MODULO 05 - EJERCICIO 02

ALEXIS YURI M.

Mapa Comparativo: Data Lake, Data Warehouse y Data Mart

1. Comparación por aspectos clave

*Data Lake*

Tipo de datos: Estructurados, semiestructurados y no estructurados (JSON, audio, logs).

Aplicación del esquema: Schema-on-read (esquema aplicado al leer los datos).

Casos de uso ideales: Ciencia de datos, machine learning, IoT, almacenamiento masivo crudo

Madurez técnica requerida: Alta: requiere conocimientos en big data, Python, Spark.

Costos y tiempo implementación: Bajo costo inicial, pero alto mantenimiento (datos en crudo, limpieza).

Usuarios típicos: Científicos de datos, analistas técnicos.

*Data Warehouse*

Tipo de datos: Principalmente estructurados (tablas, relaciones, OLAP).

Aplicación del esquema: Schema-on-write (estructura definida al ingresar los datos).

Casos de uso ideales: Inteligencia de negocios, reporting corporativo, KPI históricos.

Madurez técnica requerida: Media: experiencia en modelado, SQL, herramientas BI.

Costos y tiempo implementación: Alto costo de implementación, pero óptimo para análisis confiable.

Usuarios típicos: Analistas de negocio, ejecutivos, equipos de BI.

*Data Mart*

Tipo de datos: Estructurados, enfocados en un dominio específico del negocio.

Aplicación del esquema: Schema-on-write (estructura rígida definida por el negocio).

Casos de uso ideales: Análisis departamental: ventas, finanzas, RRHH.

Madurez técnica requerida: Baja a media: ideal para usuarios de negocio con soporte técnico básico.

Costos y tiempo implementación: Bajo a medio: implementación rápida y ágil para necesidades específicas.

Usuarios típicos: Usuarios departamentales, analistas operativos.

2. Casos reales ejemplificados

Data Lake – Sector salud (InfoHealth)

Se utiliza un Data Lake en la nube (como AWS S3 o Azure Data Lake) para almacenar señales crudas provenientes de sensores IoT, documentos PDF de exámenes, imágenes médicas y formularios clínicos. Estos datos son usados por científicos de datos para entrenar modelos predictivos.

Data Warehouse – Retail

Una cadena de supermercados implementa un Data Warehouse con Amazon Redshift para consolidar datos de ventas, inventarios y clientes, permitiendo el análisis histórico y planificación de campañas promocionales.

Data Mart – Finanzas (empresa mediana)

El área de finanzas de una empresa de logística crea un Data Mart exclusivo para analizar flujos de caja, presupuestos y desviaciones mensuales, utilizando Power BI conectado a una base SQL Server específica.

3. Conclusiones.

• Los tres enfoques tienen propósitos distintos y no son excluyentes. En una arquitectura moderna, pueden coexistir como parte de un ecosistema de datos integral.

• Data Lakes son altamente flexibles, ideales para exploración y ciencia de datos, pero requieren gobernanza y limpieza para no convertirse en 'data swamps'.

• Data Warehouses siguen siendo esenciales para reportes consistentes y decisiones estratégicas basadas en datos consolidados y confiables.

• Data Marts son una solución ágil para empoderar a áreas específicas del negocio con información relevante sin necesidad de acceder a todo el warehouse.

• La elección entre ellos depende del nivel técnico de los usuarios, el tipo de análisis, el volumen de datos y los objetivos del negocio.