
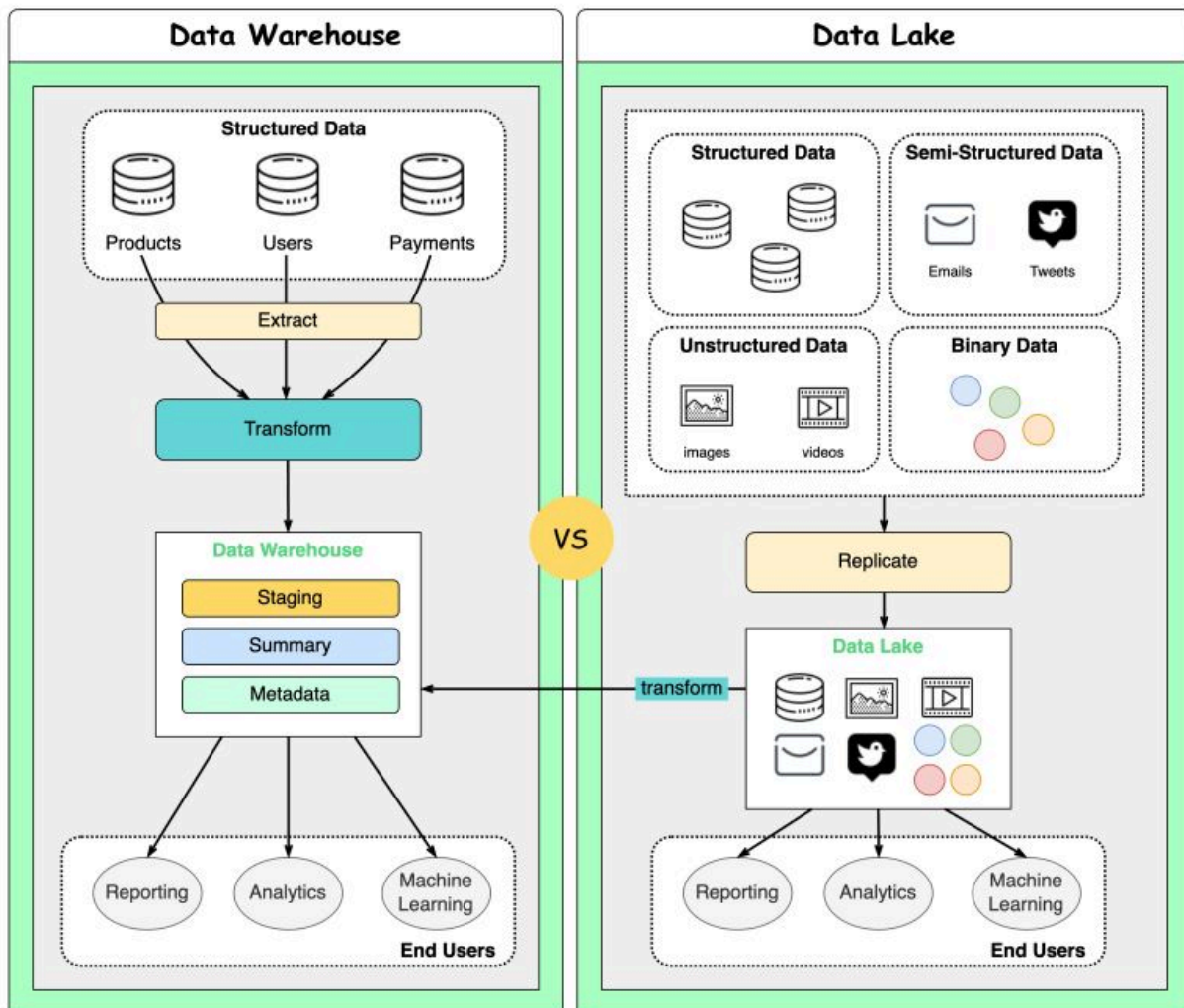


What are the differences between a data warehouse and a data lake?

The diagram below shows their comparison.

Data Warehouse vs Data Lake

 blog.bytebytego.com



- A data warehouse processes structured data, while a data lake processes structured, semi-structured, unstructured, and raw binary data.
- A data warehouse leverages a database to store layers of structured data, which can be expensive. A data lake stores data in low-cost devices.
- A data warehouse performs Extract-Transform-Load (ETL) on data. A data lake performs Extract-Load-Transform (ELT).

- A data warehouse is schema-on-write, which means the data is already prepared when written into the data warehouse. A data lake is schema-on-read, so the data is stored as it is. The data can then be transformed and stored in a data warehouse for consumption.

Over to you: Do you use a data warehouse or a data lake to retrieve data?