Database vs Data Warehouse vs Data Lake
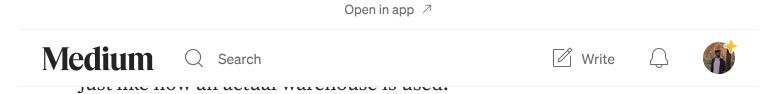
Jatin Shetty

2 min read · Aug 28, 2024

Its confusing at first, but all it takes is a few minutes to understand what they are and the differences.

If you are not a "data" person at your job, the high level is all you need. Also, give it another year or so, there will be many more new data-somethings that show up.

Alright, getting straight to the point now.

Database:

- Usually what we are most familiar with.

- Databases are usually Relational Databases. Meaning, data is stored in tables with rows are columns. And multiple tables can have relationships (references) with other tables.

- They use OLTP (Online Transaction Processing). This is basically a database system that is used to store transactional data. Think sale of an item.

- Databases are designed to be fast for read and write operations for the application it supports

Open in app ↗

Medium   Search   Write

just like how an actual warehouse is used.

- These are usually used for analytical purposes. Think running reports.

- Similar to how databases use the OLTP system, data warehouses use the OLAP system, which stands for Online Analytical Processing.

- Apart from running reports/analytics, warehouses are also used to store historical data.

- Not all the data from the databases make it to the warehouse, only some important ones make it.

- Because many databases feed into the warehouse it is important to have a standard schema that almost all databases can easily send their data to.

- A warehouse is built to be very fast and reliable for a large amount of read operations, usually for reports and analytical purposes

Data Lake:

If you are still with me, this is where it gets kinda interesting. The above two are mostly relational in nature. Data Lakes are usually not.

- It is designed to capture basically any kind of data and for large amounts.

- Think images, videos, csv etc any format of files essentially.

- The data is structured or semi-structured.

- Lakes are usually helpful for use in ML models.

- These data can also be aggregated, cleaned up and put into a database in a structured format

Closing Notes

As you can tell all 3 are completely different from each other. The use cases for what technology you would use for each scenario is different.

Most med-large sized organizations will have all types, so it upto you to design your solution accordingly. Or there is almost always a Data Governance team in an organization that you can consult with.

Thats it for this topic.

Cheers!

Jatin

Data   Database   Data Warehouse   Data Lake

Written by Jatin Shetty

4 Followers · 19 Following

Edit profile

Soldier in the Cyber Security Warfare.

No responses yet

What are your thoughts?

Respond