MODULO 08 - EJERCICIO 06-A

ALEXIS YURI M.

Diseña un flujo de datos para procesar altos volúmenes de información de ventas multicanal utilizando una plataforma de automatización. Deberás contemplar las fuentes de entrada, etapas de procesamiento, reglas de decisión y destino de los datos, aplicando buenas prácticas.

1. Componentes del Flujo.

- Fuente de datos: La información proviene de múltiples canales de venta: e-commerce, puntos físicos, aplicaciones móviles y marketplaces.

- Procesador: Las transformaciones a realizar son:

Limpieza de datos: Eliminar valores nulos o caracteres incorrectos.

Normalización: Estandarizar formatos (ej. fechas, monedas).

Enriquecimiento: Añadir información de contexto, como la ubicación de la tienda

física o el tipo de canal de venta.

- Enrutador o decisor: Se aplicarán reglas de enrutamiento basadas en la calidad de los datos. Por ejemplo, los datos limpios y estandarizados se envían al destino principal, mientras que los registros con errores se desvían a un sistema de registro de errores para su revisión.

- Destino: Los datos deben almacenarse en un Data Warehouse (DWH) para el análisis de negocio y la generación de reportes consolidados.

2. Plataforma de Automatización.

En este caso seleccionaría Apache NiFi como la plataforma de automatización porque es una herramienta de código abierto que permite diseñar flujos de datos de manera visual y flexible, lo que es ideal para un caso con fuentes de datos diversas y formatos variables. Además, su capacidad para manejar la ingesta de grandes volúmenes de datos y aplicar reglas de enrutamiento la hace elegible para este tipo de flujo multicanal.

3. Beneficios Esperados.

- Reducción de errores: Al automatizar el proceso de integración, se eliminan los errores humanos y los cuellos de botella del procesamiento manual.

- Disponibilidad en tiempo real: Aunque se procesa por lotes diarios, los datos pueden estar disponibles de forma más rápida que con un proceso manual, lo que permite un análisis más oportuno.

- Estandarización del proceso: Se garantiza que los datos de todos los canales de venta tengan el mismo formato y calidad antes de ser almacenados.

- Monitoreo Continuo: La plataforma de automatización permite un seguimiento y una supervisión constante del flujo de datos, lo que facilita la identificación de problemas y la gestión de fallos en tiempo real. Este monitoreo garantiza la fiabilidad y la resiliencia del proceso de integración.

4. Propuesta de Flujo Paso a Paso.

Extracción de Datos: Un procesador de NiFi se configura para extraer los datos de cada canal de venta (e-commerce, puntos físicos, etc.). Este procesador puede leer desde APIs, bases de datos o archivos CSV.

Validación y Limpieza: Los datos se pasan a un grupo de procesadores que aplican reglas para limpiar y validar la información (ej. eliminar registros duplicados o con campos incompletos).

Enrutamiento Condicional: Los datos limpios se enrutan a la siguiente fase, mientras que los datos con errores se dirigen a un flujo de manejo de errores para su revisión.

Transformación y Normalización: Un procesador de NiFi unifica el formato de las fechas, los montos y los nombres de las columnas para asegurar la consistencia de los datos entre todos los canales.

Carga en el Destino: Los datos ya estandarizados se cargan en el Data Warehouse de la empresa.

Monitoreo: El equipo de datos puede monitorear el flujo en tiempo real desde la interfaz de NiFi, garantizando que el proceso se ejecute de manera exitosa y sin interrupciones.