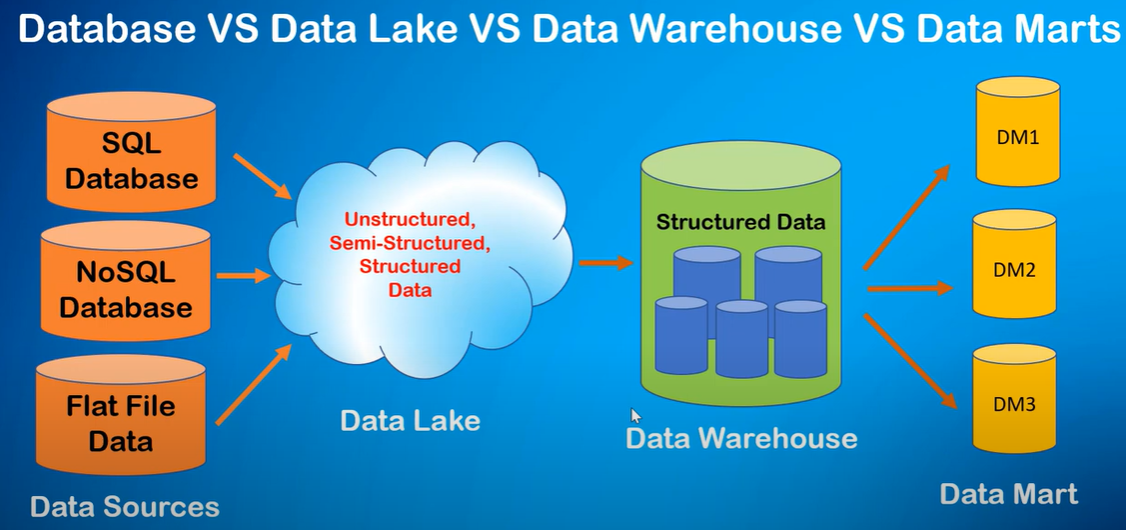
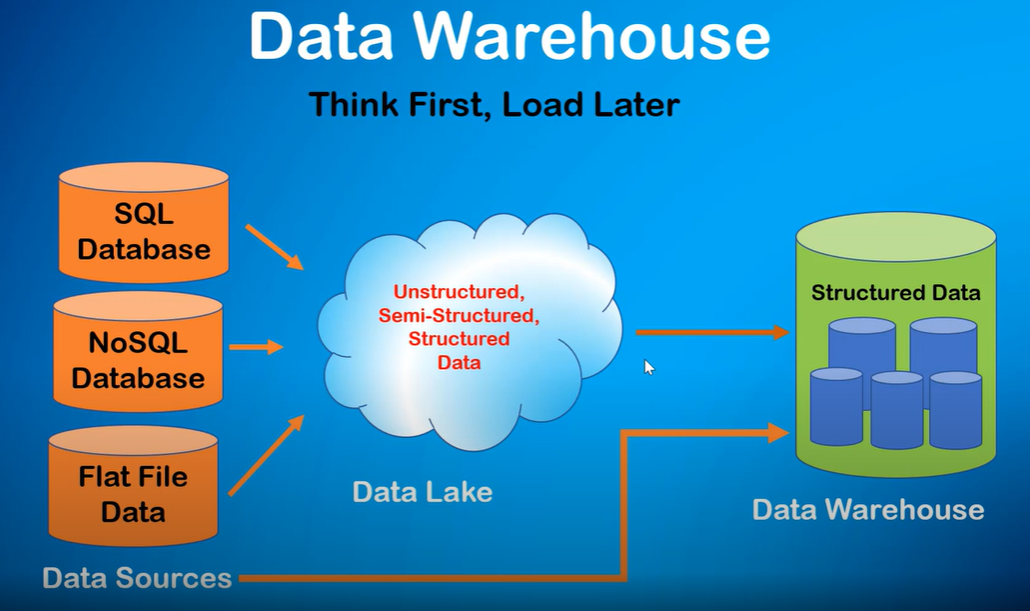
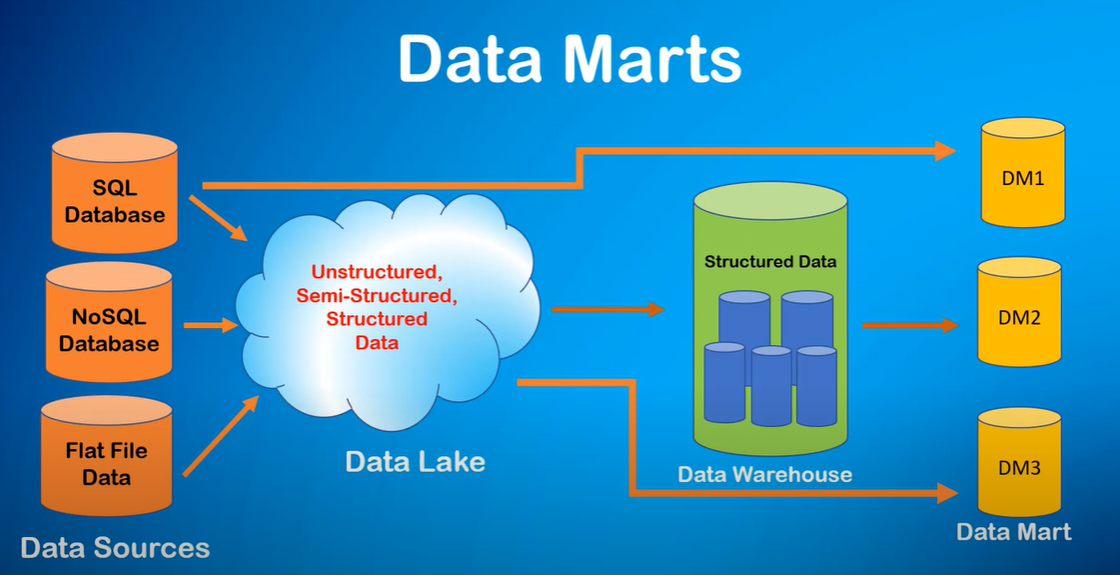
1. **What is Data Warehouse?**

Data warehousing (DW) is a method of gathering and [analysing data](https://www.interviewbit.com/data-analyst-interview-questions/) from many sources in order to get **useful business insights.** Typically, a data warehouse is used to integrate and analyze corporate data from many sources. The data warehouse is the heart of the business intelligence (BI) system, which is designed to analyze and report on data.



**Data Lake**

A data lake is the place where you dump all forms of data like structured, unstructured and log data in the form ELT methodology.

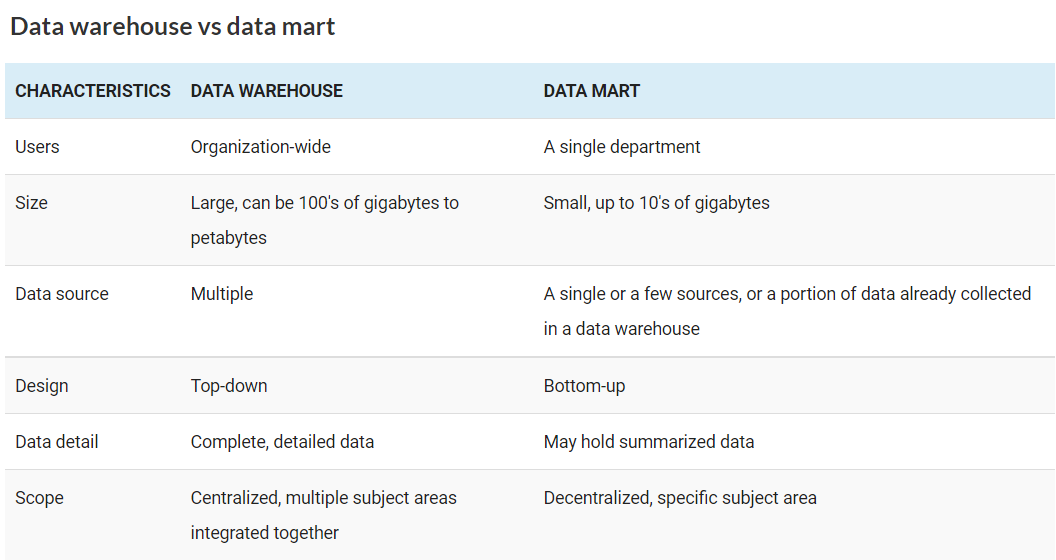
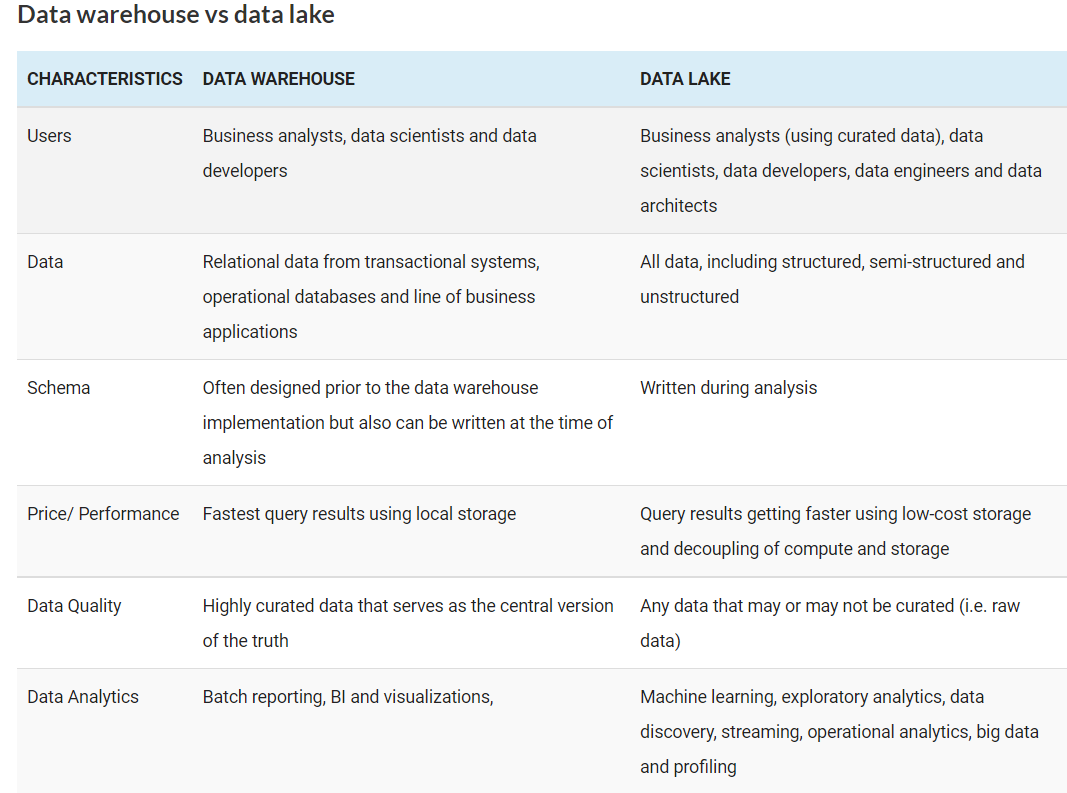
**Why data lake?**

A data lake is relevant in two contexts:

1- Your organization is so big and your product does so many functions that there are many possible ways to analyze data to improve the business. Thus, **you need a cheap way to store different types of data in large quantities.**

2- **You don't have a plan for what to do with the data, but you have a strong intent to use it at some point.** Thus, you collect data first and analyze later.

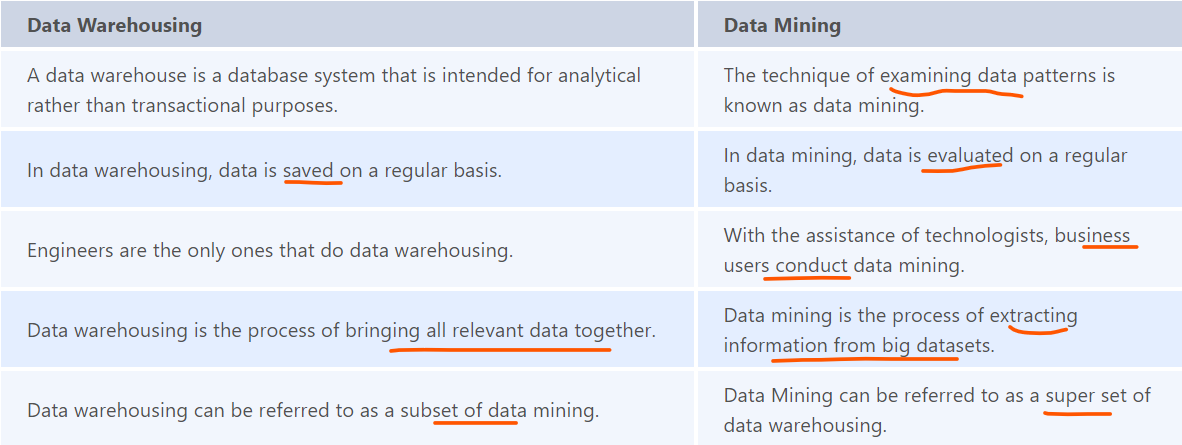
Also, the volume is so high that traditional DBs might take hours if not days to run a single query. So, having it in a Massively Parallel Processor (MPP) infrastructure helps you analyze the data comparatively quickly.



1. What do you mean by data mining? Differentiate between data mining and data warehousing.

Data Mining is the method of **analyzing hidden patterns of data** from various perspectives for **categorization** into useful data, in data warehouses environment for efficient analysis, assisting decision making, and other data requirements, ultimately resulting in **cost-cutting** and **revenue generation**.

* Examining enormous amounts of data for patterns and trends that go beyond simple analysis.
* Data mining estimates the probability of future events by utilizing advanced mathematical algorithms for data segments.



1. What is OLAP?

OLAP is abbreviated as Online Analytical Processing, and it is set to be a system which collects, manages, processes multi-dimensional data for analysis and management purposes.

|  |  |
| --- | --- |
| OLTP | OLAP |
| Data is from original data source | Data is from various data sources |
| Simple queries by users | Complex queries by system |
| Normalized small database | De-normalized Large Database |
| Fundamental business tasks | Multi-dimensional business tasks |

1. What do you understand about a fact table in the context of a data warehouse? What are the different types of fact tables?

 In a Data Warehouse system, a Fact table is simply a table that holds all **the facts or business information** that can be **exposed** to reporting and analysis when needed. Fields that reflect direct facts, as well as **foreign fields** that **connect the fact table to other dimension tables** in the Data Warehouse system, are stored in these tables. Depending on **the model type** used to construct the Data Warehouse, a Data Warehouse system can have **one or more fact tables.**

* **Transactional Fact Table:**
* **Snapshot Fact Table:**
* **Accumulating Fact Table:**

1. What do you mean by dimension table in the context of data warehousing? What are the advantages of using a dimension table?

Dimension table is a table which contain attributes of measurements stored in fact tables. This table consists of hierarchies, categories and logic that can be used to traverse in nodes.

