Phillip Yarbrough
10/6/2021
A590

Database Types

Throughout the years database technology has changed and evolved. The main purpose of database technology is to store information within a computer. This core purpose has adapted and changed due to modern life and the huge growth in data. For a user, client, and career seeking person, it is imperative to understand and learn each database type. The following paper will compare database types in terms of popularity, performance, speed and data latency, security, system compatibility, scalability, capability of handling big data, and column-based and row-based databases.

The first database that will be discussed is relational databases. These types of databases have been around since 1970 and the name comes from the way the data is stored in related tables. The data stored in these tables are made up of rows and columns. According to Matillion, "The relational database management system (RDBMS) is the program that allows you to create, update, and administer a relational database. Structured Query Language (SQL) is the most common language for reading, creating, updating, and deleting data." [1] According to db-engines.com, the following graph shows the popularity of relational databases in the last two years. [2]

380 systems in ranking, October 2021

Rank			DBMS	Database Model	Score		
Oct 2021	Sep 2021	Oct 2020			Oct 2021	Sep 2021	Oct 2020
1.	1.	1.	Oracle	Relational, Multi-model	1270.35	-1.19	-98.42
2.	2.	2.	MySQL	Relational, Multi-model	1219.77	+7.24	-36.61
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	970.61	-0.24	-72.51
4.	4.	4.	PostgreSQL	Relational, Multi-model	586.97	+9.47	+44.57
5.	5.	5.	MongoDB	Document, Multi-model	493.55	-2.95	+45.53
6.	6.	↑ 8.	Redis	Key-value, Multi-model	171.35	-0.59	+18.07
7.	7.	↓ 6.	IBM Db2	Relational, Multi-model	165.96	-0.60	+4.06
8.	8.	↓ 7.	Elasticsearch	Search engine, Multi-model	158.25	-1.98	+4.41
9.	9.	9.	SQLite	Relational	129.37	+0.72	+3.95
10.	10.	10.	Cassandra	Wide column	119.28	+0.29	+0.18

According to the table above, the top four DBMS systems are relational. Popularity is at an all time high. The performance of a relational database system is very high due to its integrated structure and data storage system. [3] The speed and latency of a relational database is extremely good due to its high strict structure demand. If a system administrator is scaling a database across multiple servers, speed and latency will decrease. The security of a relational database is very good due to its enormous growth, developer and community support, and its use in everyday business. Relational databases are a good choice for building and supporting complex software solutions, where any interaction has a range of consequences. One of the SQL fundamentals is ACID compliance (Atomicity, Consistency, Isolation, Durability). The ACID-compliance is a preferred option if you build, for instance, eCommerce or financial applications, where database integrity is critical. [3] When it comes to relational databases, they can handle enormous loads of big data. Some relational databases that users use is MySQL, MariaDB, and Oracle. The one downside of these types of databases are they are not as scalable and can become a challenge across multiple servers. The next section will discuss the database type NoSQL.

NoSQL is a broad category that describes any databases that do not use SQL as their primary data access language. The difference between NoSQL databases and relational database is that NoSQL databases do not have to conform to pre-defined schemas. The benefit for these types of databases is that they are great for organization of unstructured and semi-structured data. The popularity of NoSQL databases is rising quickly and are expanding more and more. According to the graph above, MongoDB is the fifth most popular DBMS and it is part of the NoSQL movement. The performance of a NoSQL database is very good with high speed and latency. One major disadvantage is that NoSQL databases lack security, and a user will usually have to pay a premium fee for any security. The system capabilities of a NoSQL database are endless and provide simple data access, storage, input, and retrieval. These systems cannot handle as much big data due to no support for joins. These databases can have data oversupply, which results in big memory waste and lower performance.

The third database type is cloud databases. According to Matillion, "A cloud database refers to any database that's designed to run in the cloud. Like other cloud-based applications, cloud databases offer flexibility and scalability, along with high availability. Cloud databases are also often low-maintenance since

many are offered via a SaaS model." [1] The popularity for cloud databases is on the rise and have grown from 1.6% in 2016 to 3.7% in 2019, based on the ranking scores of DBMS. The performance of cloud databases is very good due to users not having the worry of maintaining hardware, handling drive and network failures, dealing with internet outages, etc. You are leaving these problems in someone else's hands whose sole purpose is to deal with these problems. The speed and latency are fast, reliable, and consistently tested due to DBA's working on these issues outside of the clients work. The security behind a cloud database can be high, but there are still some significant gaps that cloud databases lack. Most cloud databases will have security, but not a robust up to date security system. According to fauna, "Look for cloud vendors who can offer attack detection, isolation for mission-critical data/systems, audit logs and other records to detect and trace the magnitude of data breaches, etc." [4] The system compatibility of a cloud database is fantastic because the burden heavily relies on an outside party. Cloud databases means a system analysis must deal with installing, maintaining, backing up, planning/testing recoveries, etc. This is all done by somebody else. Cloud databases handle big data very well but come at a high cost.

The fourth database type is the columnar database. According to Matillion, "Also referred to as column data stores, columnar databases store data in columns rather than rows. These types of databases are often used in data warehouses because they're great at handling analytical queries. When you're querying a columnar database, it essentially ignores all the data that doesn't apply to the query, because you can retrieve the information from only the columns you want."[1] These database types are considered the future of business intelligence and are often used in data warehouses. The popularity of these is continuing to grow very fast and the performance is higher due to storing data by column than rows. If data is kept closer together, minimizing seek time, systems can deliver that data faster. This will also improve latency. The security for most columnar databases is relatively good, but many costs additional fees for enhanced security measures. Most columnar databases are not suitable for every case. A user cannot insert a new record into a row with a single operation. Using a columnar database will take more computing and resources. One huge benefit for this database type is that it can handle huge amounts of data due to less pulling from physical disk drives by the having the head move less. Below is an example of a columnar database.

512,513,514
Seabiscuit,Bowler,Cuphead
Book,Apparel,Game
10.95,59.95,20.00
201712241200,201712241200,201712241201
goodreads.com,google.com,gamerassaultweekly.com

The fifth database type is the wide column database. This is where data is stored in column families, rather than rows and columns. Wide column databases can handle petabytes of data, thus making this database a huge contender for big data applications. The popularity for these databases is climbing rapidly. Some databases that use wide columns are Apache Cassandra and Scylla. The speed and latency are extremely high due to each data element being referenced by the row key but querying for a value is optimized like querying an index in a RDBMS, rather than a slow table scan. The security for wide column databases is not as good as traditional large-scale databases. Additional security features will cost a fee. The system capabilities are quite good due to the speed of querying, scalability, and a flexible data model.

The sixth database type is object-oriented databases. According to Matillion, "An object-oriented database is based on object-oriented programming, so data and all its attributes, are tied together as an object. Object-oriented databases are managed by object-oriented database management systems (OODBMS). These databases work well with object-oriented programming languages, such as C++ and Java. Like relational databases, object-oriented databases conform to ACID standards." [1] The popularity of object-oriented databases can be summed up from this graph below:

21 systems in ranking, October 2021

Rank			DBMS	Database Model	Score		
Oct 2021	Sep 2021	Oct 2020			Oct 2021	Sep 2021	Oct 2020
1.	1.	1.	InterSystems Caché	Multi-model	3.08	-0.14	-0.63
2.	2.	↑ 4.	InterSystems IRIS	Multi-model	1.72	-0.02	+0.04
3.	3.	↑ 5.	Db4o	Object oriented	1.67	+0.00	+0.07
4.	4.	↓ 3.	ObjectStore	Object oriented	1.47	-0.02	-0.26
5.	5.	↓ 2.	Actian NoSQL Database	Object oriented	1.32	-0.03	-1.03

The performance of object-oriented databases is high because reading and mapping an object database data to the objects is direct without any API or OR tool. Speed and latency are very good features with object databases and there are used frequently for real-time systems, architectural and engineering for 3D modeling, and telecommunications. Object oriented databases have a well-developed built-in security model. The system compatibilities for object-oriented databases are a great feature since object databases can be used in multiple languages, such as C++, Java, and Smalltalk. Handling big data is no issue for object-oriented databases.

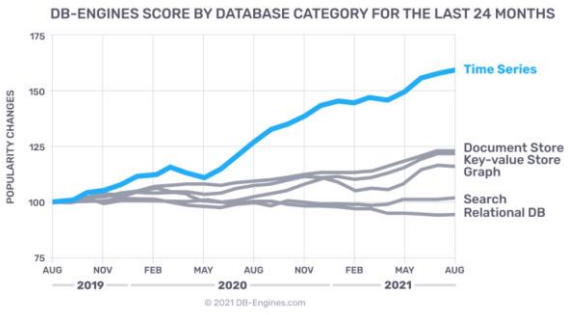
The seventh database type is the key-value database. According to Matillion, "One of the simplest types of NoSQL databases, key-value databases save data as a group of key-value pairs made up of two data items each. They're also sometimes referred to as a key-value store. Key-value databases are highly scalable and can handle high volumes of traffic, making them ideal for processes such as session management for web applications, user sessions for massive multi-player online games, and online shopping carts." [1] These databases are high in popularity due to consumer growth at an all time high. The performance, speed, and latency are extremely well and great characteristics for a high-volume traffic type of database. Key-value databases can handle huge amounts of data at ease and are currently running some of the largest consumer websites throughout the world, such as Amazon.

The eighth database type is hierarchical databases. These types of databases use a parent-child model to store data. To understand this, think of a family tree and each data structure branches off into more data structures. This was originally developed by IBM in the early 1960s and are commonly used to support high-performance and high availability applications. The popularity in hierarchical databases is high, but not growing rapidly. Speed and latency are very high with hierarchical databases and support high-performance and high availability applications. The security is good with hierarchical databases, but it can get hard for a system analysis to move data around since it must start at the node. This can case security gaps within the infrastructure. The system capability is slim since hierarchical structures, unlike, relational databases, do not describe many-to-one relationships or many-to-many relationships since child records can only have a single parent. These database types can handle large amounts of data, but it usually is not wise to choose this over another database type. There is a big lack of flexibility.

The last database type is the time series. According to Matillion, "A time series database is a database optimized for time-stamped, or time series, data. Examples of this type of data include network data, sensor data, and application performance monitoring data. All of those Internet of Things sensors that are getting attached to everything put out a constant stream of time series data."[1] These types of databases are a fast-growing segment and are gaining popularity. Below are the top ten time-series databases and a graph on the dramatic rise of time-series databases in the last year:

RANK	DBMS	SCORE		
AUG 2021		AUG 2021	24 MOS ▲	12 MOS ▲
1	**InfluxDB**	**29.56**	**+10.91**	**+6.22**
2	Kdb+	7.99	+2.49	+0.56
3	Prometheus	6.20	+2.75	+0.51
4	Graphite	4.86	+1.54	+0.55
5	TimescaleDB	3.41	+2.01	+0.68
6	ApacheDruid	3.00	+1.27	+0.71
7	RRDtool	2.43	-0.14	-0.62
8	OpenTSDB	1.90	+0.04	-0.40
9	FaunaDB	1.53	+1.11	-0.33
10	GridDB	1.37	+0.89	+0.62

Source: DB-Engines — 23 Systems in Ranking, August 2021



The speed and latency of a time series database is extremely fast and time stamps can be measure in second, millisecond, microsecond, or nanosecond precision. The security is very good, and most time series databases have a built-in security model. The system capabilities are limited to time series, but if a user is solely using these types of databases it is very efficient. The best part of time series databases is that they can handle a tremendous amount of data with ease. It is a very efficient system.

There are countless database types, and they are growing rapidly every day. Numerous applications are being developed daily and in the next couple of years a new database design will be discovered. It is imperative for a system analysis to incorporate a database and a specific database application that will be the most efficient, secure, and productive approach towards their objective.

REFERENCES

[1] Author Julie Polito. (2021, September 22). *The types of databases (with examples)*. Matillion. Retrieved October 9, 2021, from https://www.matillion.com/resources/blog/the-types-of-databases-with-examples.

[2] *Engines ranking*. DB. (n.d.). Retrieved October 9, 2021, from https://db-engines.com/en/ranking.

[3] Editor. (2019, October 15). *Comparing database management systems: Mysql, PostgreSQL, MSSQL server, mongodb, Elasticsearch and others*. AltexSoft. Retrieved October 9, 2021, from https://www.altexsoft.com/blog/business/comparing-database-management-systems-mysql-postgresql-mssql-server-mongodb-elasticsearch-and-others/.

[4] *What is a cloud database?* Fauna. (n.d.). Retrieved October 9, 2021, from https://fauna.com/blog/cloud-databases-types-advantages-and-considerations.