

Prueba técnica - Ingeniero de datos

Fecha:

Nombre:

Cédula:

La siguiente prueba técnica es abierta, los entregables corresponden al código fuente, un README con el detalle de implementación y justificación de diseño y un informe cuyo contenido se detalla más adelante. Esto se debe entregar en un repositorio git a elección.

Fecha límite de entrega: Lunes, 02 de diciembre de 2024, hora: 12:00 medio día

¡Muchos éxitos!

Descripción:

Se ha suministrado un archivo llamado **Films2.xlsx** donde encontrará el MER (Modelo Entidad - Relación) compuesto por 5 tablas, en cada pestaña del archivo se encuentra la información de cada una de las tablas.

Tener en cuenta el archivo *Films2* para formular preguntas relevantes que puedas responder utilizando las tablas del archivo. Adicionalmente, tener en cuenta los siguientes requisitos:

1. Implementar una aplicación que permita gestionar y ejecutar ETLs utilizando Python como lenguaje de programación. Esta aplicación debe ser escalable e implementar módulos de observabilidad. El paradigma de programación es a elección, pero es preferible POO.
2. Se deben implementar los ETL en esta aplicación usando un framework de procesamiento de datos (preferible Spark).
3. (Opcional) Desplegar la prueba en herramientas de la nube (GCP, AWS, Azure).
4. Realizar informe de presentación de resultados (max 3 páginas):
 1. Explicación de la arquitectura de datos y arquetipo de la aplicación.
 2. Análisis exploratorio de datos.
 3. Formular 5 posibles preguntas de negocio a las cuales los datos procesados puedan dar respuesta. Responde las preguntas.
 4. Conclusiones.

Nota: Es importante verificar la información, con el propósito de comprobar su consistencia. Dado el caso de encontrar irregularidades, por favor realizar un proceso de limpieza.