

## 4.7: Data Warehouse

As organizations have begun to utilize databases as the centerpiece of their operations, the need to fully understand and leverage the data they are collecting has become more and more apparent. However, directly analyzing the data needed for day-to-day operations is not a good idea; we do not want to tax the company's operations more than we need to. Further, organizations also want to analyze data in a historical sense: How does the data we have today compare with the same data set this time last month or last year? From these needs arose the concept of the data warehouse.

### Definition: Data Warehouse

A data warehouse is a centralized repository of integrated data from multiple sources, organized and optimized for reporting and analysis. Data warehouses consolidate data from different systems and sources, such as transactional systems, relational databases, and other sources into a single site to provide business users fast access to analytical and reporting data.

The data warehouse concept is simple: extract data from one or more of the organization's databases and load it into the data warehouse (which is itself another database) for storage and analysis. However, the execution of this concept is not that simple. A data warehouse should be designed so that it meets the following criteria:

- It uses non-operational data. This means that the data warehouse uses a copy of data from the active databases that the company uses in its day-to-day operations, so the data warehouse must pull data from the existing databases on a regular, scheduled basis.
- The data is time-variant. This means that whenever data is loaded into the data warehouse, it receives a timestamp, which allows for comparisons between different time periods.
- The data is standardized. Because the data in a data warehouse usually comes from several different sources, it is possible that the data does not use the same definitions or units. For example, our Events table in our Student Clubs database lists the event dates using the mm/dd/yyyy format (e.g., 01/10/2013). A table in another database might use the format yy/mm/dd (e.g., 13/01/10) for dates. For the data warehouse to match up the dates, a standard date format would have to be agreed upon, and all data loaded into the data warehouse would have to be converted to use this standard format. This process is called extraction-transformation-load (ETL).

There are two primary schools of thought when designing a data warehouse: bottom-up and top-down. The bottom-up approach starts by creating small data warehouses, called data marts, to solve specific business problems. As these data marts are created, they can be combined into a larger data warehouse. The top-down approach suggests that we should start by creating an enterprise-wide data warehouse and then, as specific business needs are identified, create smaller data marts from the data warehouse.

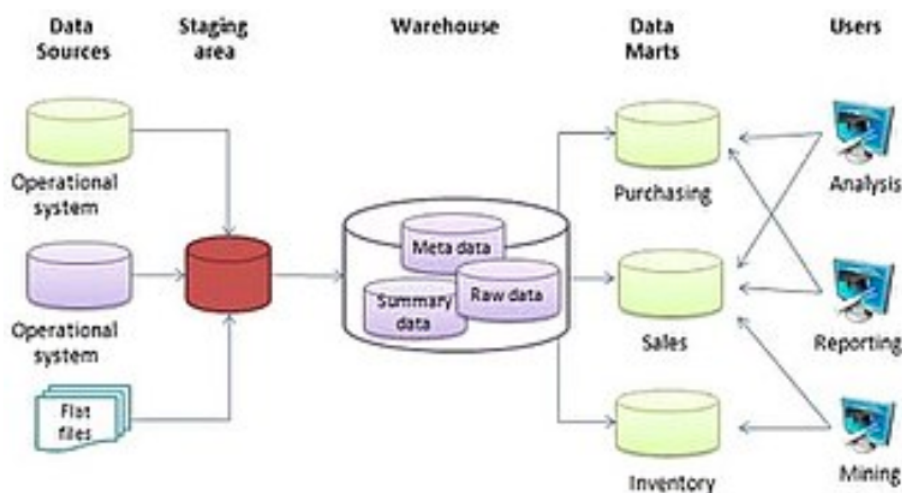


Figure 4.7.1: Data

warehouse process (top down). [Data warehouse process \(top down\)](#) by [Soha jamil](#) is licensed under [CC BY-SA 4.0](#)

### 4.7.1: Benefits of Data Warehouses

Organizations find data warehouses quite beneficial for many reasons:

- Ability to integrate data from multiple systems formatted with different software and compile it to gain deeper insight.
- The process of developing a data warehouse forces an organization to understand the data better than it is currently collecting and, equally important, what data is not being collected.
- A data warehouse provides a centralized view of all data being collected across the enterprise and provides a means for determining inconsistent data.
- Once all data is identified as consistent, an organization can generate one version of the truth. This is important when the company wants to report consistent statistics about itself, such as revenue or employees' numbers.
- By having a data warehouse, snapshots of data can be taken over time. This creates a historical record of data, which allows for an analysis of trends.
- A data warehouse provides tools to combine data, which can provide new information and analysis.

#### ✓ Use Case Data Warehouse 4.7.1

A retail company sets up a data warehouse to analyze sales performance. The data warehouse pulls in transaction data from all store checkout systems, inventory data from the ERP system, and customer information from loyalty program databases. This disparate data is integrated and transformed into a unified structure optimized for queries.

What do business analysts do to analyze sales performance?

#### **Solution**

Business analysts run reports and organize sales data in the data warehouse by time, region, product, customer attributes, and other dimensions. This enables them to spot trends, measure promotional performance, and identify challenges across the retail operation. Having all the data aggregated for reporting in the warehouse is more efficient than trying to query multiple sources and deal with the different types of reports from different systems.

With the sales data warehouse, the business gets a full picture of its performance from high-level indicators to granular details. This allows managers to make strategic decisions around pricing, inventory, promotions, or staffing.

In essence, the sales data warehouse transforms raw transactional data into an analytical asset for driving growth.

This page titled [4.7: Data Warehouse](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Ly-Huong T. Pham and Tejal Desai-Naik \(Evergreen Valley College\)](#) .

- [4.7: Data Warehouse](#) by Ly-Huong T. Pham, Tejal Desai-Naik, Laurie Hammond, & Wael Abdeljabbar is licensed [CC BY 3.0](#).