Data Dilemma:

Decoding Warehouses, Lakes, and Lakehouses for Healthcare

Businesses today have three main options for structuring their data: data warehouses, which are designed for structured analytics and reporting; data lakes, which store raw and diverse data for flexibility and exploration; and data lakehouses, which integrate the strengths of both. Each approach serves different needs, but the lakehouse offers a versatile solution by combining flexibility with structure.

Data Warehouse

Think of a well-organized library where every book (data) is cataloged and placed on specific shelves (structured schema).

Data Type: Structured

Schema: Schema-on-Write

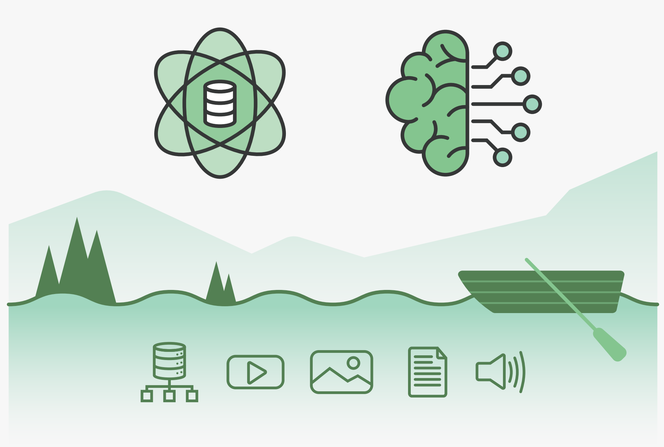
Use Case: Reporting and BI

Performance: Optimized for complex queries

Storage: Fixed schema, pre-defined

Data Lake

Imagine a vast, unorganized repository where all types of documents, books, and media are stored in their original form (raw data), and you decide how to organize them when you need them.



Data Type: Unstructured, semi-structured, structured

Schema: Schema-on-Read

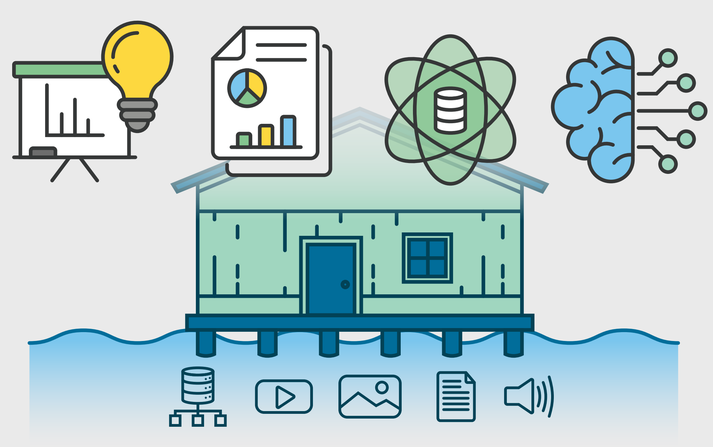
Use Case: Big data analytics, data science

Performance: Scalable storage, flexible data processing

Storage: Raw, native format

Data Lakehouse

Picture a hybrid space that combines the library's organization with the repository's flexibility, allowing you to access and organize books and media as needed for different purposes.



Data Type: Both structured and unstructured

Schema: Schema-on-Write and Schema-on-Read

Use Case: Unified analytics platform, BI, data science

Performance: High performance for varied workloads

Storage: Unified platform, flexible schema

HealthCatalyst®