Transactional and Analytical Databases

by Sophia

WHAT'S COVERED

In this lesson, you will explore the similarities and differences between analytical databases and transactional databases. Specifically, this lesson will cover.

1. Introduction
2. Transactional Databases
3. Analytical Databases

1. Introduction

An organization's data management and decision making depend heavily on understanding the difference between **transactional databases** and **analytical databases**. Everyday operations require transactional databases, such as handling financial transactions, managing inventory, and dealing with customers. These applications provide real-time transaction processing, ensure data integrity, and enforce business rules. Transactional databases ensure operational data accuracy, consistency, and reliability, with streamlined operations and satisfied customers as a result.

Large volumes of data can be transformed into meaningful insights using analytical databases. They enable businesses to identify trends, patterns, and correlations in their data through complex data analysis, reporting, and advanced analytics tasks. In order to make better decisions, organizations rely on analytical databases that provide insight into their business performance, customer behavior, and market trends. Organizations can harness data-driven insights through analytical databases to optimize operations, improve decision making, and gain a competitive advantage.

TERMS TO KNOW

Transactional Database

A database system designed for managing the day-to-day operations of an organization.

Analytical Databases

A specialized type of database designed for storing, retrieving, and analyzing large volumes of data.

2. Transactional Databases

Transactional databases, also called **online transaction processing (OLTP) databases** or operational databases, are used for handling real-time transactions daily. The system is designed to efficiently process many individual transactions, such as creating, retrieving, modifying, and deleting data.

Transactional databases are traditionally relational ones that use tables, rows, and columns as their basic building blocks. Schemas define the structure of the tables, the data types, and the relationships between them. A structured approach facilitates the organization of data, the indexing of data, and the processing of queries.

Because of the many frequent changes made in a transactional database, software and processes must be in place to ensure that each change is successfully made, even when the unexpected happens. For example, what if the power goes out as a change is being written to the database?

Transactional databases maintain data integrity and reliability by adhering to the properties of **ACID** (Atomicity, Consistency, Isolation, Durability). Using ACID properties ensures that transactions are treated as indivisible units, guaranteeing the successful completion of all transaction operations. These measures ensure data consistency and prevent data loss or corruption.

Transactional databases enable multiple users to simultaneously access the data, and that can potentially create situations where multiple sources or users are updating the same record (differently) at the same time. To avoid such problems, transactional databases implement data integrity mechanisms such as:

- Locks: Mechanisms used to prevent multiple transactions or processes from simultaneously accessing or modifying the same data.
- Concurrency control: Processes for enabling and releasing locks and resolving conflicts when multiple transactions contend for the same data.
- Isolation levels: Rules about the degree to which concurrently running transactions are isolated from each other.

Because transactional databases tend to be used in real-time activities such as sales and banking, they are optimized for fast response times and low latency. These systems are designed to handle high transaction volumes and prioritize quick data retrieval and updating.


KEY CONCEPT

An important feature of transactional databases is transaction management. Transactions are initiated, committed (that is, written to the database), rolled back, and recovered using well-defined procedures designed to ensure data integrity if a failure, error, or system crash occurs.

A transactional database is well suited for OLTP workloads, which typically involve frequent and short-duration transactions. A database like this is ideal for storing, updating, and frequently querying data.

Many applications and industries use transactional databases, including e-commerce, banking, inventory management, airline reservation systems, and healthcare. Data consistency, integrity, and responsiveness are

based on this kind of database, ensuring efficient and reliable transaction processing.

TERMS TO KNOW

Online Transaction Processing (OLTP) Databases

Operational databases, which are used for handling real-time transactions daily.

ACID

An acronym that stands for atomicity, consistency, isolation and durability of data.

3. Analytical Databases

The purpose of analytical databases is to store and analyze large amounts of data. They are optimized to respond to queries quickly and perform advanced analytics. Analytical databases have several key characteristics that help define the database as analytical rather than transactional.

Analytical databases typically store data in columns rather than in rows; this is called **columnar data storage**. Each column of a table is stored separately, with all the values for that column grouped together. This arrangement enables queries that scan large data pools for specific columns to run faster and more efficiently.

Query performance is further improved by in-memory processing, which stores and manipulates data entirely in RAM rather than on traditional disk storage. This enables much faster query execution.

Performance can be further enhanced by **parallel processing** in analytical databases. Parallel processing refers to the simultaneous execution of multiple operations across multiple CPU cores in a distributed computing environment to improve performance. This is especially useful for large-scale analytics, where a single server cannot handle the data.

Analytical databases are designed to handle complex queries efficiently, including those that involve joining tables, summarizing and grouping data, and embedding queries within other queries. These capabilities help users extract useful information from the data without long delays.

The ability to scale an analytical database is essential if it must be able to handle large amounts of data and complex queries. Distributed architectures are typically used for this purpose. A distributed architecture is a design in which data is stored across multiple interconnected servers, to avoid bogging down any individual server.

Databases with analytical capabilities can compress data, save space, and improve their performance over time. Some of the techniques used to achieve these benefits include:

- Data deduplication: eliminating duplicate values.
- Data pruning: eliminating outdated or irrelevant data.
- Data compression: running data through a compression algorithm that reduces the space it occupies.
- Indexing: creating a list of values and references to the locations where they can be found.

KEY CONCEPT

Analytical databases are invaluable to businesses whose data must be stored and analyzed in large quantities. Companies can benefit from them by making better decisions, improving efficiency, and identifying new opportunities. This kind of database supports **business analytics**, decision support systems, and group decision support systems.

Analytical databases generally consist of two main components: the **data warehouse** and the **online analytical processing (OLAP)** front end. The data warehouse is a specialized database that stores the underlying data in a format optimized for decision support rather than transactions. As such, data is typically loaded from transactional databases or other sources mostly with SELECT statements in SQL. It's rare to see any UPDATE, INSERT, and DELETE statements, as the data has already been verified and validated when it has gone into the transactional database. Online analytical processing (OLAP) generally contains a set of tools used to process that data from the data warehouse so that it can be used for analytics. This includes formatting, clearing, and removing data that does not fit specific patterns for analysis.

Analytical databases are used to help with business intelligence by generating information that is used for business decision making. The volume of data in these analytical databases can be massive due to the historical data and the data sets that may be added from third parties. Since the data does not often change, space is not as much of a concern.

TERMS TO KNOW

Columnar Data Storage

Storing data in columns rather than in rows.

Parallel Processing

The simultaneous execution of multiple operations across multiple CPU cores in a distributed computing environment to improve performance.

Online Analytical Processing (OLAP)

Software technology that is used to analyze business data from different points of view.

Business Analytics

The practice of using data analysis and statistical techniques to drive informed business decision making.

Data Warehouse

A backend enterprise-level system used for storing data that is used for analysis and reporting.

Front End

A user interface or application that enables users to interact with a database without directly interacting with the underlying DBMS.

Back End

The part of the database system responsible for managing and storing data, including the DBMS and the physical storage devices. The back end is not directly accessible to end users.

SUMMARY

In this lesson, you learned about the different purposes and characteristics that distinguish **transactional databases** from **analytical databases**. An OLTP (online transaction processing) database handles and manages daily operational transactions in real time. Data integrity, concurrency, and high-speed transactions are prioritized, ensuring efficient data entry, retrieval, and modification. It's important to note that transactional databases are optimized for handling multiple concurrent transactions with ACID properties (Atomicity, Consistency, Isolation, Durability), ensuring reliable and consistent data.

You also learned that a database that enables complex analytical queries and data analysis is known as an analytical database, also called an OLAP (online analytical processing) database. These systems are designed to handle large volumes of data and provide fast query response times. Businesses can gain insights and make informed decisions through data exploration, reporting, and complex calculations. An analytical database often involves advanced analytics techniques, facilitating in-depth analysis and understanding of data trends.

Source: THIS TUTORIAL WAS AUTHORED BY DR. VINCENT TRAN, PHD (2020) AND FAITHE WEMPEN (2024) FOR SOPHIA LEARNING. PLEASE SEE OUR TERMS OF USE.

TERMS TO KNOW

ACID
An acronym that stands for atomicity, consistency, isolation and durability of data.

Analytical Databases
A specialized type of database designed for storing, retrieving, and analyzing large volumes of data.

Back End
The part of the database system responsible for managing and storing data, including the DBMS and the physical storage devices. The back end is not directly accessible to end users.

Business Analytics
The practice of using data analysis and statistical techniques to drive informed business decision making.

Columnar Data Storage
Storing data in columns rather than in rows.

Data Warehouse
A backend enterprise-level system used for storing data that is used for analysis and reporting.

Front End
A user interface or application that enables users to interact with a database without directly interacting with the underlying DBMS.

Online Analytical Processing (OLAP)

Software technology that is used to analyze business data from different points of view.

Online Transaction Processing (OLTP) Databases

Operational databases, which are used for handling real-time transactions daily.

Parallel Processing

The simultaneous execution of multiple operations across multiple CPU cores in a distributed computing environment to improve performance.

Transactional Database

A database system designed for managing the day-to-day operations of an organization.