



Manoj Kumar



Data Warehouse VS Lakehouse





Manoj Kumar

Caveat

This post provides a general overview of data warehouses and lakehouses, and is not intended to be a comprehensive resource on these topics.

There are many other tools and technologies that can be used for these purposes, and the specific tools and technologies that are used will depend on the specific needs and requirements of the organization.



Manoj Kumar



Data Warehouse

A data warehouse is a special type of database that is used to store and analyze large amounts of data.

It is designed to support fast querying and analysis of the data, and is typically used to help make business decisions.





Manoj Kumar

Example

A retail company might use a data warehouse to store and analyze sales data. The data warehouse could store information such as sales figures, customer information, and product information.

The company might use this data to identify trends and patterns in sales, and to make data-driven decisions about how to improve sales and customer satisfaction.





Manoj Kumar

Typical tools used

A data warehouse might be implemented using a relational database management system (RDBMS) such as Oracle, and loaded with data using an ETL (extract, transform, load) tool such as Informatica.

Business intelligence tools such as Tableau or Power BI could be used to query and analyze the data stored in the warehouse.





Manoj Kumar

Lakehouse

A lakehouse is a hybrid approach that combines the benefits of data lakes and data warehouses.

It allows to store both structured and unstructured data in its raw form, and allows for real-time querying and analytics on both structured and unstructured data.

It combines the scalability and flexibility of data lakes with the performance and governance capabilities of data warehouses, allowing for a single source of truth and a centralized location for data, rather than scattered across data silos





Manoj Kumar

Example

A financial services company might use a lakehouse to store and analyze large amounts of financial data from multiple sources such as trading systems, news feeds and social media.

The lakehouse could store data in a variety of formats, including structured, semi-structured, and unstructured data. The company might use this data to identify trends and patterns in financial markets and to make data-driven decisions about investment strategies.



Typical tools used



Manoj Kumar

A lakehouse might be implemented using big data platforms like Apache Hadoop, Apache Spark or Delta lake and could store data in a variety of formats, including structured, semi-structured, and unstructured data.

Data could be ingested into the lakehouse using tools such as Apache Nifi, Apache Kafka and processed using tools such as Apache Hive, Apache Pig, Apache Spark SQL or Delta lake.

Business intelligence tools such as Tableau or Power BI could be used to query and analyze the data stored in the lakehouse.





Manoj Kumar

Looking for Real-World Experience in Data Analytics/BI?

IM me on LinkedIn to know more
Or
[click here to book a call](#)