

alumno:Salvatierra De La Cruz Jesus

Cuadro Comparativo: Data Lake vs Data Warehouse vs Base de Datos NoSQL

Criterio	Data Lake	Data Warehouse	Base de Datos NoSQL
1. Estructura de Datos	Datos sin estructura definida (raw data)	Datos altamente estructurados con esquemas fijos	Flexible: documentos, clave-valor, grafos, columnas
2. Escalabilidad	Escalabilidad masiva horizontal (petabytes)	Escalabilidad vertical limitada	Escalabilidad horizontal excelente
3. Velocidad de Consulta	Lenta para análisis complejos	Muy rápida para consultas predefinidas	Rápida para operaciones simples
4. Costo de Almacenamiento	Muy bajo (almacenamiento commodity)	Alto (requiere hardware especializado)	Medio (depende del proveedor)
5. Tipos de Datos	Estructurados, semi-estructurados, no estructurados	Solo datos estructurados	Principalmente semi-estructurados

6. Tiempo de Implementación	Rápido (almacenar primero, esquema después)	Lento (requiere diseño previo)	Medio (configuración específica)
7. Complejidad de Mantenimiento	Media (requiere gobernanza de datos)	Alta (ETL, índices, optimización)	Baja a media
8. Herramientas de Análisis	Spark, Hadoop, Databricks, Python/R	SQL estándar, BI tools tradicionales	APIs específicas, algunos soportan SQL
9. Modelo de Consistencia	Eventualmente consistente	ACID completo	Varía: desde eventual hasta ACID
10. Casos de Uso Principales	Machine Learning, exploración de datos	Reportes empresariales, BI, dashboards	Aplicaciones web/móviles, tiempo real
11. Procesamiento	Batch y streaming	Principalmente batch	Tiempo real y transaccional
12. Esquema de Datos	Schema-on-read (esquema al leer)	Schema-on-write (esquema al escribir)	Schema-less o schema flexible
13. Gobernanza y Seguridad	Desafiante (datos diversos y raw)	Excelente (datos curados y controlados)	Buena (controles granulares)
14. Capacidad de Backup	Compleja (volúmenes masivos)	Estándar (herramientas maduras)	Buena (replicación automática)
15. Madurez Tecnológica	Emergente (últimos 10 años)	Muy madura (30+ años)	Madura (15+ años)

Resumen por Escenario de Uso

Usa Data Lake cuando:

- Necesites almacenar grandes volúmenes de datos diversos
- No sepas exactamente cómo usarás los datos
- Hagas machine learning o análisis exploratorio
- Tengas presupuesto limitado para almacenamiento

Usa Data Warehouse cuando:

- Necesites reportes empresariales consistentes

- Tengas consultas predefinidas y recurrentes
- Requieras alta velocidad en consultas complejas
- La gobernanza de datos sea crítica

Usa NoSQL cuando:

- Desarrolles aplicaciones web/móviles
- Necesites respuestas en milisegundos
- Los datos cambien de estructura frecuentemente
- Requieras alta disponibilidad y escalabilidad horizontal

Tendencias Actuales

Arquitecturas Híbridas:

- **Lakehouse:** Combina Data Lake + Data Warehouse
- **Multi-model:** Bases de datos que soportan múltiples paradigmas
- **Data Mesh:** Arquitectura descentralizada con múltiples tecnologías