

DESCRIPCIÓN DEL PROBLEMA

La empresa UNALWater, con sede en Medellín y dedicada a la venta de botellas de agua, está buscando implementar un área especializada en arquitecturas Big Data debido a sus altos volúmenes de datos, con el objetivo de potenciar sus capacidades analíticas. Para iniciar este piloto, desean realizar una prueba inicial utilizando datos simulados para identificar las comunas de Medellín con mayores ventas, con el fin de optimizar sus estrategias de marketing. El objetivo es utilizar datos de transacciones diarias para determinar las áreas de mayor demanda de botellas de agua y mejorar la productividad de la empresa mediante el análisis de métricas clave.

Por esta razón, UNALWater ha contratado sus servicios para llevar a cabo una Prueba de Concepto (PoC). Uno de sus requerimientos iniciales es que no se utilizarán datos reales proporcionados por la empresa; en su lugar, se simularán eventos en tiempo real con el formato especificado:

```
{
  "latitude":6.291649401918482,
  "longitude":-75.60110546619606,
  "date":"12/05/2024 10:43:19",
  "customer_id":1888,
  "employee_id":9438,
  "quantity_products":23,
  "order_id":"d8b9b417-b098-4344-b137-362894e4dcccb"
}
```

Nota: Es importante aclarar que todos los procesos y datos nuevos que se generen en las diferentes zonas de almacenamiento partirán del cruce del ejemplo anterior con otras fuentes de información.

UNALWater tiene algunas consideraciones importantes que deben tenerse en cuenta:

- El objetivo es que las latitudes y longitudes estén dentro del rango de la ciudad de Medellín; los demás datos pueden ser simulados según las reglas definidas para demostrar valor al presentar la PoC.
- Se espera que se utilicen formatos y técnicas adecuadas para el almacenamiento de datos.
- Se requiere elaborar diagramas de arquitectura para entender cómo fue diseñada la solución.

- Es necesario que la información guardada en todas las zonas sea accesible mediante SQL.
- La simulación de los datos debe realizarse cada 30 segundos, con la posibilidad de ajustar estos intervalos según sea necesario. Se espera gestionar un proceso en tiempo real para ejecutar estas acciones.
- Se requiere realizar procesos orquestados, es decir, programas que puedan generar configuraciones y activar otros programas.
- Al procesar el evento en una arquitectura Medallion(bronze, silver, gold), para la zona bronze se esperan los siguientes campos: partition_date, order_id, commune, customer_id, employee_id, event_date, event_day, event_hour, event_minute, event_month, event_second, event_year, latitude, longitude, neighborhood, quantity_products. Para las zonas silver y gold, la decisión está a cargo de los ingenieros responsables. Ejemplo:

```
"partition_date": "14052024",  
"order_id": "65c477ea-937c-4338-a903-dc42901afacd",  
"commune": "COMUNA 10",  
"customer_id": 7530,  
"employee_id": 1114,  
"event_date": "14/05/2024 11:11:26",  
"event_day": 14,  
"event_hour": 11,  
"event_minute": 11,  
"event_month": 5,  
"event_second": 26,  
"event_year": 2024,  
"latitude": "6.251188000704114",  
"longitude": "-75.57616810955338",  
"neighborhood": "LA CANDELARIA",  
"quantity_products": 17
```

- Se esperan adiciones creativas para evitar problemas futuros.

FUENTES DE DATOS

Para esta investigación se proporcionarán los polígonos de:

- La ciudad de Medellín
- Las comunas de Medellín

CRITERIOS DE EVALUACIÓN

- El profesor realizará preguntas asumiendo varios roles (técnicos, usuarios funcionales, administrativos, etc.) en un lapso de 5 minutos.

- El trabajo no es requerido entregarse, pero se debe evidenciar el código en formato notebook (no Google Colab o Jupyter) o scripts de Python y/o Linux. Se evaluará principalmente la capacidad de análisis, generación de información nueva a partir de los datos, conocimiento del negocio (respuesta a las preguntas) y la habilidad para transmitir el nuevo conocimiento generado (documentación y exposición).
- Al estar simulando un problema típico de la vida real, las empresas normalmente solo quieren ver en qué se ha invertido, documentación de guía, cómo lo hicieron y, principalmente, ver cómo funciona. Por esta razón, los entregables o la documentación deberán ser definidos por el equipo de trabajo.

RECOMENDACIONES

- Es importante tener en cuenta que algunos componentes pueden no ser claros desde el inicio, ya que requieren investigación o se integran en fases diferentes. Se recomienda abordar estrategias modulares para avanzar adecuadamente.
- Las Pruebas de Concepto (PoC) se realizan para demostrar la viabilidad de una idea o solución en un entorno controlado. Ayudan a identificar problemas, validar la funcionalidad y evaluar el rendimiento. Además, permiten a las empresas evaluar su capacidad para llevar a cabo el proyecto y la calidad del análisis, desarrollo e ideas, de los cuales dependerá el éxito y la continuidad del proyecto.
- Trabajo en equipo es la clave