

What is RDBMS, its usage and structure?

Edris Safari

Bellevue University, Nebraska U.S.A.

### Abstract

Ever since their advent, computers we have been used to store data in such a way so as it is persistent in keeping the data up to date and safe from loss and also to give us the ability to retrieve the data in such a way so as it is fast, reliable, and easy.

Obviously, computers themselves do not offer any of these features. It is the software application running on the computer that is designed to offer these features. The software application is widely known in the industry as database management system.

Relational database management systems (RDBMS) are based on a mathematical modeling that involves set theory and calculus. This paper describes RDBMS, its usage and structure.

*Keywords:*

RDBMS: Relational Database Management System

What is RDBMS, its usage and structure?

### **Introduction**

In the early days, data were stored in files. The files resided on the hard disk of the computer and the application that managed the data (Database Management system) manipulated the files in the process of various transactions. The transactions were analogous to adding data, updating data, and removing data. The files were structured in what was called ISAM (Indexed Sequential Access Method). Data was structured in a hierarchical fashion where the parent nodes lead to leaf nodes by keys. The keys were unique and would point to other keys down the hierarchy until the data is found at the leaf level. This structure allowed retrieval of data based on equality or range criteria (i.e. name = "john" or employees with ID between 1 to 100). While it served a lot of purposes, ISAM was not suitable for large scale data management.

IBM pioneered relational database management with their introduction of DB2 after Edgar F. Cobb proposed it in 1970(Ref. 6). Along with this new concept came structure query language or SQL (Sequel in the lingo). SQL was and still is the "front-end" to the underlying database management system. The language is based on a mathematical framework that allows us to manipulate data in fast and reliable ways.

Relational databases are composed of tables (as opposed to files in ISAM). The tables have columns and rows. Certain columns are designated such that their value is unique. These columns are known as primary keys. Two tables could have the same column defined as their respective primary key. For example, we could have table A with primary key EmployeeID and columns for FirstName, LastName, Address. Table B also stores EmployeeID as primary key, but it has employee's hire data, rank, and salary. Storing data in tables in such a form allows us to combine data from several tables based on their primary key relationship. One table could be

related to multiple other tables. This allows for a one-to-one relationship between two tables, one-to-many, many-to-one, and many-to-many relationships. With this, we could store and retrieve multiple records for one key. For example, we could get all traffic violations by a person based on her/his driver's license number. In one table indexed by driver's license number and another table by drivers license number and TicketID where there could be 0 or more records for each driver's license number.

The database management systems such as Oracle, Ingress, SQL server, and the like are all based on the same relational model and use SQL as the interface. They all proprot to have features that support requirements such as fault tolerant, scalable, failover mechanism, and other features that are mission critical. The usage of RDBMS and for that matter database management systems has been ubiquitous in every sector that deals with the storage and retrieval of data and this includes data science. However, data science has some new requirements that the traditional database management systems cannot offer. This is mainly due to the increase in the volume of data and the need to be able to process that data effectively.

### **Conclusion**

Relational databases have been in existence for decades. While the underlying architecture is simple and the SQL language provides a good interface, the management of the data in a distributed environment poses a big challenge. New open source applications have been emerging and are challenging RDBMS. Data warehousing, Lake or Swamp in large data centers have given way to new applications such as Hadoop, Spark and MPP (Massively Parallel Processing) (Ref. 3) that are distributed and can process in parallel. As the volume of inevitably increases, so will the demand for change and new innovations.

## References

1. Explore the differences between ISAM and relational databases -  
<https://www.techrepublic.com/article/explore-the-differences-between-isam-and-relational-databases/>
2. What is a Relational Database Management System? -  
<https://www.codecademy.com/articles/what-is-rdbms-sql>
3. Everything a Data Scientist Should Know About Data Management -  
<https://towardsdatascience.com/everything-a-data-scientist-should-know-about-data-management-6877788c6a42>
4. Sarkar, Dr. Tirthajyoti. Data Wrangling with Python: Creating actionable data from raw sources. Packt Publishing. Kindle Edition.
5. DBMS - Index Sequential Access Method (ISAM) -  
<https://www.youtube.com/watch?v=EiW1VVPor10>
6. Relational Database - <https://www.ibm.com/ibm/history/ibm100/us/en/icons/reldb/>