

### Actividad 3 Módulo 5

1. Se propone una estructura híbrida que combine:
  - a. Data lake : Para datos sin estructura, de alta variedad y volumen.
  - b. Data warehouse: Para análisis histórico y datos estructurados ya depurados y transformados.
  - c. Data Marts: Para acceso a información específica por área de negocio.

Esta arquitectura nos permite aprovechar las fortalezas de cada solución, también cubrir todos los requerimientos técnicos y analíticos.

Almacena grandes volúmenes de datos brutos y variados, por otro lado disponer de información procesada estructurada y optimizada

2. De acuerdo con la necesidad se establece que los datos se almacenarán de la siguiente manera:
  - a. Data lake:
    - i. Ventas online
    - ii. Sensores IoT
    - iii. Campaña de marketing de RRSS
    - iv. Imágenes y videos
    - v. Archivos semiestructurados (JSON,XML,CSV)
  - b. Data warehouse:
    - i. Datos de clientes
    - ii. Datos de stock físico
    - iii. Datos de transacciones de ventas
    - iv. Métricas de campaña de marketing directo (Puntos de venta, tiendas)
    - v. KPIs Corporativos
  - c. Data Marts:
    - i. Data Mart de Ventas (Metricas de ventas por región, tienda, periodo,etc)
    - ii. Data Mart de Operaciones (Gestión de inventario y logística)
    - iii. Data Mart de Marketing (Histórico de campañas de marketing, rendimiento de campaña ya ligado a resultados respecto del stock)
    - iv. Data Mart de Finanzas (Rentabilidad, márgenes, etc)

3. Perfiles de usuarios que consumirá los datos:

- a. Data lake:
  - i. Científicos de datos (Análisis de datos y modelos predictivos)
  - ii. Ingenieros de datos (Desarrollo de pipelines de procesamiento)
  - iii. Equipos de Tecnología de la Información
- b. Data warehouse:
  - i. Analistas de BI (Creación de reportes consolidados y análisis históricos)
  - ii. Gerencia general (Para análisis de KPIs y decisiones estratégicas)
- c. Data Marts:
  - i. Jefes de ventas (Métricas de desempeño)
  - ii. Gerentes de tiendas (Análisis de Stock, rotación y performance del local)
  - iii. Equipo de marketing (Desempeño de campañas)
  - iv. Gerencia de finanzas (Análisis de rentabilidad y costos)

4. ¿Qué herramientas o capas deben considerarse?

- a. Capa operacional (Captura de los datos desde diferentes fuentes)
  - i. Sensores IoT
  - ii. RSS
  - iii. ERP y CRM
  - iv. Sistemas contables
- b. Capa de Data Lake
  - i. Almacenamiento crudo y flexible de datos (Amazon S3, Azure Data Lake)
- c. Capa de staging (Recibir los datos y prepararlos para su procesamiento)
  - i. Validación
  - ii. Corrección de errores
  - iii. Normalización de formatos
- d. Capa Data Warehouse
  - i. Almacenamiento estructurado y consolidado (Snowflake, BigQuery, dbt, Airflow)
- e. Capa Data Marts (Segmentación por áreas de negocio)  
se podría alinear de un modelo dependiente de manera que a través del DW segmentando la información por área de negocios, generando campañas de marketing producto del estudio de diferentes métricas, finanzas, y ventas sea cual sea el caso
- f. Capa Herramienta BI (Visualización y análisis para usuarios finales)
  - i. Power BI para dashboards de los diversos datos estudiados

- ii. Grafana para monitoreo en tiempo real de sensores IoT

- g. Gobernanza y Metadatos:

- i. Control, catalogación y calidad de datos (AWS Lake Formation)

- 5. Beneficios para el negocio:

- a. Toma de decisiones más rápida y eficaz

- b. Escalabilidad y flexibilidad

- c. Autonomía analítica por área de negocio

- d. Marketing personalizado basado en datos de comportamiento