

# Comparativo entre Data Lake, Data Warehouse y Data Mart

## Mapa comparativo entre 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:** 📝

En equipos, 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.

**Tiempo:** ⌚ 30 minutos

**Paso a paso:** 📌

1. En grupos, 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)
  - Casos de uso ideales
  - Nivel de madurez técnica necesario
  - Costos y tiempo de implementación
2. Agreguen un caso real que ejemplifica el uso de cada uno (pueden inspirarse en los del manual).
3. Presenten sus conclusiones en plenario con justificación.

 alkemy

Aspecto	Data Lake	Data Warehouse (DW)	Data Mart (DM)
Tipo de datos	Estructurados, semi-estructurados y no estructurados (logs, imágenes, videos, IoT).	Estructurados y altamente normalizados/desnormalizados para consulta.	Estructurados, especializados por área o dominio.
Aplicación del esquema	<i>Schema-on-read</i> (se define al leer los datos).	<i>Schema-on-write</i> (se define al cargar los datos).	<i>Schema-on-write</i> (enfocado al área específica).
Casos de uso ideales	Análisis exploratorio, ciencia de datos, machine learning, almacenamiento masivo.	Inteligencia de negocios, reportes corporativos, análisis históricos.	Análisis departamental (finanzas, marketing, operaciones) con foco en KPIs específicos.
Nivel de madurez técnica necesario	Medio-alto (requiere conocimientos en big data y procesamiento distribuido).	Medio (requiere conocimientos en modelado de datos y BI).	Bajo-medio (más sencillo si ya existe un DW).

<b>Costos</b>	Menor costo de almacenamiento, pero mayor costo de procesamiento y gestión.	Mayor costo de almacenamiento y licencias, optimizado para consultas rápidas.	Menor costo (infraestructura y mantenimiento reducidos).
<b>Tiempo de implementación</b>	Rápido para almacenar datos; más tiempo si se requiere estructuración.	Mayor tiempo por el modelado y ETL.	Menor tiempo si deriva de un DW existente.
<b>Ejemplo real</b>	<b>Netflix:</b> almacena logs, datos de streaming, interacciones de usuario para análisis y personalización.	<b>Amazon:</b> utiliza Redshift para análisis de ventas globales y reportes corporativos.	<b>Banco Galicia:</b> DM financiero para análisis de indicadores de créditos y cartera.