The Difference Between a Database and a Data Warehouse

What's the difference between a database and data warehouse? Discover some key differences to help you distinguish between the two.

What is a Database?

A database is a collection of information stored in an organized manner. Many enterprises and organizations create and manage databases using a database management system. Special DBMS software can be used create and store product inventory and customer information, for example.

Organizations most often use databases for online transaction processing (OLTP). Database software needs to provide easy access to information and fast querying so that transactions can be carried out efficiently. Databases are often referred to as operational systems, meaning they are used to process day-to-day transactions in an organization.

What is a Data Warehouse?

A data warehouse is a system that pulls together data from many different sources within an organization for reporting and analysis. The reports created from complex queries within a data warehouse are used to make business decisions.

Data warehouses are used for online analytical processing (OLAP), which uses complex queries to analyze rather than process transactions.

Data Warehouse vs Database

The following points highlight some of the key differences between databases and data warehouses.

Optimization

A database is optimized to maximize the speed and efficiency with which data is updated (added, modified, or deleted) and enable faster analysis and data access. Response times from databases need to be extremely quick for efficient transaction processing. The most important aspect of a database is that the write operation is recorded in the system. A business that sells products online wouldn't be in business very long if its database didn't make a record of every purchase.

A data warehouse, on the other hand, is optimized for reading and aggregating large data sets. Data warehouses can handle complex analytical queries that would use too much of a database system's resources.

Data Structure

Most databases use a normalized data structure. Data normalization means reorganizing data so that it contains no redundant data, and all related data items are stored together, with related data separated into multiple tables. Normalizing data ensures the database takes up minimal disk space while response times are maximized.

Product	
ProductID	Desc
1	Mtn Bike #726
2	Road Bike #871
3	Touring Bike #378

Color	
ColorID	Desc
1	Red
2	Black
3	Silver
4	Mauve

Product-Color	
ProductID	ColorID
1	1
1	2
2	1
2	2
2	3
3	1
3	3
3	4

The more normalized your data is, the more complex the queries needed to read the data because a single query combines data from many tables. This puts a huge strain on computing resources.

The data in a data warehouse does not need to be organized for quick transactions. Therefore, data warehouses normally use a denormalized data structure. A denormalized data structure uses fewer tables because it groups data and doesn't exclude data redundancies. Denormalization offers better performance when reading data for analytical purposes.
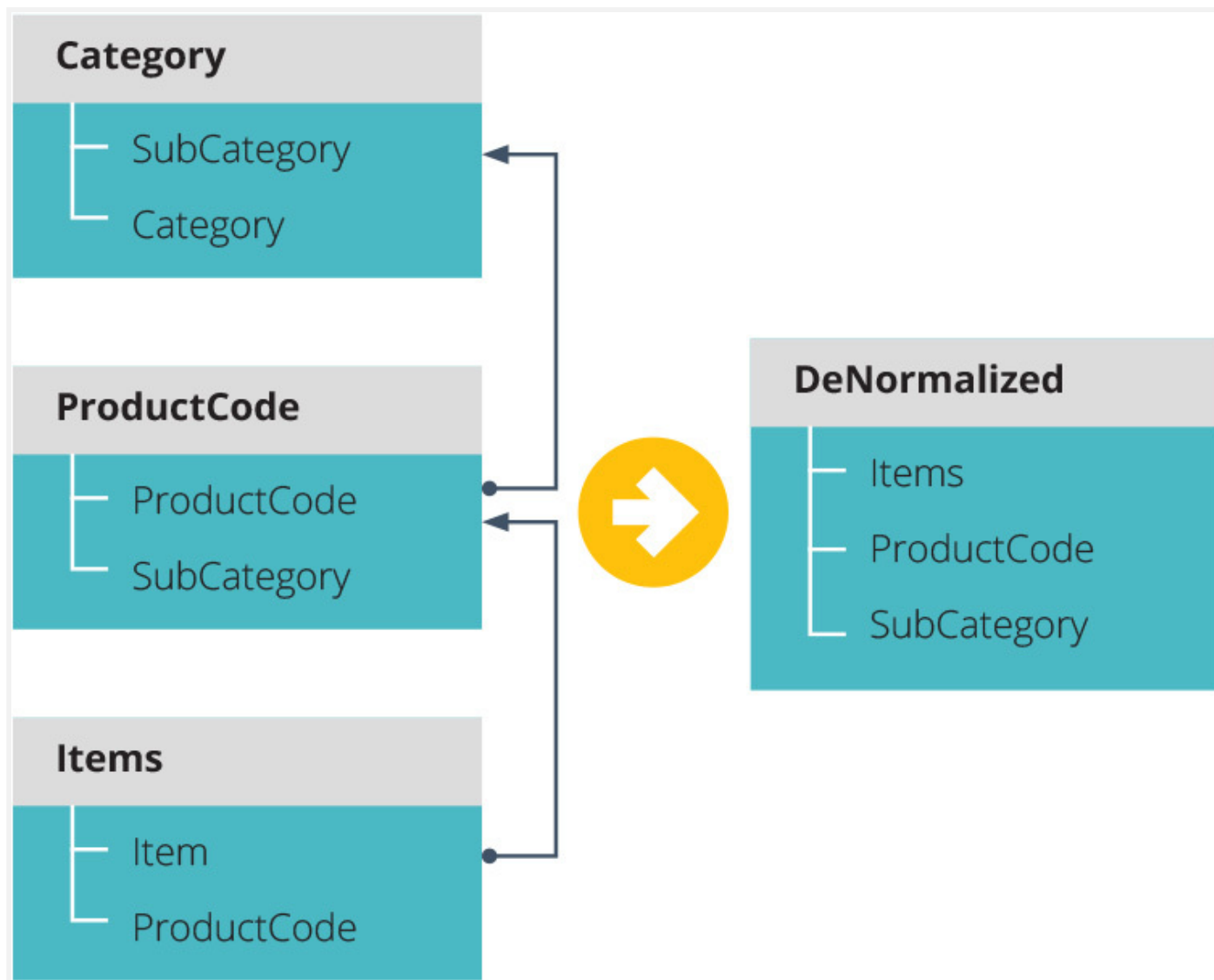
Data Timeline

A database processes day-to-day transactions within an organization. Therefore, databases typically don't contain historical data—current data is all that matters in a normalized relational database.

Data warehouses are used for analytical purposes and business reporting. Data warehouses typically store historical data by integrating copies of transaction data from disparate sources. Data warehouses can also use real-time data feeds for reports that use the most current, integrated information.

Analysis

While databases are normally used for transactional purposes, analytical queries can still be performed on the data. The problem is that the complexity of the data's normalized organization makes analytical queries difficult to carry out. A skilled developer or analyst will be required to create such analytical queries. The depth of analysis is limited to static one-time reports because databases just give a snapshot overview of data at a specific time.

The structure of data warehouses makes analytical queries much simpler to perform. No advanced knowledge of database applications is required. Analytics in data warehouses is dynamic, meaning it takes into account data that changes over time.

Concurrent Users

An OLTP database supports thousands of concurrent users. Many users must be able to interact with the database simultaneously without it affecting the system's performance.

Data warehouses support a limited number of concurrent users compared to operational systems. The data warehouse is separated from front-end applications and it relies on complex queries, thus necessitating a limit on how many people can use the system simultaneously.

Database vs. Data Warehouse Use Cases

Database Use Cases

Database use cases are related to the day-to-day transactional requirements in an organization. Some examples of database applications include:

- An airline using an online booking system

- A hospital registering a patient

- A bank adding an ATM withdrawal transaction to an account

- A website creating an order for a product it has sold

Data Warehouse Use Cases

Data warehouse use cases focus on providing high-level reporting and analysis that lead to more informed business decisions. Use cases include:

- Carrying out data mining to gain new insights from the information held in many large databases

- Conducting market research by analyzing large volumes of data in-depth

- An online business analyzing user behavior to make business decisions

Learn More About Data Warehouses

- [Data Warehouse Architecture: Traditional vs. Cloud](#)

- [Data Warehouse Concepts](#)

- [Data Mart vs. Data Warehouse](#)

Built by Panoply