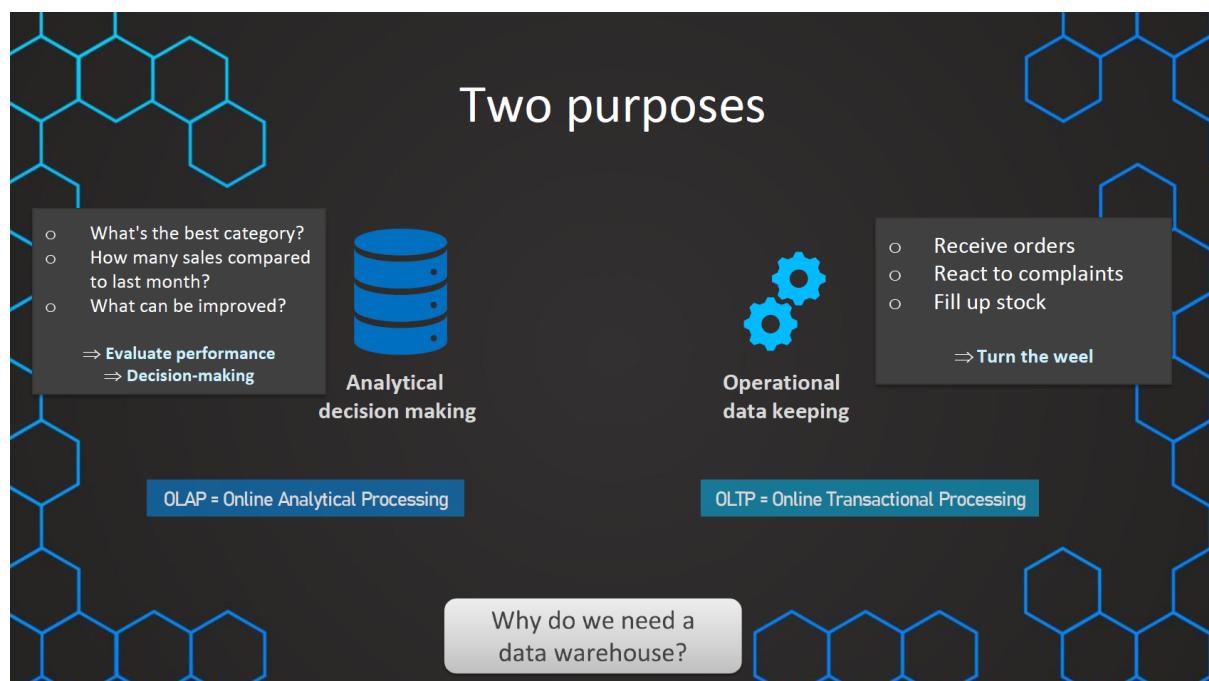


# 01. Introduction

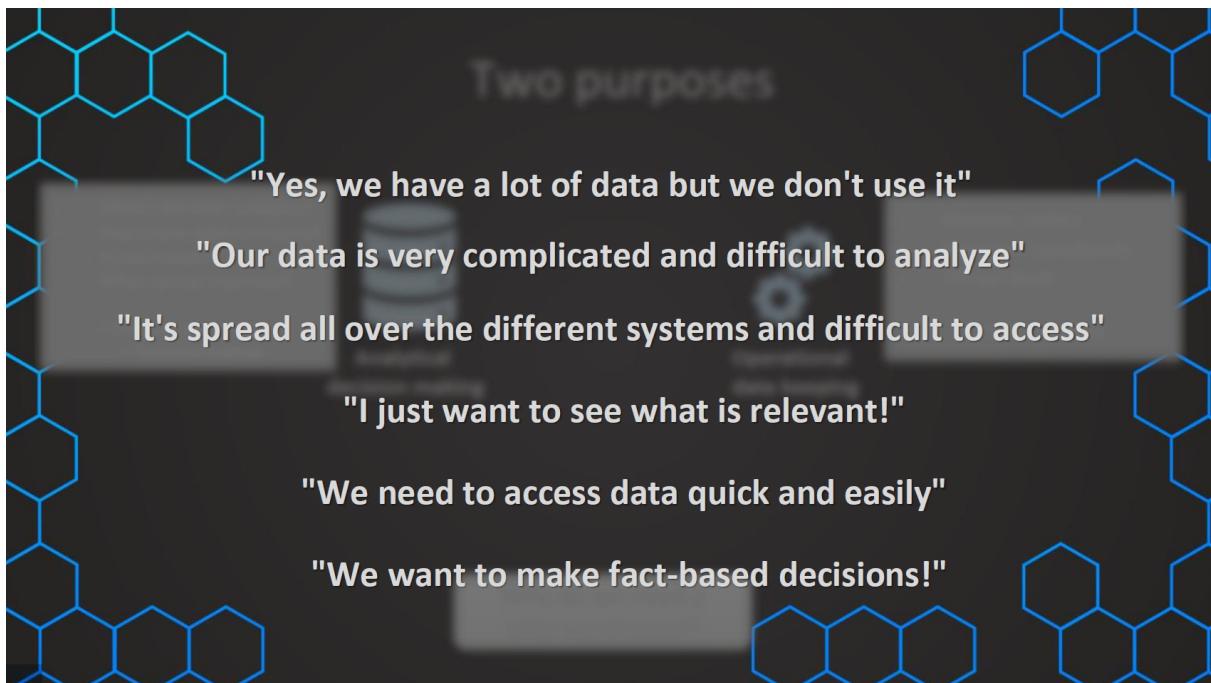
## Why do we need a data warehouse?

- To answer this questions we need to understand why and how we use the data in a company
  - Operational purpose (OLTP) : Things like receive orders, react to complaints, fill up stock basically to keep the wheels turning of the company i.e. keep the company operating. Also called **OLTP Online Transactional processing**
  - Analytical purpose (OLAP) : We also want to use the data we have to understand our company better and make better decisions and answer to the questions like What is the best category of products? Which is the best seller? Which is least seller? Number of sales compared to last month? Here we want to evaluate the performance and make better decision of the future. Also called **OLAP Online Analytical processing**

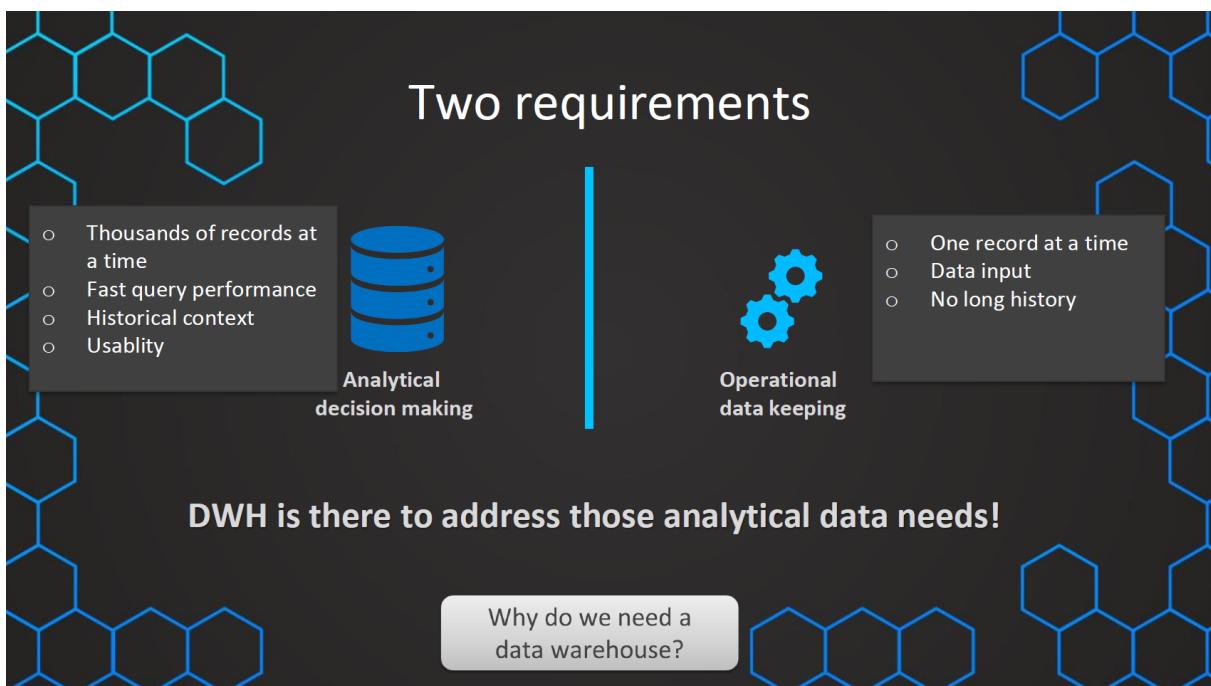


## Lack of Data Warehouse

Lack/Need of a data warehouse in a company is usually showcased by statements like:



## Requirement of a data warehouse



- Since we have those two different purposes we also have very different requirements to those systems.
- So while in the operational data processing (OLTP), we usually process one record at a time to keep the company operating, and we also usually want to enter some data or edit some data, and usually we are only concerned

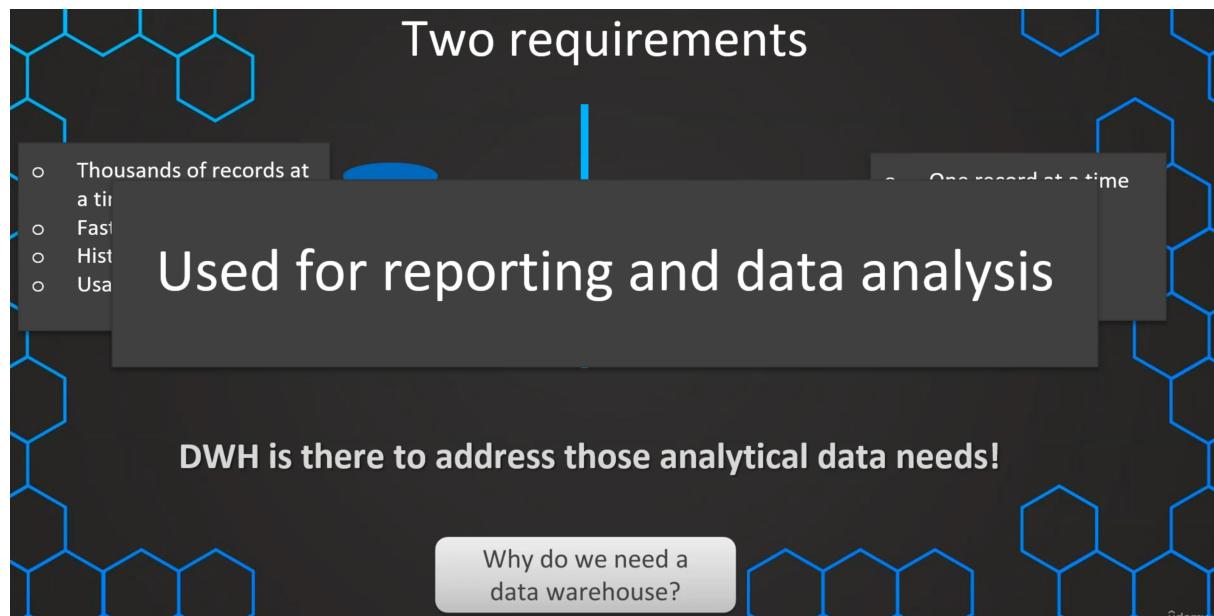
with current data and for this reason, we usually don't keep a very long history of the data.

- For the analytical data processing (OLAP), we analyse and get usually thousands and millions of records all at the same time. So for example, we want to analyse the average of the sales in the last six months, and in order to get this information quickly, it's very important to have a fast query performance.
- And also we want to make sense of this data and therefore we need context. So we want to analyze data over time or across multiple categories, and since we have those very different requirements it makes a lot of sense to keep those different systems separated.

---

***Data warehouse now is there to address the analytical data needs (OLAP side).***

So this is basically what a data warehouse is, it's a data location that is used for reporting and data analysis purposes.



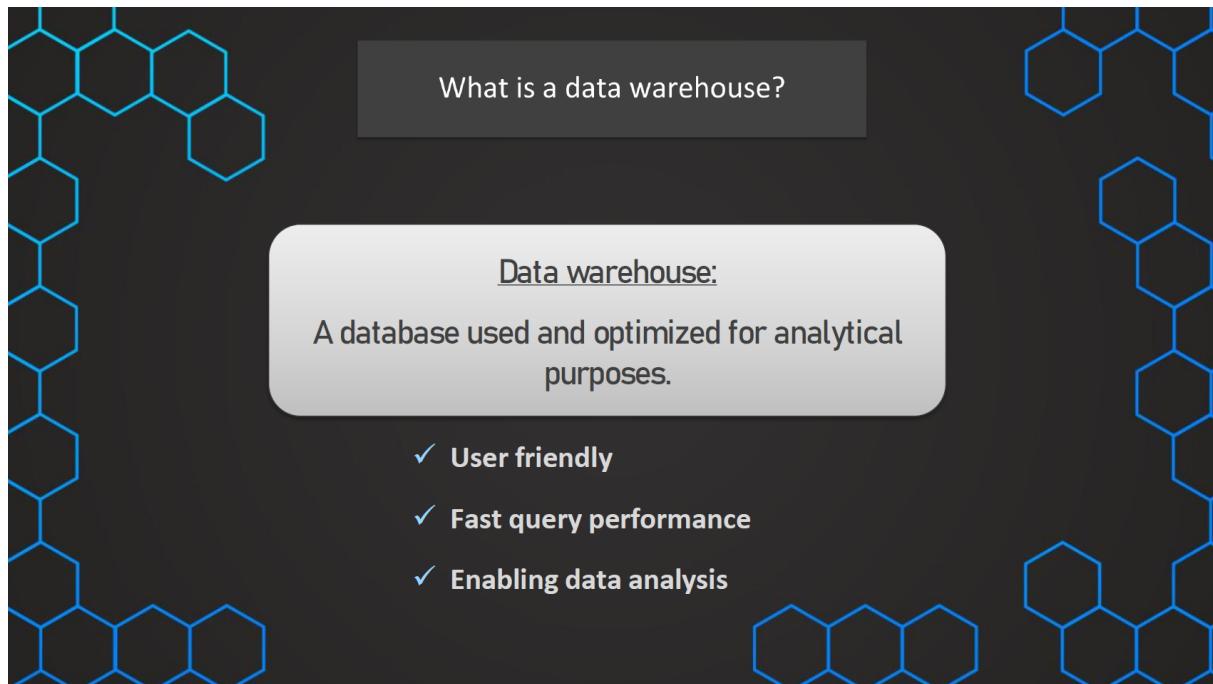
## What is a data warehouse?

***A database used and optimised for analytical purposes***

- Some of the things that are very important in a data warehouse is that it's very user-friendly, means it shouldn't be super technical, and also, all of the names, this should be all very easy to understand and optimised for the

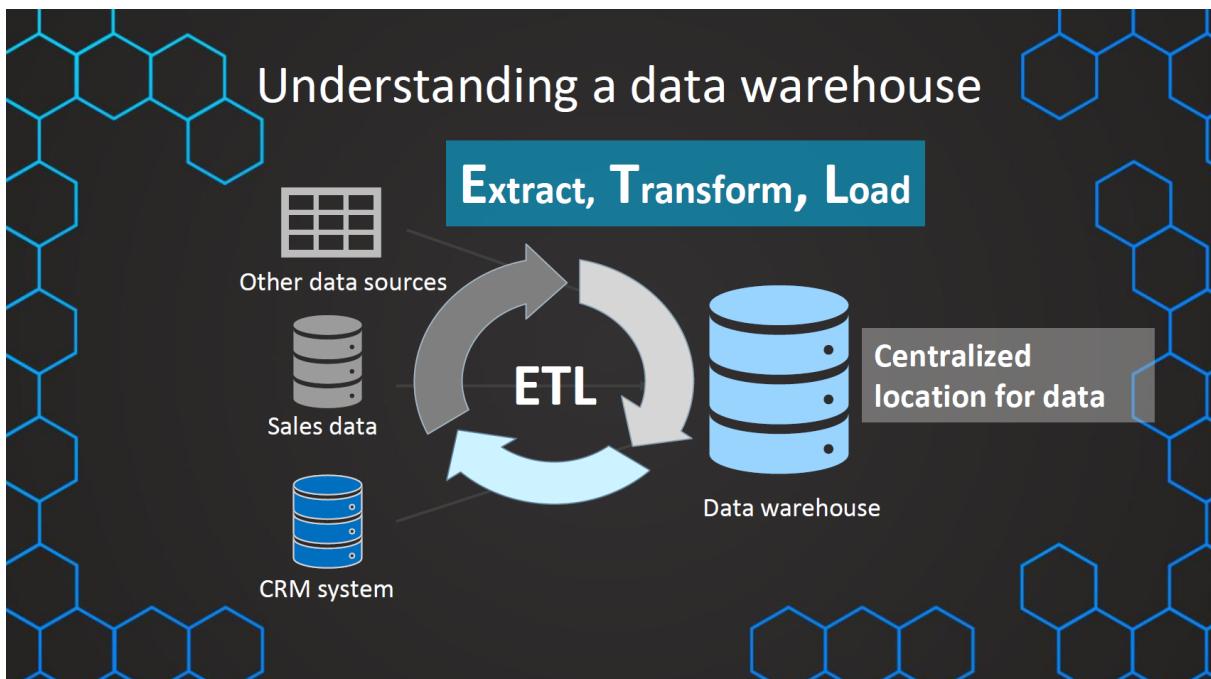
users or data analysts to retrieve this data, and process and work with this data.

- We also need a very fast query performance in this database, so that we can pull a lot of data, and process this data very quickly.
- In general, this database should be as good as possible, and optimize things to enable best and easy data analysis.



- We have the two different systems. So we have the operational data systems (OLTP). So this can be different data sources like sales data, our HR system, our CRM system and all of these data systems have different formats. The data is in a different structure and what we need to do is we need to bring all of that relevant data together and store it in a centralized location. This is the process of **data warehousing**, and also, this process is called **ETL process**.
- This is the most important process 90% of our time goes in if we are creating a data warehouse.
- So this ETL process stands for **Extract, Transform, Load**.
  - So basically we extract the data from these different sources, so that we don't use their query performance, and the resources and slow these systems down.

- We want to then transform the data, means that we want to integrate all of these different sources so that they are all in the same structure, that we can work with them, maybe the data is aggregated.
- And then, of course, in the last step, we want to load the data into our centralized location for data which is our data warehouse that is now optimized for this purpose of data analysis.



## Quiz



### Good job!

The two categories of data processing are analytical and operational processing.

Question 1:

What are usually the two main purposes of processing data in a company?

**Transactional and operational**

**Analytical and operational**

**Analytical and transactional**



### Good job!

This is analytical data processing, since data is analyzed, aggregated and used for analytical purposes.

Question 2:

What is *not* an example of operational data processing?

**A banking app receives and processes a transfer request from a user**

**An employee in a warehouse scans a product that was ordered and is picked**

**A regional manager calculates how many items an employee has picked in average per hour**

**A billing system calculates the monthly charges for a customer and sends the invoice per email**



### Good job!

Yes, OLTP stands for Online Transactional Processing

Question 3:

What is a term used to describe *operational data processing*?

**OLAP**

**OLTP**

Question 4:

What is *not* a key characteristic of a data warehouse?

**Centralized location for data**

**Fast query performance**

**User-friendly**

**Data recovery system**