iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.45873

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: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Review of Database Management Systems and Their Categories with Use Cases

Raman Kumar[1], Prof. Vanishree K[2]
[1, 2]*Information Science and Engineering, RV College of Engineering, Bengaluru, India*

Abstract: ***Data has become very important in today's world for industries, education and research, healthcare, and many different services. With the advancements in technologies and tools, data can be preserved and manipulated to draw meaningful conclusions and accomplish tasks. The database management systems provide the platform to store and manage the data and thus perform the expected tasks efficiently. In the paper, popularly used databases are reviewed for the features and the use case scenarios are identified for application. Five common Database management systems categories namely, document store, spatial, graph, search engine and time series database management systems are reviewed with the types of data that has to be stored in that category.***

Keywords: SQL, distributed system, ACID, data, scaling, DBMS

I. INTRODUCTION

The data is becoming increasingly important nowadays. For online businesses, data has become the new currency these days. For research, data is required. Data is required for consistency in operation of software used in Phones/Laptops. Data is utilized in applications associated to businesses, healthcare, education, entertainment and recreation, Due to such widespread usage, there are large number of database providers. The giant technology companies own data centers or buy the data storage from service providers. Startups need to select the service plans from the database management system vendors to subscribe from the requirements and service quality. MySQL databases store the records in many tables. NoSQL is often attributed to its faster speed of execution compared to MySQL

II. DATABASE MANAGEMENT SYSTEMS

The widely used database management systems for storing and querying the data are chosen and reviewed.

A. MySQL

MySQL is now owned by Oracle Corporation. MySQL and Oracle database lie at the top spots in most used databases worldwide. The features of MySQL are standard and robust justifying its position in the top spots. It is used by Facebook, Uber and Youtube. Youtube stores all the metadata for videos on MySQL. The basic features are available for use free of cost. However, the paid versions cost around $USD 2000-1000/month. It is open source and thus opened to advanced development. The MySQL can be used when data is to be structured with guarantee ACID properties in transactions. It is preferred choice if the application has to be scaled horizontally and for Write heavy data. It provides high security of data. It may not be the best to opt for advanced features for data protection (e.g. masking and throttling) and semi structured data like JSON. MySQL is used as a service by ScaleGrid, MySQL Database service with OLTP, OLAP and Multi-Cloud support, Vitess supporting horizontal scaling, Aiven for MySQL with Multi-Cloud service, Amazon RDS for MySQL, Google Cloud SQL for MySQL, Azure MySQL database. ScaleGrid, Vitess, and Oracle MySQL Cloud Service supporting horizontal scaling.

B. Postgre SQL

PostgreSQL popularity has risen in the last few years. It is attributed for its features and innovations. PostgreSQL is open source database which can operate on Windows, macOS and versions of Linux and Unix. It is being used in companies like Spotify, Instagram, Reddit, IMDb, Apple, Skype and NASA organisation. If multiple model databases (including Geospatial data type) are required, the PostgreSQL is a choice. It guarantees the ACID transactions maintaining the data integrity. For extremely relational data or semi structured data, PostgreSQL may not be the best choice of database. PostgreSQL is used as a service in ScaleGrid facilitating horizontal scaling,

©IJRASET: All Rights are Reserved

Azure PostgreSQL database,Amazon RDS for PostgreSQL, EDB for PostgreSQL facilitating horizontal scaling, CitusData facilitating distributed SQL, Google Cloud SQL for PostgreSQL, Aiven for PostgreSQL with multi-cloud support.

C. MongoDB

MongoDB is an object-oriented database where the data is stored as document collection. It is NoSQL database and allows storing data inside documents collection and saves from rigid tabulated data storage in SQL. It is becoming increasingly a choice for storing data in the organizations. MongoDB comes with various licenses like MongoDB Atlas, Community Server and Enterprise server. The Community server is free of cost and Enterprise server is used for commercial requirements. The MongoDB Atlas is managed on cloud. MongoDB provides support for Map-Reduce query, graph, text and geo searching. Atlas Search and Atlas Data Lake is the MongoDB Inc.'s full text search engine. MongoDB is used as a service in MongoDB Atlas, ApsaraDB for MongoDB, ScaleGrid MongoDB.

D. Oracle

Oracle was developed Larry Page in 1979. It is based on structured query language and is one of the most commonly used relational databases in enterprises. Oracle is feature advanced database. User can create OLAP and OLTP transactions in one instance of the database. It provides features for both structured and semi structured data. The blockchain tables support is available and very fast transactions are facilitated. Oracle database is not open source, and special database features are included in the paid version of the database. Oracle is not the best option if users want to save money. LinkedIn, Netflix, ebay, Intuit and many other companies are using Oracle database. Oracle database is preferred for Master database where single database is used for OLTP and OLAP. The Oracle database provides support for Blockchain tables, so it can preferred for using Blockchain table. Oracle database can be used where data warehousing is required. Oracle database is being used as a service in Amazon RDS for Oracle and Oracle Converged Database.

E. Microsoft SQL Server

It is the Microsoft's SQL based database offering, which is paid. It has received support from Corporate organizations and offers services at cheaper license than Oracle database. Microsoft has also released cloud based database to save users from on premise installation. It is Azure SQL and is built similar to the Microsoft SQL Server. It is platform as a service offering by Microsoft. SQL Server guarantees ACID transactions and is very closely coupled with other products of Microsoft. Azure Cloud can be used for taking the data off the user's premises, thus shifting to cloud. It may not be the best choice for converged or distributed SQL databases and where the data is semi structured. Microsoft SQL Server is used as a service in Azure SQL, SQL server on Google cloud, Amazon RDS for SQL server.

F. Cassandra

Cassandra was designed keeping in mind the handling of massive data. It is made on the distributed wide column store design. Cassandra is open core and distributed database with the Apache License 2.0. It can be used as a time series database and is capable of handling massive amount of data. It is AP implying Available and Partition tolerant and is used for online analytical processing (OLAP) as is performed in data warehousing. It is highly scalable and provides automatic sharding. Cassandra becomes fault tolerant due to decentralized nature. The decentralization, multi-datacenter and automatic replication makes Cassandra databases fault tolerant. It has its own query language called Cassandra Query Language (CQL) similar to the SQL query language. Cassandra can be used for Online analytical processing, time series databases, massive amount of data handling, high durability cases. It may not be the best option for use as the primary database where ACID transactions must be ensured, in cases which require frequent data deletion, to save on Cloud bill as JVM needs many nodes. Cassandra is offered as a service through DataStax Astra, Aiven for Apache Cassandra(Multi-cloud), Apache Cassandra managed by Amazon.

G. Redis

In order to improve scalability of the startup, a developer made significant improvements in the existing database system and created the distributed store for in-memory key-values. Redis is open core NoSQL database. It supports strings, lists, hashes, sorted sets with range queries, geospatial indexes, hyperlogs, streams and bitmaps.It is Consistent and Partition tolerant(CP). It offers advanced functionality for caching(Limited TTL keys, policy for LRU eviction). It is good to be used in real time use cases, for huge databases, session managements, to improve access performances and when distributed structure for in-memory is required.

©IJRASET: All Rights are Reserved

Redis may not be the best option to use when one store for server in-memory data structure is required, ACID transactions are necessary on primary storage, for small dataset single server caching is required or the data has lots of relations or Online analytical processing is required.
It is used as a service in ScaleGrid Redis for Managed Redis, Redis Enterprise Cloud with multi cloud support, Aiven for Redis with multi cloud support, Google Memory store, Azure Cache for Redis, Amazon ElastiCache.

H. Elasticsearch

Elasticsearch analytics engine supports all the datatypes including structured, unstructured, numerical, textual, and geospatial data. The Elasticsearch is free, distributed. It is fast and scalable with simple APIs. It is open core and distributed search engine which provides horizontal scaling. Elasticsearch search engine is the most scalable. It is Available and Partition tolerant and offers REST API. It supports structured and schemaless data(JSON), hence employed for monitoring and logging tasks. It is used with Kibana very often. It provides support for cross-cluster replication and automatic replication. Elasticsearch may not be the best option to use when dataset is small and data is structured, required full-text search features are basic, expensive features(e.g. advanced ML) are not necessary. Elasticsearch is used where moderate to advanced full-text search is performed, usage of semi structured data like JSON or Log Data are involved, dataset volume is huge and advanced ML features requiring near full text search in real-time. It is used as a service in Elastic Cloud for multi-cloud support, Amazon Elasticsearch Service for multicolor support, Aiven for Elasticsearch.

III. RANKING OF DATABASES

The DB-engine is an independent website which ranks the database on the basis of search engine popularity, job postings, social media mentions, and technical discussion volume. The database ranks are for July 2022 is provided. The database ranking reveals Oracle to be the most popular database worldwide. The ranking of databases on DB-Engine are: Oracle(1), MySQL(2), Microsoft SQL Server(3), PostgreSQL(4), MongoDB(5), Redis(6), Elasticsearch(7) and Cassandra(8). The high availability of the database makes it a popular choice for use in the organizations.

IV. CATEGORIES OF DATABASE MANAGEMENT SYSTEM

The database management systems belong to different categories. Each category of database management system is optimized to handle certain types of data and process the computation. One database management system may fall into multiple categories. The common categories for database management systems are reviewed with example databases for categories.

A. Document Store

Document database is a non-relational database. It queries the data as JSON-like documents. It facilitates querying to store and fetch data in the document format as used in the application code. Due to documents being are hierarchical, flexible and semi-structured, the documents can be modified according to the application requirements. It allows for the ad hoc queries analytics and flexible indexing on the documents collection. Oracle, MySQL, Microsoft SQL Server, PostgreSQL, MongoDB, Redis, Elasticsearch have document store.

B. Spatial Database Management System

A spatial database management system can query, store and manipulate the spatial data efficiently. Spatial data represent objects in a geometric space, for example points and polygons. The data types are dedicated while storing the data and spatial indices used for optimal access to the stored data. Using Spatial DBMS, the operations can be performed on the objects and objects can be manipulated e.g. calculating object property like polygon area. Geospatial data and spatio-temporal data are common variations of spatial data. Oracle, MySQL, Microsoft SQL Server, PostgreSQL, MongoDB, Redis, Elasticsearch provides spatial database management.

C. Graph Database Management System

Graph DBMS represent data as nodes and edges of graph. This facilitates data processing with advantages of graphical representation and simple calculation of certain graph properties like the nodes distance. Generally indices are not provided to the nodes, so direct node access cannot happen. Oracle, Microsoft SQL Server and Redis provide graph database management.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

D. Search Database Management System

Search engine DBMS are used for searching the content in high quality in cost effectively. These are optimized for querying keywords with specialized methods of complex search expressions, ranking of results, full text search etc.

Search engine DBMS can be preferred for application which need powerful aggregation and queries, full text search support complex search expressions, result rankings, or geospatial searches. Elasticsearch provides full text search.

E. Time Series Database Management System

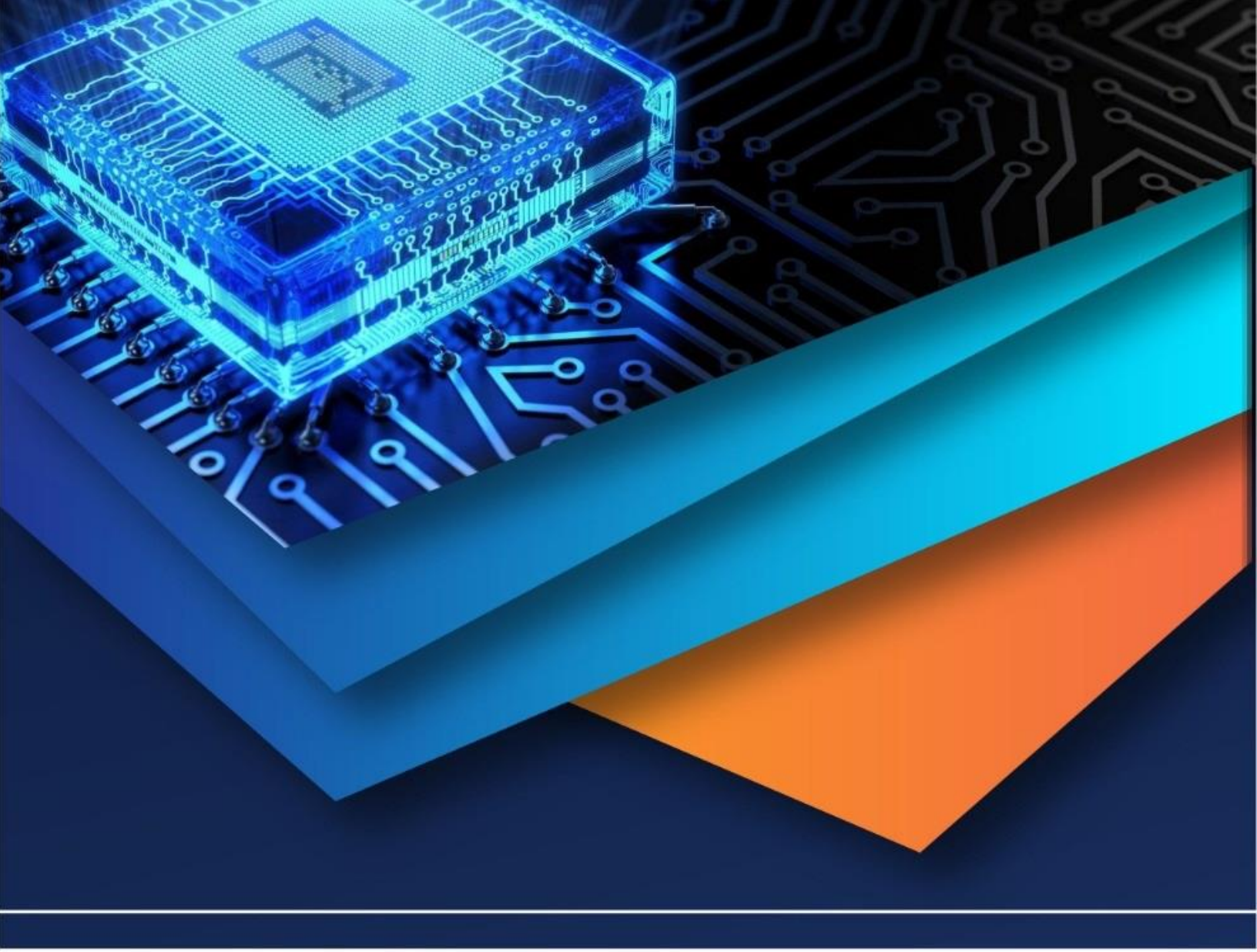
Time series database is optimized for time stamped and time series data. Such data is obtained by monitoring, tracking, downsampling the data over time and accumulated over time e.g. events, clicks, market trades, network data, application performance monitoring data etc. With time series data, summarizing queries are common. It requires traversing many entries on the database and generating the results. Optimization of this type of workload is difficult for regular databases. Time series databases are optimized for such querying. For time series data, the high precision data needs to be stored for aggregation and downsampling to generate meaningful results. After certain point of time, this data requires to be deleted. Time series database provide schemes to cheap eviction of large datasets and summarization at scale. MongoDB and Redis provide time series database management.

V. CONCLUSION

The paper reviews the top database management systems. The identified areas are usage areas, identification of scenarios where the database is not the best choice and the services to access the DBMS. The different database management systems for the databases specified are reviewed in the next section. These databases are supporting the data in different formats which require DBMS operations to be optimized for query and storage.

REFERENCES

[1] "DB-engines" https://db-engines.com/
[2] "What is a document database", https://aws.amazon.com/nosql/document/
[3] "Time series database explained", https://www.influxdata.com/time-series-database/
[4] "What is a Search-Engine Database? " https://redis.com/nosql/search-engine-databases/
[5] Md Kamaruzzaman , "Top 10 databases to use in 2021", www.medium.com
[6] "Most Popular Databases in the World", https://www.c-sharpcorner.com/article/what-is-the-most-popular-database-in-the-world/
[7] "Top 5 reasons to use an Oracle Database", www.dbta.com
[8] "Spatial Database Management Systems", https://db-engines.com/en/blog_post/88



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)