**BASES DE DATOS AVANZADAS**

*DATA WAREHOUSE*

# **PLANIFICACIÓN**

## **Componentes del equipo**

* Miguel Millán Alhambra
* Desiré Saponaro
* Antonio Manjavacas Lucas

## **Fuente de datos**

La fuente de datos elegida corresponde a un repositorio de The Movie Database (TMDb) con información sobre 5000 películas (<https://www.kaggle.com/tmdb/tmdb-movie-metadata>). Del dataset original se han extraído las siguientes características para la elaboración de nuestro almacén de datos:

* Identificador
* Título original
* Presupuesto
* Géneros
* Idioma original
* Popularidad
* Productor
* Fecha de estreno
* Ingresos
* Nota media

## **Herramientas y tecnologías a utilizar**

Para elaborar nuestro almacén de datos emplearemos las siguientes herramientas:

* Para la carga y procesamiento de datos utilizaremos el lenguaje de programación **Python** y el entorno de trabajo **Google Collab**. El uso de cuadernos colaborativos facilita el trabajo en grupo y la documentación en tiempo de desarrollo.
* Para el almacenamiento de datos utilizaremos una base de datos no relacional **MongoDB** almacenada en el servidor *cloud* **MongoDB Atlas**. Las principales ventajas que este tipo de base de datos puede aportarnos es flexibilidad y almacenamiento en la nube.
* La documentación será realizada utilizando **Google Docs**, igualmente facilitando el trabajo colaborativo en la nube.
* Finalmente, la presentación del proyecto se elaborará con **Google Presentations** por las mismas razones.

## **Herramientas y tecnologías a utilizar**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **11/2** | **12/2** | **13/2** | **14/2** | **15/2** | **16/2** | **17/2** |
| Extracción de datos y atributos (ETL) |  | Planificación del proyecto |  | **Fecha límite planificación** |  | Limpieza de datos y carga en MongoDB Atlas |
| **18/2** | **19/2** | **20/2** | **21/2** | **22/2** | **23/2** | **24/2** |
| **Fecha límite entrega parcial** | Diseño de la base de datos y consultas | Diseño de la base de datos y consultas |  |  | Hacer presentación | **Fecha límite almacén de datos** |