizada para la reproducción.
  - ms\_played: Duración en milisegundos de la reproducción.
  - reason\_start: Motivo por el cual comenzó la reproducción.
  - reason\_end: Motivo por el cual terminó la reproducción.
  - shuffle: Indica si la reproducción estaba en modo aleatorio.
  - skipped: Indica si la canción fue saltada.
  - id\_track: Clave foránea hacia dim\_track.
  - id\_album: Clave foránea hacia dim\_album.

- id\_artist: Clave foránea hacia dim\_artist.
- id\_time: Clave foránea hacia dim\_time.
- id\_platform: Clave foránea hacia dim\_platform.

## Desarrollo de la Aplicación en Java

Se creó un programa en Java (Main.java) para manejar la carga de datos desde un archivo CSV y realizar consultas sobre la base de datos. Las principales funciones del código incluyen:

1. **Conexión con la Base de Datos:** Se utiliza JDBC para conectarse a MySQL.
2. **Carga de Datos:**
  - Se lee un archivo CSV (spotify\_history\_trimmed.csv).
  - Se extraen los valores relevantes y se insertan en la base de datos.
  - Se implementan métodos para verificar si un registro ya existe antes de insertarlo.
3. **Consultas:**
  - Se implementa un menú interactivo en consola que permite ejecutar consultas sobre los datos almacenados, como obtener todos los artistas, álbumes, canciones, fechas y plataformas.

## Ejecución del Proyecto

Para ejecutar el sistema, se deben seguir los siguientes pasos:

1. Crear la base de datos en MySQL utilizando el script DWH.sql.
2. Ejecutar Main.java, asegurándose de que el archivo CSV esté en la ubicación correcta.
3. Interactuar con el menú para explorar los datos almacenados.

## Ejecución de las Consultas

El programa permite realizar cinco tipos de consultas a la base de datos mediante un menú interactivo:

### 1. Mostrar las canciones reproducidas y el mes de reproducción

Track: Old Thing Back (feat. Ja Rule and Ralph Tresvant), Artist: Matoma, Album: Old Thing Back (feat. Ja Rule and Ralph Tresvant), Month Played: 8

Track: Old Thing Back (feat. Ja Rule and Ralph Tresvant), Artist: Matoma, Album: Old Thing Back (feat. Ja Rule and Ralph Tresvant), Month Played: 8

Track: Dolphin Style, Artist: Death Team, Album: Dolphin Style, Month Played: 8

Track: Carnaval - Niels Van Gogh Radio Edit, Artist: Alex Divello, Album: Carnaval, Month Played: 8

Track: Drop It Again (Original Mix), Artist: Vigiland, Album: Drop It Again, Month Played: 8

Track: Love Me Back, Artist: LunchMoney Lewis, Album: Bills - EP, Month Played: 8

Track: Kryptonite - Original Mix, Artist: Goh, Album: LaBoracay 2015, Month Played: 8

Track: Deep End (feat. Eastside and Mayer Hawthorne), Artist: Coucheron, Album: Deep End (feat. Eastside and Mayer Hawthorne), Month Played: 8

Track: Bills, Artist: LunchMoney Lewis, Album: Bills - EP, Month Played: 8

Track: California - Chris Lake & Matroda Remix, Artist: SNBRN, Album: California, Month Played: 8

Track: Mentirás - Kiko Carmona Mix, Artist: Salma, Album: Mentirás, Month Played: 8

Track: Drop It Again (Radio Edit), Artist: Vigiland, Album: Drop It Again, Month Played: 8

Track: Fucking Bitches In The Hood, Artist: Death Team, Album: Fucking Bitches In The Hood, Month Played: 8

2. **Mostrar la canción más reproducida en el mes de enero.**

```
Consulta 2: Canción más reproducida en enero
Most Played Song in January: Little Wing, Artist: Stevie Ray Vaughan, Album: The Best Of, Total ms Played: 5730662
Seleccione la consulta a ejecutar (1-5) o 0 para salir:
```

### 3. Muestra las 10 canciones más reproducidas por cantidad de reproducciones

Consulta 3: Top 10 canciones más reproducidas por cantidad de reproducciones

Track: Thousand Eyes, Artist: Of Monsters and Men, Album: Beneath The Skin, Play Count: 126

Track: Stole the Show, Artist: Kygo, Album: Stole the Show, Play Count: 120

Track: Happy, Artist: The Rolling Stones, Album: Exile On Main St, Play Count: 115

Track: Take Me To Church, Artist: Hozier, Album: Hozier, Play Count: 104

Track: "Purple Haze - Live 10/11/68 WinterLand, Artist: San Francisco, Album: CA", Play Count: 98

Track: Sun Is Shining, Artist: Axwell /\ Ingrosso, Album: Sun Is Shining, Play Count: 96

Track: Photograph, Artist: Ed Sheeran, Album: x, Play Count: 90

Track: "Earned It (Fifty Shades Of Grey) - From The ""Fifty Shades Of Grey"" Soundtrack", Artist: The Weeknd, Album: "Earned It (Fifty Shades Of Grey) - From The ""Fifty Shades Of Grey"" Soundtrack", Play Count: 89

Track: Locked Away (feat. Adam Levine), Artist: R. City, Album: Locked Away (feat. Adam Levine), Play Count: 85

Track: Paper Doll, Artist: John Mayer, Album: Paper Doll, Play Count: 84

**4. Muestra las 10 canciones más reproducidas por tiempo total de reproducción**

Consulta 4: Top 10 canciones más reproducidas por tiempo total de reproducción

Track: Paper Doll, Artist: John Mayer, Album: Paper Doll, Total ms Played: 18001018

Track: Can't You Hear Me Knocking - 2009 Mix, Artist: The Rolling Stones, Album: Sticky Fingers, Total ms Played: 14406396

Track: Take Me To Church, Artist: Hozier, Album: Hozier, Total ms Played: 14289420

Track: Mirrors, Artist: Justin Timberlake, Album: The 20/20 Experience (Deluxe Version), Total ms Played: 14049336

Track: "In Your Atmosphere - Live at the Nokia Theatre, Artist: Los Angeles, Album: CA - December 2007", Total ms Played: 13110724

Track: Young And Beautiful, Artist: Lana Del Rey, Album: Young And Beautiful, Total ms Played: 12609744

Track: Paradise City, Artist: Guns N' Roses, Album: Appetite For Destruction, Total ms Played: 10945504

Track: The Way You Look Tonight, Artist: Frank Sinatra, Album: Ultimate Sinatra, Total ms Played: 10927561

Track: Born To Die, Artist: Lana Del Rey, Album: Born To Die - The Paradise Edition, Total ms Played: 10090094

Track: To Be Alone, Artist: Hozier, Album: Hozier, Total ms Played: 9392015

## 5. Muestra las plataformas con más reproducciones

```
Consulta 5: Plataformas con más reproducciones
Platform: android, Play Count: 21950
Platform: iOS, Play Count: 3415
Platform: windows, Play Count: 2148
Platform: web player, Play Count: 1330
Platform: cast to device, Play Count: 14
Seleccione la consulta a ejecutar (1-5) o 0 para salir:
```

El usuario puede seleccionar una opción ingresando el número correspondiente en la consola, lo que ejecutará la consulta y mostrará los resultados en pantalla.

## Conclusiones

Este proyecto permite almacenar y analizar información sobre el uso de Spotify de manera eficiente, aprovechando la estructura de un Data Warehouse para futuras consultas y visualización de tendencias. La implementación en Java facilita la automatización del proceso de carga y consulta de datos.