TAHANI ALMUSHAILI. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 13, Issue 3, March 2023, pp. 109-114

RESEARCH ARTICLE OPEN ACCESS

Database Management Systems.

TAHANI ALMUSHAILI
**Public Authority for Applied Education and Training*

ABSTRACT
I have authored this paper to illustrate the History, Types, and Structure of Database Management Systems and their dependencies. Aiming to give a solid overview of the terms database and DBMS. Also, to help the reader create a global, comprehensive picture of the database and DBMS.We start the paper by listing the history of databases and storing data techniques. Then the terms data and types of data. After that, we start defining database and DBMS and their types. We have an extensive illustration of database objects. We give an overview of SQL, DML, TCL, DQL, and DDL as Structured Query Language. Finally, we ended the paper by talking about database normalization and its levels.

Keywords- Database, DBMS, SQL, Normalization, Database Objects.

Date of Submission: 12-03-2023 Date of acceptance: 29-03-2023

I. INTRODUCTION

The first database concept was introduced in the 1960s.In the 1950s and early 1960s Magnetic tape was introduced for data storage in the mid-1960s.

Reading data from multiple tapes and writing data to a new tape involve the data storage process. Process punched cards were used for input, and the printer was used for output. Salary increases, for example, were managed. Storage data on tapes are read sequentially. The data is much larger than the main memory. The database was accessed via low-level pointer operations. The type of data to be stored dictated the storage information. Thus, adding a new field to your database necessitates rewriting the underlying access. In 1972, two major relational database systems were developed at UBC, and system R. E.F Codd proposed the relational model for the database in a seminal paper on database thinking. This provided a new perspective on the database. He separates the schema of a database's logical organization from its physical storage methods. It also aided in the development of SQL/DS. Between 1974 and 1977, two key relational database system prototypes were constructed: Ingres, developed at UBC, and System R, developed at IBM San Jose. Ingres employed the QUEL query language, which inspired the development of systems like Ingres Corp., MS SQL Server, Sybase, Wang's PACE, and Britton-Lee. System R, on the other hand, used the SEQUEL query language and contributed to the development of SQL/DS, Oracle, Non-Stop SQL DB2, and so on.P-Chen creates the ER Entity-Relationship database concept in 1976. The data applications are the core of this architecture. It provided another crucial understanding of the theoretical data model. The designers may concentrate on the facts rather than a logical framework thanks to the higher-level molding.

SQL (Structured Query Language) became mainstream in the 1980s. IBM released DB2. The evolution of the IBM PC gave rise to various DB businesses and products such as RIM, OS/2, PARADOX, RBASE, and Database Management System. As computer sales increased rapidly, the relational database system became a success. This expanded the adoption of network and hierarchical database models.Much effort was done in the early 1990s on client tools for application development, such as PowerBuilder (Microsoft) and Oracle Developer. The client-server computer approach established the standard for future business choices. Work on the object database system began this year as well. The more complicated database system was introduced at a great cost. This year, new client tools for application development were launched. Personal productivity tools such as ODBC and Excel/Access were also created.In the mid-1990the internet was launched in the middle of this year. It enables remote access to computer systems. The notion of Web/DB grew. The database system was upgraded. The client-server database was first used by ordinary PC users.

The late 1990s saw an upsurge in the need for internet database connectors such as Active server page, Java Servlets, FrontPage, and Dream Weaver. Enterprise Java Beans, Oracle Developer 2000, and other internet firms working on

Web/Internet connections this year. With the advent of online analytic processing (OLAP), many shops began to use point-of-sale (POS) technology on a regular basis.

In the early twenty-first century, the expansion of database applications continued. PDAs enabled more interactive applications to emerge. IBM dominated the huge DB market. Microsoft and Oracle.

II. BASIC CONCEPTS

II.1. Data and types of data

We must understand what data is, what sorts of data there are, and how to classify it before we can define a database. Data are pieces of information that have been transformed into a format that can be easily moved about or processed by computers. In terms of today's computers and transmission technologies, data is information that has been converted into binary digital form. Data may be used as either a singular or plural topic. The phrase "raw data" refers to information in its simplest digital form.Structured data, unstructured data, and semi-structured data are the three types of data from a database perspective.Let us explain them more:

II.1.1. Structured data.

Most of us are accustomed to working with structured data, which is most frequently categorized as quantitative data. Consider the type of information that neatly fits into the defined rows and columns of relational databases and spreadsheets.Names, dates, addresses, credit card numbers, stock information, geolocation, and other information are examples of structured data.Machine language can quickly and readily comprehend highly ordered data. A relational database management system enables users of relational databases to input, search for, and manipulate structured data extremely fast (RDBMS). The most appealing aspect of organized data is this.The computer language used to manage structured data is known as a structured query language, or SQL. This language was created by IBM in the early 1970s and is particularly good for dealing with database relationships.

II.1.2. Unstructured data.

Most frequently classified as qualitative data, unstructured data cannot be handled or evaluated using standard data tools and techniques.Text, audio and video files, social network posts, mobile activity, satellite photography, surveillance imaging, and more are examples of unstructured data. Since unstructured data lacks a predetermined data model and cannot be arranged in relational databases, it is challenging to deconstruct. For managing unstructured data, non-relational or NoSQL databases are preferable.Allowing unstructured data to flow into a data lake and remain there in its unprocessed state is another approach to handle it.

II.1.3. Semi-structured data

Semi-structured data is a sort of structured data that falls somewhere in between structured and unstructured data. It does not use a relational or tabular data format, but it does feature tags and semantic markers that scale data into records and fields in a dataset. JSON and XML are common instances of semi-structured data. Semi-structured data has a higher level of complexity than structured data but is less complicated than unstructured data. It is also easier to store than unstructured data and bridges the gap between the two.

II.2. What is a database?

User data, metadata, and any other data are stored in a common, integrated computer structure. Computer database is essential for effective data management.

II.2.1. Database Definition.

Another definition is that database is a structured arrangement of material that has been arranged and is often stored virtually in a computer system. A database management system often controls a database (DBMS). The term "database system," which is sometimes abbreviated to "database," refers to the combination of the data, the DBMS, and the applications that are connected to it.To facilitate processing and data querying, the most popular types of databases now in use usually describe their data as rows and columns in a set of tables. The data may then be handled, updated, regulated, and structured with ease. For creating and querying data, most databases employ structured query language (SQL).

II.2.2. Types of Databases.

There are several varieties of databases. The ideal database for a given company will depend on how it plans to use the data.

II.2.2.1. Relational databases

Relational databases took over. A relational database's items are arranged as a collection of tables with rows and columns. The most effective and adaptable method of gaining access to structured data is through relational database technology.

II.2.2.2. Object-oriented databases

Like object-oriented programming, information is represented as objects in an object-oriented database.

II.2.2.3. Distributed database

Two or more files spread over several separate sites make up a distributed database. The database may be

spread across several computers, one physical location, or a number of separate networks.

II.2.2.4. Data warehouses

A data warehouse is a database created primarily for quick query and analysis. It serves as a central store for data.

II.2.2.5. NoSQL databases

A NoSQL, or nonrelational database, stores and manipulates unstructured and semi-structured data (in contrast to a relational database, which defines how all data inserted into the database must be composed). As online applications got more frequent and complicated, NoSQL databases gained popularity.

II.2.2.6. Graph databases

A graph database holds data in terms of entities and their relationships.Relationships may be stored and accessed using graph databases. In graph databases, relationships are treated as first-class citizens and provide most of their value. In graph databases, data items are stored as nodes, while connections between nodes are stored as edges. A parent-child connection, an action, ownership, and other concepts may all be described by an edge, which always has a start node, an end node, a type, and a direction. The quantity and variety of connections that a node can have are both unlimited.

II.2.2.7. OLTP databases.

An OLTP database is a fast, analytical database built to handle high quantities of transactions from various users.

II.2.2.8. Cloud databases

A cloud database is a collection of organized or unstructured data stored on a private, public, or hybrid cloud computing platform. Traditional cloud database models and database as service models exist (DBaaS). Administrative and maintenance activities are handled by DBaaS.

II.3. Database Management System Definition.

A group of programs that manage the database and create access to it, is known as a database management system.In other words, A DBMS, which is a complete database software application, is often necessary (DBMS). A DBMS acts as a conduit between the database and the applications that will use it, enabling users to access, modify, and control how the data is arranged and optimized. A DBMS also makes it easier to monitor and manage databases, enabling a number of administrative tasks including performance tweaking, backup, and recovery. Popular database management systems (DBMSs) include dBase, MySQL, Microsoft Access, Microsoft SQL Server, and Microsoft Access.

II.4. Database Objects

A data structure used to either store or refer to data is referred to as a database object in a relational database. The table is the item that people use the most frequently. Indexes, stored procedures, sequences, views, and many others are examples of other objects. A new object type cannot be formed while creating a database object because all existing object types are constrained by the source code of the relational database model being used, such as Oracle, SQL Server, or Access. Instances of the objects, such as a new table, an index on that table, or a view on the same database, are what is being produced.

We will list all database objects from our perspective below:

II.4.1. The Database Schema.

The skeletal structure of a database schema is the logical representation of the whole database. It outlines the structure of the data and the relationships that link them. It lays forth every restriction that must be imposed on the data. A database schema outlines its entities and their relations. It includes descriptive information about the database that may be shown using schema diagrams. The schema is created by database designers to make the database easier for developers to comprehend and utilize.

II.4.2. Indexes.

Indexes are a strong tool a database uses in the background to accelerate querying. By offering a way to rapidly find the required data, an index powers the query.A database index's main function is to enhance query performance by accelerating data retrieval. The B-TREE data structure and the references to the real data must be stored in extra storage because of this.By using indexes, a database may rapidly discover data without having to go through each row of a table each time it is searched. Given the circumstances; indexes offer a pretty effective technique to obtain sorted records.

II.4.3. Tables.

All the data in a database is stored in tables, which are database objects. Data is logically arranged in tables using a row-and-column layout akin to a spreadsheet. Each column is a record field, and each row denotes a distinct record. For instance, a table containing employee information for a business may include a row for every employee and columns for details like employee number, name, address, job title, and home phone number.

II.4.4. Fields and columns.

Fields and columns arepart of database tables that define the type of data stored in the table in a

vertical way. In another way, the table is consisting of a list of columns that store and constrain their below cells type, size, and nullability.

II.4.5. Records and rows.

As records and rows are types of structured database objects each info inserted on the table must be on records. We can say that table is consisting of fields and records which keep it structured and increase information readability.

II.4.6. Keys.

A tuple (or row) in a relationship can be uniquely identified by a collection of properties called a key in a database management system (DBMS) (or table). A relational database's many tables and columns can be linked to one another via keys. Key values are the specific values included within a key.

In DBMS, keys often come in lots of various categories. To prevent repetition, the relevant database must implement each of these types of keys in SQL correctly. Accurate identification will increase database accuracy, which will improve outcomes quickly. Let us examine these DBMS keys:

II.4.6.1. Primary Key

If you are wondering what a primary key is in a database management system (DBMS), it is a column or group of columns that help identify each item in a table individually. In a table, there can only be one primary Key. Additionally, the primary Key cannot have any row with the same values repeated. The primary key's values must all be unique and include no duplicates. To grasp what are keys and what is a primary key in DBMS, a primary key is the most important feature among many others.

A column or collection of columns with the PRIMARY KEY (PK) constraint applied will not be able to have any null values or duplicate data. There can only be one primary key restriction per table.

II.4.6.2. Foreign keys

Foreign keys are used to establish connections between two tables. A column or group of columns having a foreign key must have all of its values match the primary key of the referential table. By using foreign keys, data and referential integrity are maintained.For two entities that could be connected by the same information but do not share identifiable information to remain distinct, foreign keys are crucial. In these situations, the tables are connected to preserve the relationship but do not totally function as a substitute for one another.

II.4.6.3. Composite Key

Each tuple in a table can be uniquely identified by a collection of two or more properties called a composite key. When taken independently, the features in the collection might not be distinctive. But when combined, they will guarantee originality. A composite key is sometimes known as a "concatenated key."

II.4.7. Relationships.

When two relational database tables have a foreign key that refers to the primary key of the other table, this is referred to as a relationship in the context of databases. Relational databases may divide and store data in several tables while linking dissimilar data items thanks to relationships.There are three types of database relationships based on the number of links between records on tables.

II.4.7.1. One to One Relationship (1:1):

A single row of the first table can only be associated with one and only one record of the second table in a connection known as a one-to-one relationship (1:1). In an equivalent manner, each row of the first table may be connected to any row of the second table.

II.4.7.2. One to Many Relationships:

It is utilized to develop a connection between two tables. The rows of the second table can relate to one or more rows of the first table, but they can only relate to one row in the first table. Another name for it is a many-to-one connection.

II.4.7.3. Many to Many Relationships:

Two tables are related to one another through many to many connections. Any record from the first table may be related to any record from the second table, or not. Every record in the second table can also be related to more than one record in the first table, in an equivalent manner. An N: N connection is another way to describe it.

II.5. Structured Query Language (SQL)

Nobody can make a paper related to databases without defining SQL. A standardized programming language called Structured Query Language (SQL) is used to handle relational databases and carry out different operations on the data they contain. Originally developed in the 1970s, SQL is frequently used by database administrators as well as programmers developing scripts for data integration and data analysts setting up and performing analytical queries.

It is a language for dealing with and controlling interconnected databases by dealing with data structures and performing data entry, deletion, sorting, searching, filtering, modifying, and so on.

SQL is divided into four types:

II.5.1. Data Definition Language (DDL)

In the current database industry, DDL is a part of any formal language used to describe data. Nevertheless, it is regarded as a subset of SQL (Structured Query Language). When implementing database changes, SQL frequently pairs imperative verbs with statements that are written in standard English. DDL defines modifications to the database

schema but does not appear in a SQL database as a distinct language.By interacting with descriptions of the database schema, it is used to construct and alter the structure of objects in a database. DDL instructions, as opposed to data manipulation language (DML) commands, are used to change the database structure, such as by adding new tables or objects and all their characteristics (data type, table name, etc.).The most common DDL commands are CREATE, ALTER, DROP, and TRUNCATE.

II.5.2. Data Manipulation Language (DML)

A family of computer languages known as a data manipulation language (DML) includes instructions that let users change data in a database. This manipulation entails adding new data to database tables, accessing old data, removing old data from tables, and changing old data. The majority of DML is included in SQL databases. DML is similar to plain English and improves effective user engagement with the system. The manipulation commands SELECT, UPDATE, INSERT INTO, and DELETE FROM makeup DML's functional capacity.

II.5.3. Data Query Language (DQL)

A DQL clause in a SQL statement enables you to get and arrange data from a database. The SELECT command may be used to obtain data from a database so that you can execute operations on it. It is equivalent to the relational algebraic projection operation. A new temporary table is created from the results of a SELECT operation on a table or set of tables and is then displayed or sent to a program.

The database server looks at the chosen clause, which would be the first and one of the last clauses of a select statement. The reason for this is that we need to be aware of all the potential alternate columns before we can decide what to include in the final result set.

II.5.4. Transaction Control Languages (TCL)

The Structured Query Language (SQL) is used in Relational Database Management Systems (RDBMS) to carry out a number of actions to save, retrieve, and alter data across different tables in a database. Consider a few situations where we could have accidentally altered a record and want to restore the data, or where we might have entered a few records and want to save them. In these cases, Transaction Control Language (TCL) enters the picture. The consistency and integrity of the data kept in the database are upheld through the usage of the transaction control language.

II.6. Database Normalization.

Normalization is a way to design databases to avoid redundancy and get rid of some unwanted operations such as insertion, update, and deletion. It also divides large tables into small tables and ensures that data is stored logically.Normalization has many levels or Forms, and each level has some rules to ensure that the normalization technique is applied correctly.With the introduction of the First Normal Form, the relational model's creator Edgar Codd put out the notion of data normalization, and he later expanded it with the Second and Third Normal Forms. Later, he collaborated with Raymond F. Boyce to create the Boyce-Codd Normal Form theory.

Normalization is the process of removing anomalies and redundancies from database design.[5]

Let us list normalization Forms and their rules:

II.6.1. First normal form (1NF).

Which resembles Step one of the normalization procedures more. The first normal form assumes that you will create your table in a way that makes it simple to add to it and to get data, as necessary.

A table must go by the following four guidelines to be in the First Normal Form:

- It should only include attributes and columns with a single value.
- Column values should all belong to the same domain.
- A table's columns should each have a name that is unique.
- It is unnecessary in whatever sequence the data are stored.

II.6.2. Second Normal Form(2NF).

A table must be in the First Normal form and free of Partial Dependency to be in the Second Normal form.When an attribute in a table totally depends on a portion of a composite primary key rather than the entire primary key, this is known as partial dependence.We may separate the table, delete the property that is generating partial Dependency, and relocate it to another table where it will fit in nicely to eliminate partial dependency.

II.6.3. Third Normal Form (3NF)

An improvement in the Second Normal Form is the Third Normal Form. A table is in the Third Normal Form when it is in the Second Normal Form and has no transitive dependencies. Removing transitive dependencies has the advantage of lessening data duplication and preserving data integrity.

II.6.4. Boyce-Codd Normal Form (BCNF)

The 3.5 Normal Form, sometimes referred to as the Boyce-Codd Normal Form (BCNF), is an extension of the third normal form.

A table must meet the two requirements listed below tofulfill the Boyce-Codd Normal Form:

- It must be in the Third Normal Form

- A table T in a database schema with FD set F is said to be in Boyce-Codd normal form (BCNF) when the following property holds. For any functional dependency X → A implied by F that lies in T, where A is a single attribute that is not in X, X must be a super key for T. A database schema is in BCNF when all the tables it contains are in BCNF.[6]

III. CONCLUSION

Data are pieces of information that have been transformed into a format that can be easily moved about or processed by computers. There are three types of data from a database perspective: structured, unstructured, and semi-structured. Data may be used as either a singular or plural topic. Semi-structured data is a type of structured data that falls somewhere between structured and unstructured. Semi-structuring data does not use a relational or tabular data format but instead features tags and semantic markers that scale data into records and fields in a dataset.A NoSQL, or nonrelational database, stores and manipulates unstructured and semi-structured data. A graph database holds data in terms of entities and their relationships. An OLTP database is a fast, analytical database built to handle high quantities of transactions. The table is the item that people use the most frequently. Indexes, stored procedures, sequences, views, and many others are examples of other objects. A new object type cannot be formed while creating a database object because all existing object types are constrained by the source code of the database.All the data in a database is stored in tables, which are database objects. Data is logically arranged in tables using a row-and-column layout. A table is consisting of a list of columns that store and constrain their below cells type, size, and nullability. Data and referential integrity are preserved with the use of foreign keys. There are three types of database relationships based on the number of links between records on tables.Each row of the first table can only be associated with one and only one record of the second table in a connection called a 1:1 relationship. The Structured Query Language (SQL) is a programming language for dealing with relational databases. It was originally developed in the 1970s and is used by database administrators and data analysts. Data Manipulation Language (DML) is a subset of the formal language used to describe data. DDL defines modifications to the database but does not appear in a database as a separate language. The most common DDL commands are CREATE, ALTER, DROP, and TRUNCATE. The SELECT command may be used to obtain data from a database so that you can execute operations on it. The Structured Query Language (SQL) is used in Relational Database Management Systems (RDBMS) to carry out actions to save, retrieve, and alter data across different tables. A table must go by the following four guidelines to be in the First Normal Form. Removing transitive dependencies has the advantage of lessening data duplication. The 3.5 Normal Form, sometimes referred to as the Boyce-Codd Normal Form (BCNF) is an extension of the third normal form.

REFERENCES

[1]. C. Coronel, S. Morris, database systems: design, implementation, and management, ninth edition (Cengage Learning, 20 Channel Center Street Boston, MA 02210 USA, 2011).
[2]. C. Coronel, S. Morris, database systems: design, implementation, and management, Twelfth edition (Cengage Learning, 20 Channel Center Street Boston, MA 02210 USA, 2017).
[3]. C. Coronel, S. Morris, database systems: design, implementation, and management, Twelfth edition (Cengage Learning, 20 Channel Center Street Boston, MA 02210 USA, 2017).
[4]. RamezElmasri, Shamkant B. Navathe,fundamentals of database systems, seventh edition (Pearson Education, 221 River Street, Hoboken, NJ 07030, 2016).
[5]. Steve Conger, hands-on database an introduction to database design and development (Pearson Education, One Lake Street, Upper Saddle River, New Jersey 07458, 2012).
[6]. S. Buxton, L. Fryman, R. H. Güting, T. Halpin, Jan L. Harrington, William H. Inmon, Sam S. Lightstone, Jim Melton, Tony Morgan, Thomas P. Nadeau, Bonnie O'Neil, Elizabeth O'Neil, Patrick O'Neil, Markus Schneider, Graeme Simsion, Toby J. Teorey, Graham Witt, Database Design: Know It All (Elsevier Inc, 30 Corporate Drive, Suite 400 Burlington, MA 01803, 2009).