

## Descripción de la Necesidad

La compañía transnacional “Data Ingestion” le contrata como un Outsourcing de Ingeniería de Datos para solventar una necesidad puntual con relación al manejo de datos dentro de una infraestructura On-Premise, una necesidad en Python y una modernización a Nube.

A continuación, se definen las necesidades puntuales a trabajar:

1. Actualmente la empresa cuenta con un Modelo Relacional de Base de Datos que tiene la información de las Ventas (Ordenes de Ventas), Productos, Clientes, Ubicación, entre otros y el Modelo Relacional está en el archivo llamado “[Modelo Relacional Sistema Ordenes.jpg](#)”.

De este modelo se le pide que realice una consulta en SQL que permita identificar el top 10 de productos que más compran los clientes, adicionalmente se desea visualizar el promedio de venta de cada producto, la desviación estándar y la fecha de la venta más reciente, a continuación, se muestra un ejemplo ilustrativo de cómo se desean visualizar los datos (los datos en el ejemplo son con fines ilustrativos).

product_id	product_name	quantity	avg_total	stdev_total	recent_order_date
100	Sofá	1200	700	100	2024-02-07
90	Televisor	1191	800	130	2024-05-30
120	Lavadora	1087	400	87	2024-01-02

En el archivo “[Muestra Datos.zip](#)” encontrará una muestra de los datos que tiene cada tabla del Modelo Relacional.

Algunos puntos para tomar en cuenta:

- Toda tupla de encabezado de orden de venta (SalesOrderHeader) tiene asociado un detalle de orden de venta (SalesOrderDetail).
- Se desea obtener únicamente los datos de clientes físicos.
- Existen información de ordenes de ventas desde el 2014 hasta el 2024, a pesar de que en los archivos de muestras solo vienen datos de unos pocos años.

2. La empresa necesita consumir un API y analizar su respuesta, este contiene las ordenes realizadas por los clientes en los últimos 60 minutos, la tienda opera en 6 países de la región con diferente tipo de moneda, por lo que a continuación se comparte la tabla de tipos de cambio que se debe utilizar, en base a estos datos quieren analizar el comportamiento de los clientes y los productos más vendidos.

Con base en el enunciado, completar lo siguiente:

- Consumir y obtener el json del API, en su Endpoint Ordenes (<https://ca0589a4-4f03-493a-bd52-56f4afeb7625.mock.pstmn.io/orders>), el mismo está en un servidor no seguro, por lo que la validación SSL debe deshabilitarse.
- Aplicar la conversión de monto para dolarizarlo según la tabla de tipo de cambio.
- Mostrar un resumen con el total de Productos Vendidos y Monto Total por País.
- Mostrar un resumen con el Top 3 Categorías por país por Cantidad de Productos Vendidos, adicional agregar el Monto Total de esas Categorías.

La tabla de tipo de cambio es la siguiente:

País	Moneda	Tasa a USD
El Salvador	USD	1
Panamá	PAB	1
Costa Rica	CRC	0.0019
Nicaragua	NIO	0.027
Honduras	HNL	0.041
Guatemala	GTQ	0.13

Según los puntos indicados lo que se solicita es un script de Python que debería dar como resultado lo siguiente:

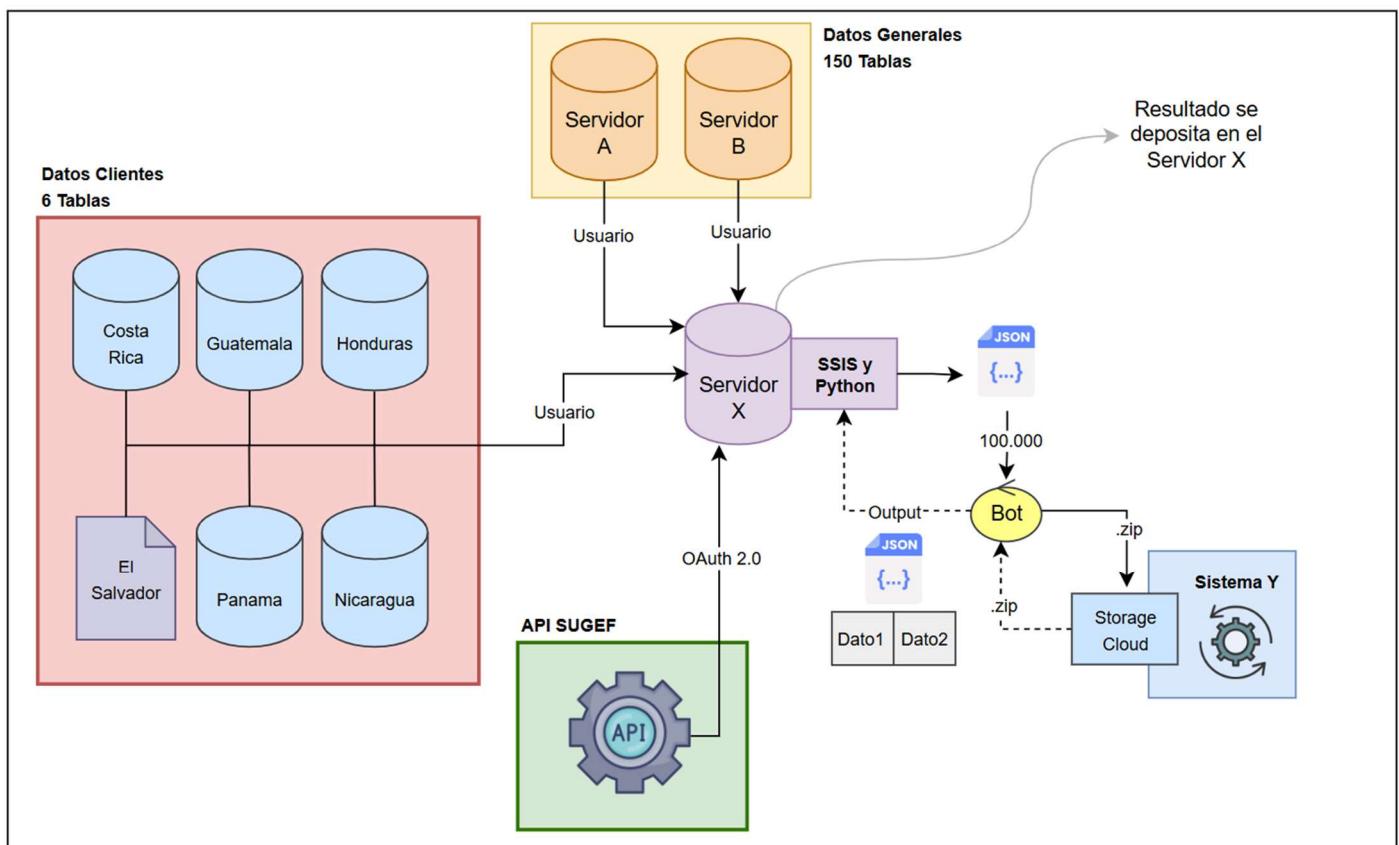
Resumen por país:			
país_envio	total_ordenes	monto_total_usd	
Costa Rica	126	\$823,804.78	
El Salvador	77	\$643,513.93	
Guatemala	235	\$1,863,072.66	
Honduras	297	\$2,558,834.81	
Nicaragua	112	\$908,675.70	
Panama	153	\$1,267,353.79	

Top 3 categorías por país:			
país_envio	category	cantidad_total	monto_total_usd
Costa Rica	Food - Snacks	124	\$75,563.36
Costa Rica	Pets	75	\$59,459.10
Costa Rica	Food - Condiments	73	\$47,271.00
El Salvador	Kitchen	79	\$57,176.41
El Salvador	Food - Condiments	58	\$37,612.24
El Salvador	Food - Beverages	47	\$20,577.86
Guatemala	Food - Snacks	203	\$157,906.06
Guatemala	Food - Condiments	197	\$158,052.40
Guatemala	Kitchen	180	\$101,627.11
Honduras	Food - Snacks	334	\$238,350.93
Honduras	Kitchen	285	\$248,585.13
Honduras	Food - Condiments	185	\$171,433.16
Nicaragua	Food - Snacks	102	\$75,526.80
Nicaragua	Kitchen	72	\$41,987.05
Nicaragua	Outdoor	71	\$70,029.61
Panama	Kitchen	179	\$105,535.48
Panama	Home	125	\$81,254.79
Panama	Food - Snacks	117	\$79,568.15

3. La empresa tiene un proceso crítico donde centraliza sus fuentes de datos y los datos del API de SUGEF en el Servidor X, genera un Dataset y lo envía mediante un archivo .json a un storage cloud del Sistema Y, el cual aplica un modelo y retorna otro archivo .json con los datos 1 y 2 calculados y por último el Servidor X procesa y valida el archivo .json y disponibiliza el dato en una tabla para el consumo de negocio.

El área desea que el proceso modernizado utilice los servicios en la nube que ya se encuentran desplegados en la empresa, estos son: un Storage Gen2, Data Factory y Databricks, adicionalmente el equipo nube ha indicado que se debe usar una Arquitectura Medallón y Data Vault para modelar los datos en una de las capas.

- ¿Cómo adaptarías el modelo/arquitectura híbrido actual a un proceso modernizado en nube incorporando los recursos actualmente desplegados?
- ¿Qué otros recursos solicitarías desplegar y por qué?
- ¿Qué otra propuesta o solución innovadora aplicarías en la modernización?



## **Entregables**

1. Consulta SQL que solventa la necesidad del Punto 1, se puede apoyar en la página: <https://www.mycompiler.io/es/new/sql> o bien con alguna herramienta de SQL que disponga.
2. Script de Python que denote la solución de la problemática descrita, puede utilizar cualquier IDE de Python, se recomienda Visual Studio Code, Bloc de Notas, o algún compilador de Python como <https://www.mycompiler.io/es/new/python>.
3. Diagrama y explicación general de la propuesta de modernización del proceso On-Premises a Nube, se recomienda utilizar <https://app.diagrams.net/>, Paint, Visio o bien utilizar cualquier herramienta de diseño que disponga.

## **Reglas y Consideraciones Importantes**

1. Al ser una prueba de un proceso de contratación, el uso de herramientas de IA Generativa (como Copilot, ChatGPT y otras) quedan prohibidas.
2. Excluyendo las herramientas mencionadas en el punto 1, se puede usar internet como apoyo (Buscar en Google).
3. El tiempo destinado para la resolución del caso de uso es de 1 hora, posteriormente se debe realizar la defensa.
4. En todo el momento debe de compartir pantalla (escritorio) y se debe validar que la configuración de pantalla indique “Duplicado” para evitar configuraciones de múltiples pantallas.
5. El Participante debe de tener la cámara encendida en todo momento.
6. La Prueba será grabada para documentar el proceso.
7. Para la defensa se tendrá 10 minutos de exposición.
8. Durante la defensa se realizarán algunas preguntas.
9. Se vale completamente cualquier idea o propuesta creativa e innovadora que amplie el alcance de los puntos de la prueba.