

Universidad Yachay Tech

Maestría en Ciencia de Datos - Taller

Título: Diseño de Arquitectura de Datos para un E-Commerce

Objetivos:

1. Entender y aplicar teoría de arquitectura de datos.
2. Identificar elementos importantes de la arquitectura del sistema
3. Visualizar la estructura en un documento HLD (High Level Design)

Escenario

Eres parte del equipo de arquitectura de datos de una empresa de comercio electrónico de tamaño medio, Yachay-Shop, que opera en múltiples canales de venta (sitio web, aplicación móvil y tiendas físicas). La empresa ha crecido rápidamente y ahora enfrenta desafíos en la gestión y el análisis de grandes cantidades de datos provenientes de diferentes sistemas. Necesitas diseñar una nueva arquitectura de datos que ayude a la empresa a integrar, almacenar, procesar y analizar mejor los datos para la toma de decisiones.

Sistemas Actuales

1. Sitio Web y Aplicación Móvil: Generan datos sobre el comportamiento del cliente (clics, tiempo en páginas, productos visualizados), transacciones en línea y perfiles de usuario.
2. Tiendas Físicas: Datos de transacciones desde los sistemas de punto de venta (POS), actualizaciones de inventario y uso de tarjetas de fidelización.
3. CRM (Gestión de Relaciones con los Clientes): Administra perfiles de clientes, historial de pedidos y campañas de marketing.
4. Sistema de Gestión de Inventarios: Rastrea los niveles de existencias en los almacenes y tiendas, proporcionando actualizaciones en tiempo real sobre la disponibilidad de productos.
5. Análisis de Marketing Externo: Datos de herramientas de marketing de terceros como Google Analytics y Facebook Ads, que proporcionan métricas sobre el rendimiento de campañas.

Desafíos

- Silos de Datos: Cada sistema está actualmente aislado, lo que dificulta obtener una visión completa del comportamiento del cliente, el rendimiento de los productos y las métricas generales del negocio.
- Volumen y Variedad de Datos: Los datos vienen en varios formatos (estructurados, semiestructurados y no estructurados), incluyendo transacciones de ventas, registros de usuarios, métricas de redes sociales y detalles de inventario.
- Procesamiento en Tiempo Real: Se necesita información en tiempo real sobre los niveles de inventario, el compromiso de los clientes y las tendencias de ventas, especialmente durante campañas de marketing o ventas flash.

Necesidades del Negocio

- Vista 360 del Cliente: Integrar datos de todos los canales para obtener una visión completa de las interacciones y el historial de compras de cada cliente.
- Actualizaciones de Inventario en Tiempo Real: Asegurarse de que los niveles de existencias se actualicen en todas las plataformas (sitio web, aplicación móvil, tiendas) en tiempo real.
- Eficacia del Marketing: Analizar la efectividad de las campañas de marketing vinculando los datos externos de marketing (anuncios, redes sociales) con las ventas y el comportamiento de los clientes.
- Pronósticos y Reportes de Ventas: Generar informes diarios, semanales y mensuales sobre ventas, tendencias de clientes y rendimiento de productos, además de análisis predictivos sobre tendencias futuras.

Tu Tarea

- Diseñar una Arquitectura de Datos que integre los datos de los sistemas mencionados, maneje tanto el procesamiento por lotes como en tiempo real, y permita un almacenamiento, consulta y análisis eficientes.
- Considera si utilizarías un almacén de datos, un data lake, o una combinación de ambos.
- Define los métodos de ingestión para cada fuente de datos y los pasos de transformación necesarios para unificar los datos.
- Describe cómo se almacenarán los datos (por ejemplo, estructurados vs. no estructurados), cómo se procesarán y cómo se pondrán a disposición para herramientas de inteligencia empresarial.
- Asegúrate de diseñar el sistema de manera que pueda escalar con el aumento en el volumen de datos.