

EXPLICACIÓN NETFLIX-ETL

El proyecto Netflix ETL (Extract, Transform, Load) es un sistema diseñado para procesar y analizar datos sobre títulos de Netflix, realizando las siguientes etapas de manera general:

Objetivo del Proyecto

El objetivo principal del proyecto es extraer datos de un archivo CSV que contiene información sobre títulos de Netflix, transformar estos datos para enriquecerlos y normalizarlos, y cargarlos en una base de datos MongoDB para su análisis y visualización.

Etapas del Proyecto

1. Extract (Extracción):
 - Tarea: Leer datos de un archivo CSV que contiene información sobre títulos de Netflix.
 - Herramienta: Utilizamos `fs` (file system) y `csv-parser` para leer y analizar el archivo CSV.
 - Archivo: `netflix_titles.csv`.
2. Transform (Transformación):
 - Tarea: Limpiar y enriquecer los datos extraídos. Esto incluye:
 - Transformar la duración de los títulos a minutos.
 - Calcular la antigüedad del contenido.
 - Convertir fechas a formato `Date` de JavaScript.
 - Herramienta: Transformaciones se realizan en el mismo script ETL utilizando JavaScript.
3. Load (Carga):
 - Tarea: Cargar los datos transformados en una base de datos MongoDB.
 - Herramienta: Utilizamos el driver oficial de MongoDB para Node.js (`mongodb`).
4. Visualización:
 - Tarea: Mostrar los datos cargados en una interfaz web, permitiendo al usuario ver y analizar la información de los títulos de Netflix.
 - Herramienta: Utilizamos `Express` para crear un servidor web y `EJS` como motor de plantillas para renderizar la vista.

Flujo del Proyecto

1. Extracción y Transformación de Datos:
 - El script `etl.js` lee el archivo `netflix_titles.csv` y procesa cada fila del archivo, transformando los datos según sea necesario (por ejemplo, convertir duraciones a minutos y calcular la antigüedad).
2. Carga de Datos:
 - Los datos transformados se insertan en la colección `titles` de la base de datos `netflix` en MongoDB. Usamos `Promise.all` para

asegurarnos de que todas las inserciones se completen antes de cerrar la conexión a la base de datos.

3. Servidor Web para Visualización:

- El servidor `index.js` se ejecuta utilizando Express, conectándose a la base de datos MongoDB para recuperar los datos cargados.
- Se renderiza una vista utilizando EJS, mostrando los datos en una tabla en una página web.

Ejecución del Proyecto

1. Ejecutar el Script ETL:

- Ejecuta `etl.js` para procesar y cargar los datos en MongoDB.

```
node etl.js
```

2. Ejecutar el Servidor Express:

- Ejecuta `index.js` para iniciar el servidor web y visualizar los datos.

```
node index.js
```

3. Ver los Datos en la Web:

- Abre un navegador y ve a `http://localhost:3000` para ver los datos de Netflix en una tabla.

Beneficios del Proyecto

- **Automatización del Proceso ETL:** Permite procesar automáticamente grandes volúmenes de datos desde un archivo CSV y cargarlos en una base de datos.
- **Enriquecimiento y Normalización de Datos:** Los datos se transforman para ser más útiles y consistentes, facilitando el análisis posterior.
- **Visualización de Datos:** Proporciona una interfaz web interactiva para ver y analizar los datos cargados.

Potenciales Mejoras

- **Paginación y Filtros:** Implementar paginación y filtros en la interfaz web para manejar grandes volúmenes de datos y permitir búsquedas específicas.
- **Gráficos y Análisis Avanzado:** Añadir gráficos y otras herramientas de visualización para realizar análisis más avanzados de los datos.
- **Automatización y Programación:** Programar el script ETL para que se ejecute automáticamente en intervalos regulares, manteniendo los datos actualizados.

En conclusión, el proyecto Netflix ETL proporciona un flujo completo desde la extracción de datos de un archivo CSV, pasando por la transformación y limpieza de datos, hasta la carga y visualización de estos en una base de

datos MongoDB, ofreciendo una base sólida para el análisis de datos de Netflix.