

# Data Architect Zenpli Challenge

**Objetivo:** El objetivo de esta prueba es evaluar tus habilidades en la captura de requerimientos, diseño, arquitectura, procesamiento, transformación y visualización de datos, así como tu enfoque hacia la calidad de los datos.

**Nota:** Puede elegir el stack tecnológico de nube que desee.

**Descripción del problema:** Se te ha proporcionado un conjunto de datos que contiene información sobre diferentes columnas que tienen tipos de datos de diferente naturaleza.

La tarea tiene 2 partes:

**Parte 1:** Modelar un pipeline de datos que ejecuta los siguientes pasos utilizando el dataset que le fue compartido:

Paso 1 (ejecución en paralelo)

- Verificar el esquema de datos;
- Verificar datos perdidos/faltantes;
- Verificar datos erróneos/anómalos;

Paso 2 (ejecución secuencial)

- Normalizar una columna (cualquiera de valores continuos);
- Filtrar una columna por cierto valor (cualquiera de valores categóricos);
- Agrupar ciertas columnas (cualquiera que correspondan a fechas).

Paso 3 (ejecución secuencial)

- Transformar una variable y agregarla al conjunto de datos. (Aplique la función  $x^3 + \exp(y)$  sobre cualquier tupla de variables continuas);
- Agregación - Conteo de registros únicos (sobre cualquier columna de valores categóricos).

Paso 4 (visualización)

- Cree alguna visualización que muestre detalles sobre calidad de datos.

**Parte 2:** Defina y diagrame una arquitectura de datos para una plataforma event-driven que recibe datos mediante una API, pasan por un Kafka. Puede usar algún patrón arquitectónico y los componentes que desee.

**Importante:** Suma puntos un enfoque de gobernanza, disponibilidad, seguridad y compliance