

Mapa Comparativo: Data Lake, Data Warehouse y Data Mart

Contexto

Comprender las diferencias clave entre estos tres enfoques de almacenamiento es fundamental para poder seleccionar el más adecuado según los objetivos y el contexto de la organización.

Consigna

Completa un mapa comparativo que contraste Data Lake, Data Warehouse y Data Mart en al menos cinco aspectos clave: tipo de datos, esquema, casos de uso, costos y usuarios típicos.

Paso a paso

1. Completa una tabla que compare los tres enfoques en estos ejes:
 - Tipo de datos (estructurado, no estructurado, etc)
 - Aplicación del esquema (En escritura o lectura)
 - Caso de Uso Ideales
 - Nivel de madurez técnica necesario
 - Costos y tiempo de implementación
2. Agrega un caso real que ejemplifica el uso de cada uno.
3. Presenten conclusiones.

Aspecto	Data Lake	Data Warehouse	Data Mart
---------	-----------	----------------	-----------

Tipo de datos	Datos estructurados, semi-estructurados y no estructurados (datos en bruto)	Principalmente datos estructurados, limpios y organizados	Datos estructurados, subset del Data Warehouse
Aplicación del esquema	Esquema en lectura (schema-on-read): se interpreta al leer los datos	Esquema en escritura (schema-on-write): datos limpios al ingresar	Esquema en escritura, enfocado en área específica
Casos de uso ideales	Análisis exploratorio, machine learning, big data	Reporting empresarial, BI, análisis estructurado	Reportes específicos para departamentos o áreas
Nivel de madurez técnica necesario	Alto: gestión de grandes volúmenes y datos variados	Medio-alto: diseño de modelos y procesos ETL	Medio: solución más sencilla para usuarios finales
Costos y tiempo de implementación	Bajo costo inicial (almacenamiento barato), crece con volumen	Más costoso y complejo, implementación más larga	Menor costo y tiempo; solución rápida y focalizada

Casos reales

- **Data Lake:** Netflix almacena datos en bruto de usuarios y eventos para análisis avanzados y machine learning.
- **Data Warehouse:** Amazon Redshift es utilizado para almacenar datos limpios para reportes y análisis estructurados.
- **Data Mart:** Un departamento de ventas con un Data Mart para reportes rápidos de clientes y productos.

Conclusiones

- Los **Data Lakes** son adecuados para almacenar grandes volúmenes de datos variados, requieren mayor madurez técnica y ofrecen gran flexibilidad.
- Los **Data Warehouses** proveen datos limpios y organizados para análisis estructurados, con mayor inversión y complejidad.
- Los **Data Marts** son soluciones rápidas y específicas para áreas o departamentos, con menor costo y alcance limitado.

