DATABASE Vs DATA WAREHOUSE

Why use a database?

If your application needs to store data (and nearly every interactive application does), your application needs a database. Applications across industries and use cases are built on databases. Many types of data can be stored in databases, including:

- Patient medical records
- Items in an online store
- Financial records
- Articles and blog entries
- Sports scores and statistics
- Online gaming information
- Student grades and scores
- IoT device readings
- Mobile application information

How database is implemented

A myriad of databases exist. Examples include:

- Relational databases: Oracle, MySQL, Microsoft SQL Server, and PostgreSQL
- Document databases: MongoDB and CouchDB
- Key-value databases: Redis and DynamoDB
- Wide-column stores: Cassandra and HBase
- Graph databases: Neo4j and Amazon Neptune

What is a database?

A database is a collection of data or information. Databases are typically accessed electronically and are used to support Online Transaction Processing (OLTP). Database Management Systems (DBMS) store data in the database and enable users and applications to interact with the data. The term “database” is commonly used to reference both the database itself as well as the DBMS.

Database characteristics

A variety of database types have emerged over the last several decades. All databases store information, but each database will have its own characteristics. Relational databases store data in tables with fixed rows and columns. Non-relational databases (also known as NoSQL databases) store data in a variety of models including JSON (JavaScript Object Notation), BSON (Binary JSON), key-value pairs, tables with rows and dynamic columns, and nodes and edges. Databases store structured and/or semi-structured data, depending on the type.

You may also find database characteristics like:

- Security features to ensure the data can only be accessed by authorized users.
- ACID (Atomicity, Consistency, Isolation, Durability) transactions to ensure data integrity.
- Query languages and APIs to easily interact with the data in the database.
- Indexes to optimize query performance.
- Full-text search.
- Optimizations for mobile devices.
- Flexible deployment topologies to isolate workloads (e.g., analytics workloads) to a specific set of resources.
- On-premises, private cloud, public cloud, hybrid cloud, and/or multi-cloud hosting options.

Why use a data warehouse?

Data warehouses are a good option when you need to store large amounts of historical data and/or perform in-depth analysis of your data to generate business intelligence. Due to their highly structured nature, analyzing the data in data warehouses is relatively straightforward and can be performed by business analysts and data scientists.

Note that data warehouses are not intended to satisfy the transaction and concurrency needs of an application. If an organization determines they will benefit from a data warehouse, they will need a separate database or databases to power their daily operations

How is Data warehouse implemented

Examples of data warehouses include:

- Amazon Redshift.
- Google BigQuery.
- IBM Db2 Warehouse.
- Microsoft Azure Synapse.
- Oracle Autonomous Data Warehouse.
- Snowflake.
- Teradata Vantage.

What is a data warehouse?

A data warehouse is a system that stores highly structured information from various sources. Data warehouses typically store current and historical data from one or more systems. The goal of using a data warehouse is to combine disparate data sources in order to analyze the data, look for insights, and create business intelligence (BI) in the form of reports and dashboards.

You might be wondering, "Is a data warehouse a database?" Yes, a data warehouse is a giant database that is optimized for analytics.

Data warehouse characteristics

Data warehouses store large amounts of current and historical data from various sources. They contain a range of data, from raw ingested data to highly curated, cleansed, filtered, and aggregated data.

Extract, transform, load (ETL) processes move data from its original source to the data warehouse. The ETL processes move data on a regular schedule (for example, hourly or daily), so data in the data warehouse may not reflect the most up-to-date state of the systems.

Data warehouses typically have a pre-defined and fixed relational schema. Therefore, they work well with structured data. Some data warehouses also support semi-structured data.

Once the data is in the warehouse, business analysts can connect data warehouses with BI tools. These tools allow business analysts and data scientists to explore the data, look for insights, and generate reports for business stakeholders.

Data Warehouse vs. Database: A Comparative Analysis

Parameter	Data Warehouse	Database
Workloads	Analytical	Transactional and Operational
Characteristics	It is subject-focused since it provides information on a certain topic rather than information about a company's current activities. The data also has to be stored in a unanimously acceptable manner and data warehouse in common.	Removes redundancy and offers security. It allows for numerous data views.
Data Type	It stores both historical and current data. It is possible that the data is out of date.	The data in the database is updated.
Orientation	Might not be updated. Depends on the frequency of ETL processes.	Real-time
Purpose	Designed to analyze	Designed to record
Tables and Joins	Tables and joins are straightforward since they're denormalized.	A database's tables and joins are complicated because they're normalized.
Availability	It is available in real-time.	Data is updated from source systems when needed.
Technique	Analyze data	Capture data
Query Type	Simple transaction queries are implemented.	Complex queries are utilized for analytical reasons.

Schema Flexibility	Fixed and pre-defined schema definition for ingest.	Flexible or rigid schema based on the type of database.
Users	Data scientists and business analysts.	Application developers
Processing Method	It uses OLAP (Online Analytical Processing).	It makes use of OLTP (Online Transactional Processing).
Storage Limit	Data from any number of apps is stored.	Generally confined to a particular application.
Usage	Data modeling approaches are employed for designing. It permits you to analyze your enterprise.	ER modeling approaches are employed for designing. It aids in the execution of basic business procedures
Applications	Healthcare sector, airline, retail chain, insurance sector, banking, and telecommunication.	Banking, universities, airlines, finance, telecommunication, manufacturing, sales and production, and HR management.
Pros	A data warehouse allows business users to access vital data from several sources in one location. It delivers consistent information on numerous cross-functional tasks. Aids in the integration of several data sources in order to alleviate the load on the production system.	It provides data security and access. A database provides a number of methods for storing and retrieving data. Databases function as an efficient handler to balance the need of various applications using the same data.

Cons	Adding additional data sources takes effort and comes at a con. Problems with the data warehouse can sometimes go undiscovered Data warehouses require a leep. Data extraction, loading, and cleaning can be time-consuming.	The cost of hardware and software for creating a database system is quite high, which might raise your organization's budget. Because many DBMS systems are complicated, training users to utilize the DBMS is essential. Data owners might lose control of their data, generating concerns about ownership, security, and privacy.

References for this article:

1. mongodb.com
https://www.mongodb.com/resources/basics/databases/data-lake-vs-data-warehouse-vs-database
2. simplilearn.com
https://www.simplilearn.com/data-warehouse-vs-database-article
3. coursera
https://www.coursera.org/articles/data-warehouse-vs-database

Happy Cloud Computing

Regards

Saransh Jain