System Databases by Fabrizio Fernandez soto

05/26/2025

Databases and Database systems

A database is an organized collection of data. The data are typically organized to model aspects of reality in a way that supports processes requiring information. For example modelling the availability of rooms in hotels in a way that supports finding a hotel with vacancies.

Introduction

Formally “ database” refers to the data themselves and supporting data structures. Databases are created to operate large quantities of information by inputting, storing, retrieving and managing that information. Databases are set up so that one set of software programs provides all users with access to all data.

Database system software

The next generation of post-relational databases in the late 2000s became known as NoSQL databases, introducing fast key-value stores and document-oriented databases. A competing “next generation” known as NewSQL databases attempted new implementations that retained the relational/SQL model while aiming to match the high performance of NoSQL compared to commercially available relational DBMSs.

The Architecture of Database System

Linking the information back together is the key to this system. In the relational model, some bit of information was used as a "key" uniquely defining a particular record. When information was being collected about a user, information stored in the optional tables would be found by searching for this key. For instance, if the login name of a user is unique, addresses and phone numbers for that user would be recorded with the login name as its key. This simple "re-linking" of related data back into a single collection is something that traditional computer languages are not designed for.

Database Management Systems (DBMSs)

The 1980s ushered in the age of desktop computing. The new computers empowered their users with spreadsheets like Lotus 1-2-3 and database software like BASE. The BASE product was lightweight and easy for any computer user to understand out of the box. C. Wayne Ratliff the creator of BASE stated.

Design and modeling

- A database built with one DBMS is not portable to another DBMS (i.e., the other DBMS cannot run it). However, in some situations it is desirable to move, migrate a database from one DBMS to another. The reasons are primarily economical (different DBMSs may have different total costs of ownership or TCOs), functional, and operational (different DBMSs may have different capabilities). The migration involves the database's transformation from one DBMS type to another. The transformation should maintain (if possible) the database related application (i.e., all related application programs) intact. Thus, the database's conceptual and external architectural levels should be maintained in the transformation.

What Is a Data System, Anyway?

- Historically, data systems have been developed by a centrally managed group, which took responsibility for all aspects of the system. None of these systems provided for the complete range of data system functionality-from discovery to analysis-over a broad range of data providers and data types. The first generation of data systems involved a single computer that could be accessed only locally. Such systems consisted of data, a search capability, and an ability to manipulate the data.
- Many useful types of information are missing in widely used databases; little incentive currently exists to (re)supply the missing data. As a standard practice, funding agencies should require the submission of fully described results to public databases. To minimize the risk of human error during data submission, databases must implement appropriate curation protocols and supporting software. Since errors in data repositories and databases are a known problem, data providers should establish reliable means of reporting, tracking, and correcting errors in a timely manner.

Hierarchical DBMS

- A DBMS is said to be hierarchical if the relationships among data in the database are established in such a way that one data item is present as the subordinate of another one or a sub unit. Here subordinate means that items have "parent-child" relationships among them. Direct relationships exist between any two records that are stored consecutively. The data structure "tree" is followed by the DBMS to structure the database. No backward movement is possible/allowed in the hierarchical database.
- Some well-known database systems that use the network model include: Integrated Data Store (IDS)
- IDMS (Integrated Database Management System)
- RDM Embedded
- RDM Server
- TurbolMAGE
- Univac DIMS-1100

Data (computing)

- Keys in data provide the context for values. Regardless of the structure of data, there is always a key component present. Data keys in data and data-structures are essential for giving meaning to data values. Without a key that is directly or indirectly associated with a value, or collection of values in a structure, the values become meaningless and cease to be data. That is to say, there has to be at least a key component linked to a value component in order for it to be considered data. Data can be represented in computers in multiple ways, as per the following examples
- Until the advent of non-volatile computer memories like USB sticks, persistent data storage was traditionally achieved by writing the data to external block devices like magnetic tape and disk drives. These devices typically seek to a location on the magnetic media and then read or write blocks of data of a predetermined size. In this case, the seek location on the media is the data key and the blocks are the data values.

Databases Use Metadata

- The term "metadata" was coined in 1968 by Philip Bagley, in his book "Extension of programming language concepts" where it is clear that he uses the term in the ISO 11179 "traditional" sense, which is "structural metadata" i.e. "data about the containers of data"; rather than the alternate sense "content about individual instances of data content" or metacontent, the type of data usually found in library catalogues. Since then the fields of information management, information science, information technology, librarianship and GIS have widely adopted the term. In these fields the word metadata is defined as "data about data"'. While this is the generally accepted definition, various disciplines have adopted their own more specific explanation and uses of the term.
- Metadata (metacontent), or more correctly, the vocabularies used to assemble metadata (metacontent) statements, are typically structured according to a standardized concept using a well-defined metadata scheme, including: metadata standards and metadata models. Tools such as controlled vocabularies, taxonomies, thesauri, data dictionaries and metadata registries can be used to apply further standardization to the metadata. Structural metadata commonality is also of paramount importance in data model development and in database design.

Metadata Usage

- Data virtualization has emerged as the new software technology to complete the virtualization stack in the enterprise. Metadata are used in data virtualization servers which are enterprise infrastructure components, alongside database and application servers. Metadata in these servers are saved as persistent repository and describe business objects in various enterprise systems and applications. Structural metadata commonality is also important to support data virtualization.
- Metadata is "information about information" and it is one of the really useful features of digital audio files. When audio went from analogue to digital, it became possible to label or encode audio files with more information than could be contained in just the file name. That descriptive information is called "metadata“.

Purposes of an MIS

- A database management system (DBMS) is an aggregate of data, hardware, software, and users that helps an enterprise manage it operational data. The main function of a DBMS is to provide efficient and reliable methods of data retrieval to many users. For example a college has 10,000 students each year and each student can have approximately 10 grade records per year, then over 10 years, the college will accumulate 1,000,000 grade records. It is not easy to extract records satisfying certain criteria from such a set, and by current standards, this set of records is quite small! Given the current concern for "grade inflation“, a typical question that we may try to answer is determining the evolution of the grade averages in introductory programming courses over a 10-year period. Therefore, it is clear that efficient data retrieval is an essential function of database systems.
- Users interact with database systems through query languages. The query language of a DBMS has two broad tasks: to define the data structures that serve as receptacles for the data of the database, and to allow the speedy retrieval and modification of data. Accordingly, we distinguish between two components of a query language: the data definition component and the data manipulation component. The main tasks of data manipulation are data retrieval and data update. Data retrieval entails obtaining data stored in the database that satisfies a certain specification formulated by the user in a query. Data updates include data modification, deletion and insertion.

Data Warehouse Overview

- Types of systems: Data mart - A data mart is a simple form of a data warehouse that is focused on a single subject (or functional area, such as sales, finance or marketing. Online Analytical Processing (OLAP) - Is characterized by a relatively low volume of transactions. Queries are often very complex and involve aggregations. Online Transaction Processing (OLTP) - Is characterized by a large number of short on-line transactions (INSERT, UPDATE, DELETE). OLTP systems emphasize very fast query processing and maintaining data integrity in multi-access environments. Predictive Analysis - Predictive analysis is about finding and quantifying hidden patterns in the data using complex mathematical models that can be used to predict future outcomes. Predictive analysis is different from OLAP in that OLAP focuses on historical data analysis and is reactive in nature, while predictive analysis focuses on the future. These systems are also used for CRM (Customer Relationship Management).

Data warehouses vs Operational systems

- Operational systems are optimized for preservation of data integrity and speed of recording of business transactions through use of database normalization and an entity-relationship model. Operational system designers generally follow the Codd rules of database normalization in order to ensure data integrity. Codd defined five increasingly stringent rules of normalization. Fully normalized database designs (that is, those satisfying all five Codd rules) often result in information from a business transaction being stored in dozens to hundreds of tables. Relational databases are efficient at managing the relationships between these tables. Data warehouses are optimized for analytic access patterns. Analytic access patterns generally involve selecting specific fields and rarely if ever 'select ' as is more common in operational databases. Because of these differences in access patterns, operational databases (loosely, OLTP) benefit from the use of a row-oriented DBMS whereas analytics databases (loosely, OLAP) benefit from the use of a column-oriented DBMS. Unlike operational systems which maintain a snapshot of the business, data warehouses generally maintain an infinite history which is implemented through ETL processes that periodically migrate data from the operational systems over to the data warehouse.