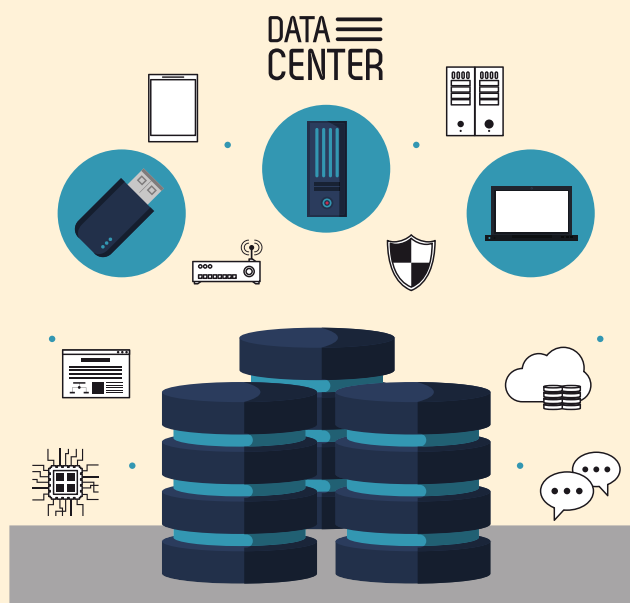
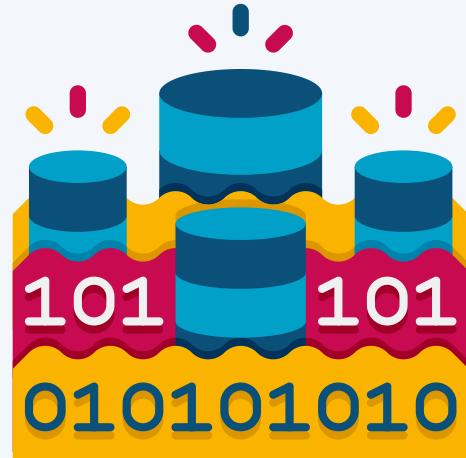




Manoj Kumar

Database VS Data Warehouse VS Data Lakes





Please note

This post provides a general overview of data warehouses, data lakes, and databases, 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



Database

A database is a collection of structured data that is stored in a computer system. It is used to store and manage data that is used by an application or system.

There are many different types of databases, such as relational databases and NoSQL databases

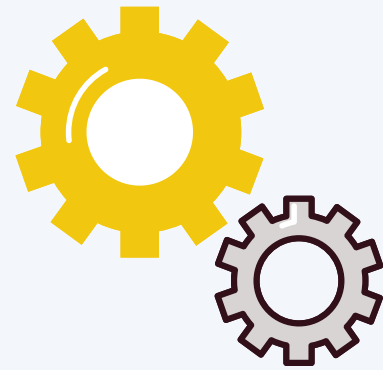


Example

A restaurant chain uses a database to store customer information and orders.

The database stores data including customer names, addresses, and phone numbers, as well as information about menu items ordered.

The restaurant uses this data to track customer preferences and analyse trends in customer behavior



Typical Tools Used

Relational Databases:

- Oracle
- MySQL
- Microsoft SQL Server

NoSQL Database:

- MongoDB
- Cassandra
- Couchbase



Data Warehouse

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

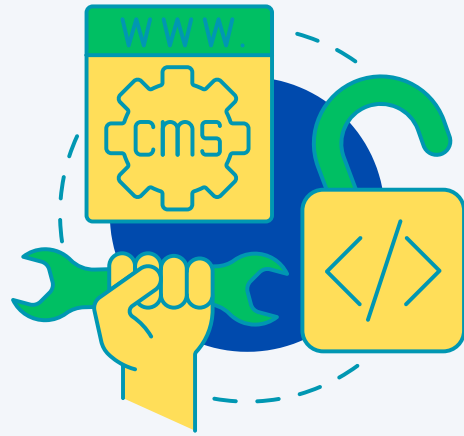
It is designed to support fast querying and analysis of the data, and is typically used to help make business decisions



Example

A hospital might use a data warehouse to store and analyse patient data. The data warehouse could store information such as patient medical records, treatment history, and outcomes data.

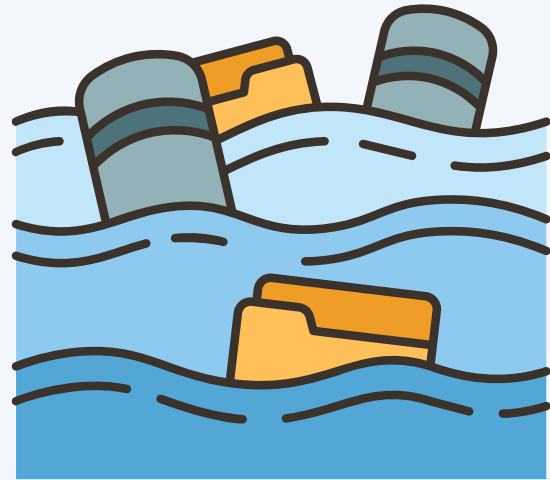
The hospital might use this data to identify trends and patterns in patient care, and to make data-driven decisions about how to improve patient outcomes



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 SSIS.

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



Data Lake

A data lake is a large, centralized repository that stores all kinds of data in its raw and original format.

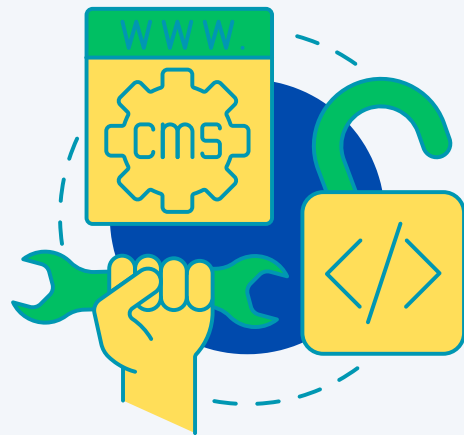
It is designed to handle very large amounts of data, and allows you to store data of any type and structure without the need to transform it before storing it.



Example

A government agency might use a data lake to store large amounts of data that is collected from a variety of sources, such as sensors, cameras, and other monitoring systems.

The data lake could store data in a variety of formats, including text, images, and video, and might be used to store data about traffic patterns, weather conditions, and other factors that are relevant to the agency ' s operations



Typical Tools Used

A data lake might be implemented using a big data platform such as Hadoop or Spark, and could store data in a variety of formats, including structured, semi-structured, and unstructured data.

Data could be ingested into the data lake using tools such as Flume or Kafka, and processed using tools such as Pig, Hive, or Spark



Manoj Kumar

Follow to Learn more..

Please like, comment and share with others

