The objective of this paper is a comparison of databases based on popularity, performance, speed and latency, security, system compatibility, and capability of handling big data; this paper will also dive into column-based and row-based database comparison. I decided to focus on two databases I’ve encountered in my career in data first as a data analyst and second as IT quality analyst. I worked as a data analyst for a growing nonprofit that used Microsoft SQL Server for the database housing client information. I am currently working for a large size for-profit company that has recently completed the transition from Microsoft SQL Server (SQLS) to using Snowflake. Both systems have their pros and cons that can benefit a company of any size; however, it comes down to the priorities of data storage and internal change management.

The popularity of SQLS is undeniable due to the impact the Microsoft Corporation has on technology from operating systems to office applications. “Why has MS SQL Server become popular? First of all, it was developed by an IT giant that dominated the market throughout the 90s. You may not remember, but everyone wanted Windows.” (Romanowski, 2021) The ability to acquire a Microsoft product that users could rely on was easy and trustworthy for companies and individuals. “…in the 90s, Microsoft was synonymous with modernity and stability.” (Romanowski, 2021) Many developers and analyst learned T-SQL the language used for SQLS based on the wide availability of the application. “…the fact that MS SQL servers are so common means that a lot of database administrators and developers have experience with them.” (Romanowski, 2021) This was true in my case as well, the small nonprofit only had Microsoft Corporation products because there is typically a smaller learning curve. I was able to take a course on T-SQL taught by Microsoft online which ensured my job security. This brings to me to the transition to using Snowflake at my current workplace. “Snowflake is a cloud-based data warehouse solution that is rendered as a Software-as-a-Service (SaaS) with full assistance for ANSI SQL.” (Times Business News , 2020) Snowflake is rapidly growing in popularity amongst companies transitioning into big data. “Snowflake can be adapted to fit various use cases- data lakes with raw and fresh, new data, data warehouse/data marts, and ODS with staged data with attractive, modeled, and presentable data.” (Times Business News , 2020) It user-friendly from a complete novice to data engineer to build or design tables. “It has out-of-the-box characteristics or features such as detachments of storage and estimates, on-the-fly adaptable compute, data cloning, data sharing, and third-party tools or devices assistance.” (Times Business News , 2020) As analyst I found that SQLS set the foundation for my quick grasping of Snowflake, however in tech the new technology typically has a higher rating in popularity. Please see graphs.

Chart, line chart

Description automatically generated (DD Engines, 2021)

Chart, line chart

Description automatically generated

(DB-Engines, 2021)

Now that we are in the Internet of Things (IoT) Microsoft has to compete with several cloud based systems and had to focus on performance. The benefit of having a stabilized product with auto updates Microsoft can make changes that impact millions of users immediately. “With built-in transparent data compression and encryption features, SQL server offers enhanced performance.” (Tek-Tools, 2020) Snowflake on the hand focuses on the performance of the system and how to make things simpler for the user. “A driving factor of Snowflake’s impressively fast performance it its automatic query optimization feature…Powered by a dynamic query optimization engine in its cloud services layer, it manages query planning and optimization based on data profiles.” (Vittal, 2020) My experience with SQLS the speed has been based on size of the database as well as if it was local or remote access. “…all the built-in capabilities that power the speed of SQL Server…From a SQLOS scheduling engine that minimizes OS context switches to read-ahead scanning to automatic scaling…” (Microsoft, 2017) Microsoft has heard complaints from users about it speed and has been making several upgrades to address it. “With Snowflake, everything is cloud-based, meaning you can deploy it on a minute scale that can later get scaled up or down based on the company’s needs.” (Zuar, 2012) Snowflake has significantly reduced the time it requires me to query databases that were formally housed in SQLS. “Snowflake databases are as user-friendly as they get and allow users to organize data in any manner they wish. Snowflake is designed to be a highly responsive platform that performs optimally on its own-there’s no need for a specialist to keep an eye on it constantly.” The importance of speed can be high priority of several companies, however the longevity that SQLS offers can supersede that. “SQL Server provides efficient permission management tools with access controls designed to help users secure sensitive business information.”

This brings me to security of the data being stored by the company and the steps that data warehouses take to ensure the contents. Most companies that use SQLS store locally instead of cloud-based even with the option to use SQLS cloud feature. “Microsoft SQL Server has recorded only 19 security vulnerabilities in 2010-2016, while other major databases have recorded 200 or more in that time period.” (Yehuda, 2018) The Microsoft Corporation has focused on auditing and protecting their code. Most systems used in workplaces are operating on Windows OS, therefore it uses it own authentication for SQLS. Whereas Snowflake authentication is based on the needs of the company and/or the federal guidelines they operate within. “In the Snowflake environment, the VPS service takes it a step further with a dedicated set of cloud resources…With Snowflake’s VPS, all data is processed privately within a dedicated Virtual Private Cloud instance for each individual customer.” (Vittal, 2020) Windows authentication can be right amount of security for the average company, however adding further enhancements is not uncommon.

Microsoft Corporation within the last decade has made significant efforts to increase their compatibility outside of their own products including adding in Linux capabilities. “In September 2017… the addition of in-database machine learning in SQL Server with support for both R and Python.” (Yehuda, 2018) If users decide that SQL language is the one they want to stick with, however are open to change in other ways Microsoft will be available. “Because compatibility is optimized when a DBMS and OS are made by the same company, its just easier to stick with Microsoft products all around.” (Yehuda, 2018) Snowflake being the newer system of the two didn’t have to make any major changes to become more accessible to other languages outside of SQL. “Snowflake data warehouse supports most DDL and DML commands of SQL. It also supports advanced DML, transactions, lateral views, stored procedures…data warehouse supports an extensive set of client connectors and drivers such as Python connector, Spark connector, Node.js driver, .NET driver, etc.” (Times Business News , 2020) Snowflake ensures that the user has all of the potential options needed to perform their duty at inception, whereas SQLS gives room for growth. “In case a database needs to be switched, reconnecting the new database’s program to the data access layer is made much easier when doing everything with .NET and SQL Server…using Microsoft languages is more straightforward.” (Yehuda, 2018)

The major focus of data storage for companies of various sizes with potential for growth is rather their database storage system can handle big data. One of the main benefits offered to SQLS users is that “…using Microsoft SQL you can host locally or remotely through a well-documented process.” (Yehuda, 2018) It has the ability to take a small operation to a larger one based on the growth of their data. “…SQL servers can be clustered to reduce single points of failure, increase the efficiency of hardware resources, and facilitate maintenance and patching.” (Yehuda, 2018) The interconnect ability of Microsoft products makes it simple to enhance without worrying about the impact it would have on the overall system. Snowflake is a cloud-based system therefore it is designed to handle big data and operate efficiently. “The snowflake architecture consists of three component layers: database storage: The innate elasticity of cloud storage means, data can be independently stored, encrypt and compressed. Snowflake also provides additional protection of geo-redundancy to the stored data. Compute clusters: Also called “Virtual Warehouses”, these can simultaneously load data and run queries. Cloud services: This layer ties together the different components of Snowflake and performs the user requested functions. The cloud service layer runs on Snowflake compute instances…three layers are designed to be independently scalable.” (Vittal, 2020) The scalability of Snowflake at varying levels can change how data survives within a company for the long run. “While scalability is expected to a degree from any reliable cloud data warehouse platform, Snowflake offers a uniquely flexible multi-cluster, shared data architecture.” (Vittal, 2020)

With that wrapping up the discussion about SQLS and Snowflake, we will move into the discussion about row-based and column-based database structures. Row-based databases are typically known as traditional storage databases used with the focus on storage. “Row oriented databases are databases that organize data by record, keeping all of the data associated with a record next to each other in memory.” (The Data School, 2021) Column- based databases are not as focused on storage but more what can be done with the data stored. “The column store is awesome for performing aggregation over large volumes of data. Or when you have queries that only need a few fields from a wide table.” (Gobla, 2016) Querying has been done with traditional based querying; however, it can be time-consuming based on the row structure of how the data is written. “At a basic level, row stores are great for transaction processing. Column stores are great for highly analytical query models.” (Gobla, 2016) The processes of writing column-based is time consuming, therefore some users find themselves integrating both options. “This column-oriented database is being used by most major providers of cloud data warehouse.” (The Data School, 2021) Most notable are Snowflake, Redshift, and BigQuery, whereas common row-based databases are Postgres and MySQL. “…it is more of the norm that companies support multiple database platforms for multiple uses…your needs might change over time.” (Gobla, 2016)

Deciding between the stable standard that is SQLS or Snowflake or between row-based or column-based is done on a case-by-case basis. The needs of a growing company versus an established one its typically the deciding factor. One of the benefits of IoT is things are typically well documented for analysts and developers to troubleshot the best practice regardless of the selection.

# References

DB-Engines. (2021). *DB-Engines Ranking - Trend of Microsoft SQL Server Popularity*. Retrieved from DB-Engines: https://db-engines.com/en/ranking\_trend/system/Microsoft+SQL+Server

DD Engines. (2021). *DB-Engines Ranking- Trend of Snowflake Popularity* . Retrieved from DB-Engines: https://db-engines.com/en/ranking\_trend/system/Snowflake

Gobla, R. (2016, 12 26). *Row Store vs. Column Store Databases*. Retrieved from DZone: https://dzone.com/articles/row-store-and-column-store-databases

Microsoft. (2017, 9 27). *SQL Server 2017: Fast, faster, and the fastest database everywhere you need it*. Retrieved from Microsoft SQL Server Blog: https://cloudblogs.microsoft.com/sqlserver/2017/09/27/sql-server-2017-fast-faster-and-the-fastest-database-everywhere-you-need-it/

Romanowski, J. (2021, 3 25). *Is MS SQL Server Still Worth Learning in 2021*. Retrieved from learnsql: https://learnsql.com/blog/learn-ms-sql-server/

Tek-Tools. (2020, 5 15). *All About SQL Server: Advantages, Best Practices, and Tools*. Retrieved from TEK-TOOLS: https://www.tek-tools.com/database/sql-server-best-practices-and-tools

The Data School. (2021, 8 9). *Row vs Column Oriented Databases*. Retrieved from The Data School: https://dataschool.com/data-modeling-101/row-vs-column-oriented-databases/

Times Business News . (2020, December 18). *Why is Snowflake Gaining Popularity So Quickly* . Retrieved from Time Business News : https://timebusinessnews.com/snowflake-gaining-popularity-quickly/

Vittal, S. (2020, May 20). *Top 10 Reasons Why You Should Consider Snowflake For Data Warehousing*. Retrieved from Archetype : https://blog.archetypeconsulting.com/top-10-why-snowflake

Yehuda, Y. (2018, 5 03). *6 Reasons Why Microsoft SQL Is Alive and Well*. Retrieved from Dzone: https://dzone.com/articles/6-reasons-why-microsoft-sql-is-alive-and-well

Zuar. (2012, 8 11). *Pros and Cons of Using Snowflake Cloud Data Warehouse*. Retrieved from Zuar: https://www.zuar.com/blog/pros-and-cons-of-using-snowflake-cloud-data-warehouse/