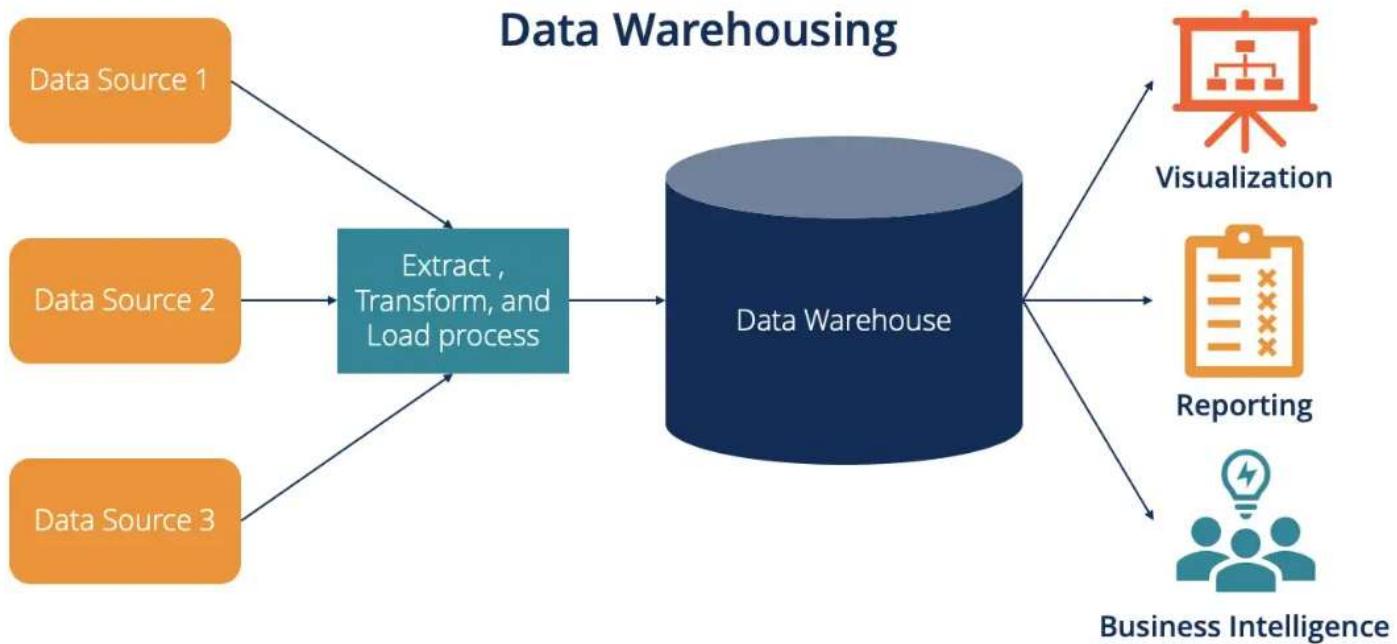


6_Data lake vs Data Warehouse vs Data Lakehouse

Tuesday, February 10, 2026 3:49 PM

What Is a Data Warehouse?

- A data warehouse is a unified data repository for storing large amounts of information from multiple sources within an organization.
- A data warehouse represents a single source of “data truth” in an organization and serves as a core reporting and business analytics component.
- Typically, data warehouses store historical data by combining relational data sets from multiple sources, including application, business, and transactional data.
- Data warehouses extract data from multiple sources and transform and clean the data before loading it into the warehousing system to serve as a single source of data truth.
- Organizations invest in data warehouses because of their ability to quickly deliver business insights from across the organization.
- Data warehouses enable business analysts, data engineers, and decision-makers to access data via BI tools, SQL clients, and other less advanced (i.e., non-data science) analytics applications.



The benefits of a data warehouse

Data warehouses, when implemented, offer tremendous advantages to an organization. Some of the benefits include:

- **Improving data standardization, quality, and consistency:** Organizations generate data from various sources, including sales, users, and transactional data. Data warehousing consolidates corporate data into a consistent, standardized format that can serve as a single source of data truth, giving the organization the confidence to rely on the data for business needs.
- **Delivering enhanced business intelligence:** Data warehousing bridges the gap between voluminous raw data, often collected automatically as a matter of practice, and the curated data that offers insights. They serve as the data storage backbone for organizations, allowing them to answer complex questions about their data and use the answers to make informed business decisions.
- **Increasing the power and speed of data analytics and business intelligence workloads:** Data warehouses speed up the time required to prepare and analyze data. Since the data warehouse's data is consistent and accurate, they can effortlessly connect to data analytics and business intelligence tools. Data warehouses also cut down the time required to gather data and give teams the power to leverage data for reports, dashboards, and other analytics needs.
- **Improving the overall decision-making process:** Data warehousing improves decision-making by providing a single repository of current and historical data. Decision-makers can evaluate risks, understand customers' needs, and improve products and services by transforming data in data warehouses for accurate insights.

For example, [Walgreens](#) migrated its inventory management data into Azure Synapse to enable supply

chain analysts to query data and create visualizations using tools such as Microsoft Power BI. The move to a cloud data warehouse also decreased time-to-insights: previous-day reports are now available at the start of the business day, instead of hours later.

The disadvantages of a data warehouse

Data warehouses empower businesses with highly performant and scalable analytics. However, they present specific challenges, some of which include:

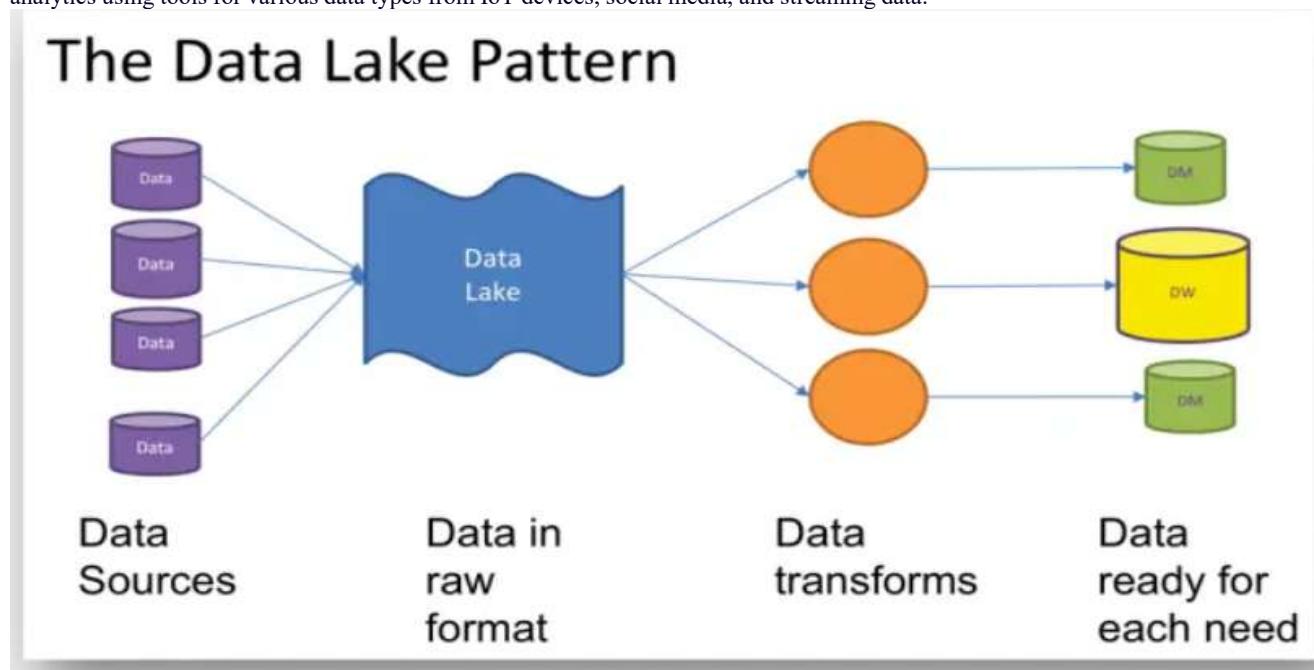
- **Lack of data flexibility:** Although data warehouses perform well with structured data, they can struggle with semi-structured and unstructured data formats such as log analytics, streaming, and social media data. This makes it hard to recommend data warehouses for machine learning and artificial intelligence use cases.
- **High implementation and maintenance costs:** Data warehouses can be expensive to implement and maintain. This article by Cooladata estimates the [annual cost of an in-house data warehouse](#) with one terabyte of storage and 100,000 queries per month to be \$468,000. Additionally, the data warehouse is typically not static; it becomes outdated and requires regular maintenance, which can be costly.

What Is a Data Lake?

A data lake is a centralized, highly flexible storage repository that stores large amounts of structured and unstructured data in its raw, original, and unformatted form. In contrast to data warehouses, which store already “cleaned” relational data, a data lake stores data using a flat architecture and object storage in its raw form. Data lakes are flexible, durable, and cost-effective and enable organizations to gain advanced insight from unstructured data, unlike data warehouses that struggle with data in this format.

In data lakes, the schema or data is not defined when data is captured; instead, data is extracted, loaded, and transformed (ELT) for analysis purposes. Data lakes allow for machine learning and predictive analytics using tools for various data types from IoT devices, social media, and streaming data.

The Data Lake Pattern



The benefits of a data lake

Because data lakes can store both structured and unstructured data, they offer several benefits, such as:

- **Data consolidation:** Data lakes can store both structured and unstructured data to eliminate the need to store both data formats in different environments. They provide a central repository to store all types of organizational data.
- **Data flexibility:** A significant benefit of data lakes is their flexibility; you can store data in any format or medium without the need to have a predefined schema. Allowing the data to remain in its native format allows for more data for analysis and caters to future data use cases.
- **Cost savings:** Data lakes are less expensive than traditional data warehouses; they are designed to be stored on low-cost commodity hardware, like object storage, usually optimized for a lower cost per GB stored. For example, [Amazon S3 standard object storage](#) offers an unbelievable low price of \$0.023 per GB for the first 50 TB/month.
- **Support for a wide variety of data science and machine learning use cases:** Data in data lakes is stored in an open, raw format, making it easier to apply various machine and deep learning

algorithms to process the data to produce meaningful insights.

The disadvantages of a data lake

Although data lakes offer quite a few benefits, they also present challenges:

- **Poor performance for business intelligence and data analytics use cases:** If not properly managed, data lakes can become disorganized, making it hard to connect them with business intelligence and analytics tools. Also, a lack of consistent data structure and ACID (atomicity, consistency, isolation, and durability) transactional support can result in sub-optimal query performance when required for reporting and analytics use cases.
- **Lack of data reliability and security:** Data lakes' lack of data consistency makes it difficult to enforce data reliability and security. Because data lakes can accommodate all data formats, it might be challenging to implement proper data security and governance policies to cater to sensitive data types.

What Is a Data Lakehouse? A Combined Approach

A data lakehouse is a new, big-data storage architecture that combines the best features of both data warehouses and data lakes. A data lakehouse enables a single repository for all your data (structured, semi-structured, and unstructured) while enabling best-in-class machine learning, business intelligence, and streaming capabilities.

Data lakehouses usually start as data lakes containing all data types; the data is then converted to [Delta Lake](#) format (an open-source storage layer that brings reliability to data lakes). Delta lakes enable ACID transactional processes from traditional data warehouses on data lakes.

The benefits of a data lakehouse

Data lakehouse architecture combines a data warehouse's data structure and management features with a data lake's low-cost storage and flexibility. The benefits of this implementation are enormous and include:

- **Reduced data redundancy:** Data lakehouses reduce data duplication by providing a single all-purpose data storage platform to cater to all business data demands. Because of the advantages of the data warehouse and the data lake, most companies opt for a hybrid solution. However, this approach could lead to data duplication, which can be costly.
- **Cost-effectiveness:** Data lakehouses implement the cost-effective storage features of data lakes by utilizing low-cost object storage options. Additionally, data lakehouses eliminate the costs and time of maintaining multiple data storage systems by providing a single solution.
- **Support for a wider variety of workloads:** Data lakehouses provide direct access to some of the most widely used business intelligence tools ([Tableau](#), [PowerBI](#)) to enable advanced analytics. Additionally, data lakehouses use open-data formats (such as Parquet) with APIs and machine learning libraries, including Python/R, making it straightforward for data scientists and machine learning engineers to utilize the data.
- **Ease of data versioning, governance, and security:** Data lakehouse architecture enforces schema and data integrity making it easier to implement robust data security and governance mechanisms.

The disadvantages of a data lakehouse

The main disadvantage of a data lakehouse is it's still a relatively new and immature technology. As such, it's unclear whether it will live up to its promises. It may be years before data lakehouses can compete with mature big-data storage solutions. But with the current speed of modern innovation, it's difficult to predict whether a new data storage solution could eventually usurp it.