



Reto Técnico:

Importación y Exportación de Datos Masivos con Python y SQL en Docker

Objetivo

Evaluar la capacidad del aspirante para trabajar con grandes volúmenes de datos utilizando Python y Pandas, así como su habilidad para integrarse con una base de datos SQL ejecutada en un entorno Docker. El reto también evalúa buenas prácticas en estructuración de scripts, uso de herramientas modernas y comprensión del flujo de datos.

Contexto

Una empresa recibe archivos CSV con miles de registros que deben ser importados eficientemente a una base de datos SQL para su posterior análisis y consulta. Se requiere que el candidato desarrolle una solución reproducible, portable y funcional para cumplir con ese flujo de trabajo.

Requisitos Técnicos

1. Base de datos

- Utilizar PostgreSQL o MySQL, ejecutándose dentro de un contenedor Docker.
- Exponer el puerto correspondiente (5432 para PostgreSQL o 3306 para MySQL).
- Crear una base de datos llamada `crm` y una tabla `clientes` (`id` (entero, clave primaria); `name` (texto); `email` (texto); `country` (texto); `created_at` (fecha); `status` (enum)):

2. Dataset

- Generar o utilizar un archivo CSV llamado `clientes.csv` que contenga al menos 1,000 registros simulados.
- El archivo debe incluir datos variados y razonables para las columnas mencionadas.

3. Script en Python

- Leer el archivo `clientes.csv` utilizando `pandas.read_csv`.
- Insertar los datos en la tabla `clientes` utilizando `pandas.to_sql` o un método equivalente mediante SQLAlchemy.



- Consultar desde la base de datos todos los registros donde el país sea "Colombia".
- Exportar el resultado de esa consulta a un nuevo archivo llamado `clientes_colombia.csv`.

4. Entregables

- Archivo `docker-compose.yml` que levante el contenedor con la base de datos.
- Script o notebook en Python que realice las operaciones descritas.
- Archivo CSV de salida con los registros filtrados.
- (Opcional) Archivo `.env` para gestión de variables de entorno.
- Instrucciones breves de ejecución en un archivo `README.md`.

Criterios de Evaluación

- Correcta implementación de la lógica de importación y exportación.
- Uso adecuado de pandas, SQLAlchemy, y conexión a base de datos.
- Manejo básico de Docker y docker-compose.
- Limpieza, claridad y estructura del código.
- Buenas prácticas: control de errores, uso de entorno virtual, manejo de dependencias, comentarios, etc.
- Reproducibilidad del entorno y facilidad de ejecución del reto.