iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4133

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

A Comparative Study on Open Source Database Management System

Naresh Purohit [1], Kuldeep Dave [2], Rupesh Sharma [3]
[1, 3] *Department of Computer Science Engineering, Mahaveer Institute of Technology & Science, Pali , Rajasthan*
[2] *Department of Electrical Engineering, Mahaveer Institute of Technology & Science, Pali, Rajasthan*

Abstract: Database stores data in an organized manner so that its retrieval becomes easy. Also, the management of data is easier when it's stored in a database. This paper gives a comparative study on some open source databases that there are many factors to be considered, however, before choosing a database for a particular software application.
Here, we are going to try to understand the core differences of some of the most commonly used and popular relational database management systems (RDBMS). We will explore their fundamental differences in terms of features and functionality, how they work, and when one excels over the other in order to help developers with choosing a RDBMS.
Keywords: Database, Open Source, Software Application, RDBMS.

I. INTRODUCTION

Databases are logically modelled storage spaces for all kinds of different information (data). Each database, other than schema-less ones, has a model, which provides structure for the data being dealt with. Database management systems are applications (or libraries) which manage databases of various shapes, sizes, and sorts. A DBMS enables users to define, create, maintain and control access to the database. It is the software that interacts with the users' application program and the database. Typically, a DBMS provides the following facilities:

A. It allows users to define the database, usually through a Data Definition Language (DDL). The DDL allows users to specify the data types and structures and constraints on the data to be stored in the database.
B. It allow user to insert, update, delete and retrieve data from the database, usually through a Data Manipulation Language (DML). Having a central repository for all data and data descriptions allow the DML to provide a general inquiry facility to this data, called a query language. The most common query language is the Structured Query Language (SQL).
C. It provides controlled access to the database. For example it may provide:
D. A security system, which prevents unauthorized users accessing the database and also prevent unauthorized transactions.
E. An integrity system, which maintains the consistency and validity of stored data.
F. A concurrency control system, which allows multiple users to access the database simultaneously.

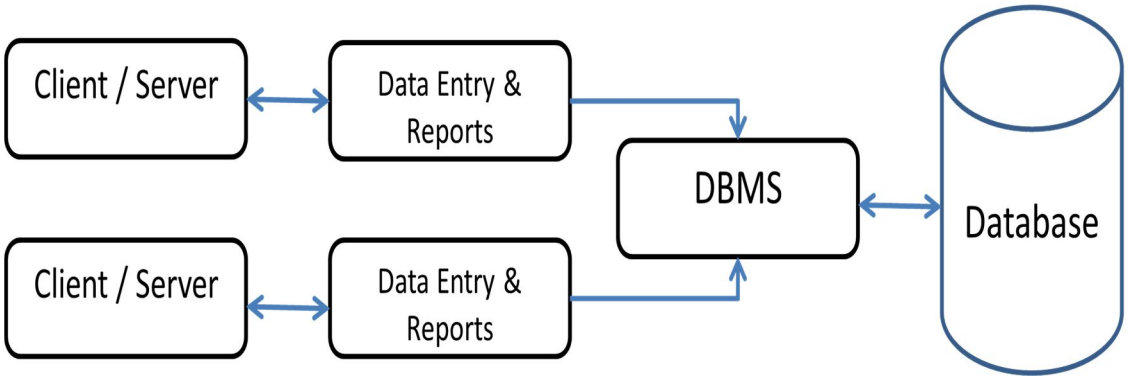


Fig. 1 Database Processing

It is really difficult and time consuming for a database to operate if it has a huge load of data or if it has a variety of data sets to operate upon. There are also several other factors like the security of data as well as the cost associated with the database, impacting the choice of database. A free database helps users to avoid huge costs. Hence open source databases have been playing quite and important role in many of the widely used software applications across the globe.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

II. DIFFERENT KIND OF DATABASES

There are various types of databases being used by different users across the globe, based on their requirement. The following is the broad classification of databases.

A. Relational Databases

Relational databases are the most common among all the types. In such databases, the data is actually stored in the form of different data tables. Each of the table has a unique key field and that key is used to connect one table to the other tables. Hence, different tables are related to each other with the help of various key fields. Such databases are widely used in industries like media, telecom, etc, and this is probably the type of database one is most likely to come across.

B. Operational Databases

An operational database is very important for organizations, as it supports the customer database and the inventory database. It help companies to keep track of inventories as well as store details of the customers who buy is products. They can contain things like payroll records, customer information and employee data. They are critical to data warehousing and business analytics operations. The key characteristic of operational databases is their orientation toward real-time operations, compared with conventional databases that rely on batch processing. With operational databases, records can be added, removed and modified in real time. Operational database management systems can be based on SQL but a growing number are using NoSQL and nonstructural data.

C. Distributed Databases

Many organizations have several regional offices, manufacturing plants, branch offices and a head office. Each such work group may have its own set of databases, which collectively from the main database of the company. Such a system of databases is known as a distributed database. So basically a distributed database is a collection of multiple interconnected databases, which are spread physically across various locations that communicate via a computer network.

D. End User Databases

There is a variety of data available at the workstations of different end users of an organization. Each workstation acts like a small database in itself, and has data in the form of spread sheets, presentations, downloaded files, word files and in notepad format. All such small databases together form an end user database

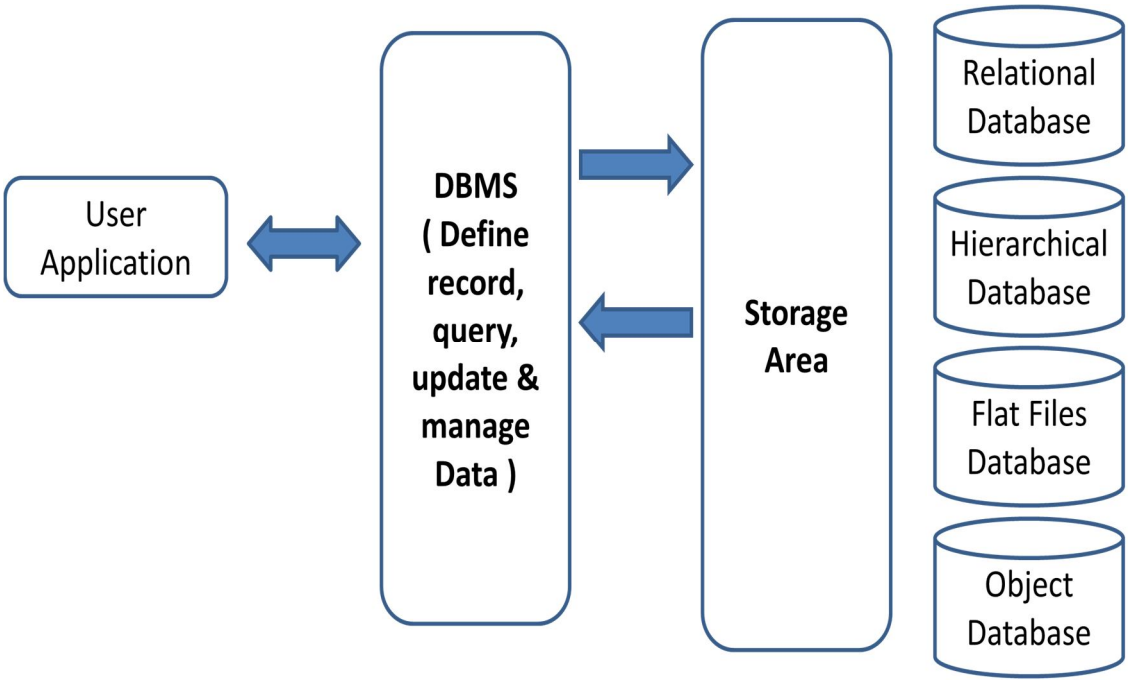


Fig. 2 Interaction of a user with a DBMS

Now let's have a look at some of the popular open source databases available in the market

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

III. MySQL

MySQL has been in use since 1995 and is currently owned by Oracle. In addition to its open source version, there are couple of paid editions available as well, which offer some additional feature, such as cluster geo-replication and automatic scaling. We all know that MySQL has become an industry standard now, as it is compatible with almost every operating system and has been written in both C and C++. The database option is great for different international users as well, since the server can provide various error messages to the clients in a number of languages.

A. *Pros*

1) Offers a flexible privilege and password system.
2) Can be used even if there is no available network.
3) Uses host-based verification.
4) It has security encryption for all the password traffic.
5) Support servers as a separate program for the client server networked environment.

B. *Cons*

1) Users feel that MySQL no longer falls under the free OS category.
2) It is no longer community driven, so bug fixes and patches do not happen in time.
3) Falls behind other similar available options due to its slow updates.

IV. POSTGRESQL

PostgreSQL, also called Postgres, is basically an object relational DBMS with much emphasis on extensibility and standards compliance. As a database server, its important functions are to securely store data and return the data or data sets in response to different requests made from other software applications. PostgreSQL can handle different workloads, ranging from that of small single- machine applications to even large internet-facing applications with many concurrent users. It has been developed by PostgreSQL Global Development Group, and is available as free and open source software.

A. *Pros*

1) It is transactional and ACID (atomicity, consistency, isolation, durability) – compliant.
2) Supports updatable views and materialized views.
3) Functions, stored procedures, triggers, etc, can be very well used in it.
4) Supports concurrency with a help of a system known as MVCC (Multi Version Concurrency Control).
5) Supports a large variety of data types like Boolean, binary, etc.

B. *Cons*

6) It does not have any bug tracker (whereas it supports a bug-submission form, which feeds into the pgsql-bugs mailing list), making it a bit difficult to know the status of bugs.

V. MARIADB

MariaDB is a database that is widely used by tech giants like Wikipedia, Facebook and even Google. It has been developed by various developers who worked as part of MySQL. It is basically a database server that provides drop-in replacement functionality for MySQL. Data security is one of the most significant concerns and priorities for developers of MariaDB, and almost in each of its solution releases the developers merge in all of MySQL's security patches and also enhance the same, if required.

A. *Pros*

1) It provides real-time access to the data sets.
2) It supports maximum number of core functionalities of MySQL.
3) It supports high scalability with easier integration.
4) It provides a couple of alternate patches, storage engines and server optimisations.
5)

B. Cons

1) It does not provide support for the memcached interface.
2) The password complexity plug-in is not available.
3) Has no optimiser trace.

VI. CHOOSING THE BEST AVAILABLE OPEN SOURCE DATABASE

A. Size of Data

When choosing a database, we should always consider the volume of data that we need to retrieve and store as critical application data in a database. The amount of data that we can retrieve and store may vary depending on the combination of the selected data structure, and the ability of any database to differentiate between various data sets available across multiple servers and file systems. Hence we need to chose our database by considering the overall volume of data generated by software application at any specific rate, and also the size of data that needs to be retrieved from the database.

B. Speed & Scalability

We must also gauge the speed that we require for reading and writing different sets of available data into the database – the time taken to service all the incoming reads and writes to any specific application. Some databases are designed to optimise read-heavy applications, whereas others are designed o support write-heavy solutions. Selecting a database that can handle our application's input/output needs can really go a long way to a scalable architecture.

C. Structure of Data

Structure of data set is all about the manner in which we need to store and retrieve our data. Since an application deals with data present in diverse formats, before selecting a database we should consider picking the right data structure for storing and retrieving the data sets. If we fail to select right data structure for persisting available data, the application may take more time to retrieve data from the database.

D. Accessibility of Data

We should also think about the number of users that may concurrently access the database in order to perform any operation on available data, and also the level of computation involved in accessing any set of data. The processing speed of software applications may get affected if the database chosen is not really good enough to handle large loads.

E. Safety & Security of Data

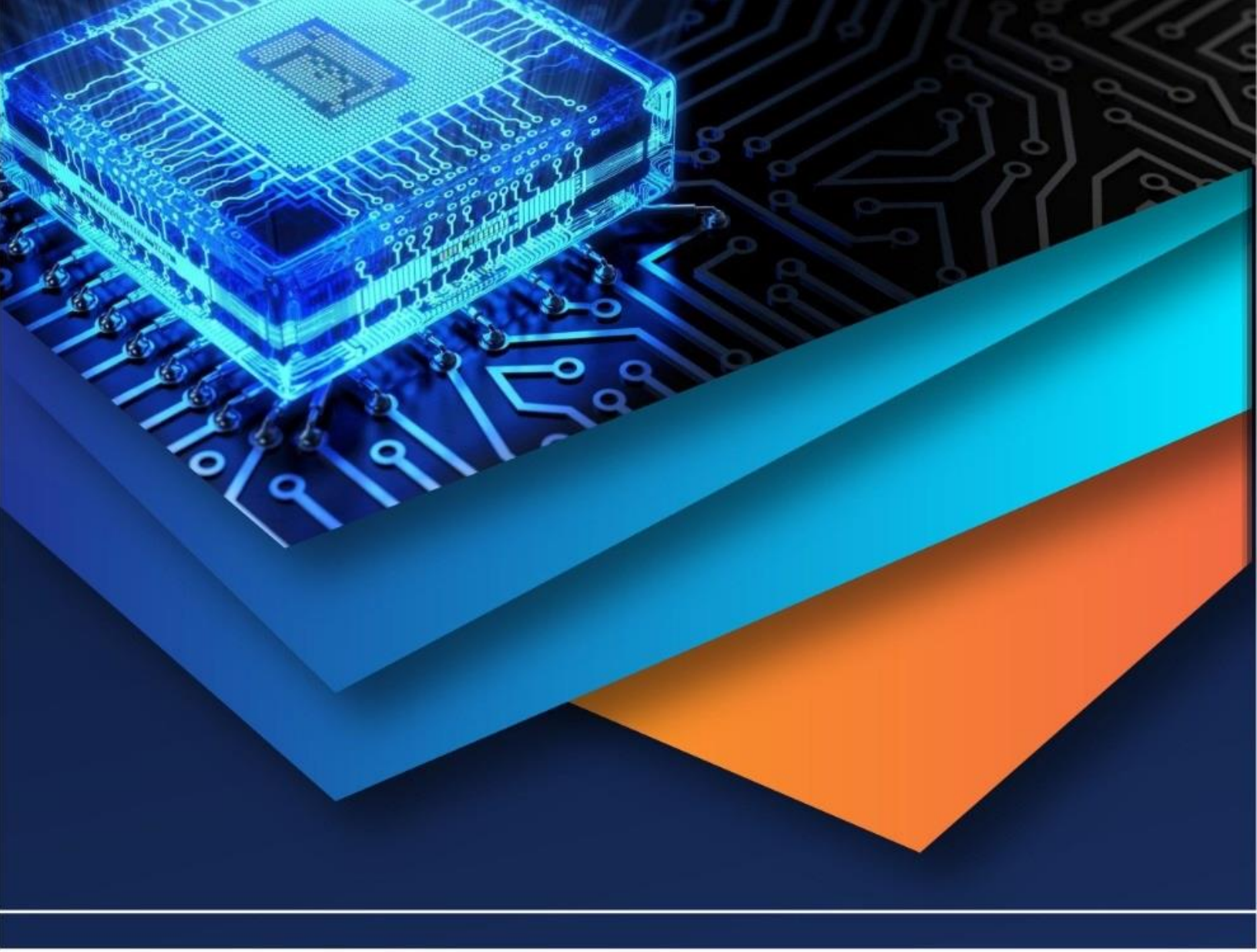
We must also check the level of security that a database provides for the data stored in it. In case of highly confidential data, we really need to have a highly secured database for the application using it. The different safety measures implemented by a database in case of ay system crash or failure are an important factor that we should look at before choosing a database.

VII. CONCLUSION

After researching all the available databases which are free and source code is available, it was found that only PostgreSQL is the most mature, most widely used and robust RDBMS SQL free database (object relational) in the world. PostgreSQL is very appealing as lot of work had already been done and it has ODBC, JDBC drivers using these it is possible to write applications independent of the databases. The applications written in PostgreSQL using ODBC, JDBC drivers are easily portable to other databases like Oracle, Sybase and Informix. And applications written for Oracle, Sybase and Informix using ODBC, JDBC drivers are easily portable to PostgreSQL database.

REFERENCES

[1] "Open Source System for Analysing, Validating, and Storing Protein Identification Data" , Robertson Craig, John P. Cortens, and Ronald C. Beavis, J. Proteome Res., 2004, 3 (6), pp 1234–1242
[2] "PostgreSQL: Introduction and Concepts" , Bruce Momjian, Pearson Education
[3] "PostgreSQL: A Comprehensive Guide to Building, Programming" By Korry Douglas, Susan Dougla
[4] "MariaDB Cookbook" By Daniel Bartholomew, PACKT publishing
[5] "Managing & Using MySQL: Open Source SQL Databases for Managing Information" By Tim King, George Reese, Randy Yarger, Hugh E. Williams, O'REILLY
[6] "MariaDB Essentials" By Emilien Kenler, Federico Razzoli, PACKT publishing.
[7] "PostgreSQL for Data Architects" By Jayadevan Maymala , PACKT publishing.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY