

Plan de Trabajo del Proyecto Final

Bases de datos no relacionales - 2025

1. Identificación del grupo

Número del grupo: 24

Integrantes: Nadia Martinez - 5.141.045-0

Elizabeth Bennett - 5.147.508-8

2. Familia de proyecto elegida

Seleccionar una de las siguientes familias:

Familia 4: Coast to coast (de la obtención al análisis, posible solución políglota)

3. Objetivos del proyecto

- Integrar información sobre artistas, álbumes, canciones, géneros y colaboraciones entre artistas, utilizando modelos no relacionales.
 - Explorar y visualizar relaciones musicales a través de un grafo de artistas y canciones, con foco en colaboraciones y géneros compartidos.
-

4. Descripción general del proyecto

La idea de este proyecto es construir un sistema de recomendación musical utilizando MongoDB como base documental para el almacenamiento detallado de canciones y sus artistas, y Neo4j como base de grafos para representar las relaciones entre canciones y artistas.

La motivación es poder recomendar canciones a partir de un artista o género, así como también según la popularidad de las mismas. A su vez, nos parece una buena forma de reforzar nuestro entendimiento tanto en las bases documentales como en las de grafos, usando las tecnologías vistas en el curso.

5. Actividades previstas y cronograma tentativo

Actividad	Breve descripción	Semana(s) prevista(s)
Recolección y normalización de datos.	Identificar y descargar datasets relevantes.	Semana 1
Definición del modelo	Modelado de los datos en las distintas bases. Base de datos de documentos para artistas y base de datos de grafos para colaboraciones.	Semana 2 y 3
Carga de datos	Cargar los datos en ambas bases y posibilidad de hacer consultas de pruebas de la carga.	Semana 4
Desarrollo de consultas	Desarrollo de consultas específicas para completar los objetivos del proyecto.	Semana 5
Desarrollo del informe	Preparar el informe completo a entregar y desarrollo de puntos faltantes.	Semana 6

6. Base teórica o artículos relacionados

No vamos a utilizar ningún proyecto en específico.

7. Datos y entornos

Juegos de datos a utilizar :

No está definido el conjunto de datos a usar aún pero tenemos varias opciones a elegir, extraídas de [Kaggle](#). Algunas opciones son:

1. [Spotify Dataset](#)
2. [Musical collaborations between top artists in 2022](#)

Entornos de prueba y herramientas:

La idea es trabajar en un Notebook de Google Colab, y para las distintas bases se va a utilizar MongoDB y Neo4j.

Se va a utilizar Python y Cypher para el desarrollo de las consultas en MongoDB y Neo4j respectivamente.

8. Otras observaciones (opcional)

Tenemos la duda de si es correcto realizar el proyecto utilizando tanto bases de datos documentales, como de grafos. O si sería mejor elegir sólo una de ellas.

Dependiendo de los datasets que vamos a usar, podría pasar que al integrar el dataset de canciones con el de artistas, no tengamos todos los artistas para las canciones. Por lo tanto, es necesario resolver estos casos en que nos falten los datos y que decisión tomaríamos para modelarlo.
