NoSQL Database Vs Relational Database

Abstract— **NoSQL databases are used more commonly compared to the relational database as NoSQL can handle a large set of data. This paper will discuss the comparison of NoSQL and relational database.**

Key words: *(NoSQL Database, Relational Database, MongoDB, SQL Server)*

I.   Introduction

The NoSQL database consists of four categories which include key-value store, document-oriented, column-oriented and graph database. Key-value stores a collection of key and value pairs. An example of a key-value store database is Memcached and Redis. Document oriented uses key and document, not all the documents need to be valid against the schema. An example of a document-oriented database is MongoDB and CouchDB. In the column-oriented database, the data is stored as columns, it is similar to the relational database were relational database stores data in rows. An example of a column-oriented database is Cassandra and Bigtable. In the graph database, the data is connected to nodes. An example of a graph database is GraphDB and OrientDB. The relational databases are based on the relational model. An example of relational databases includes MySQL, SQL Server and Postgre. (Kunda and Phiri, 2017).

In this research article, the sections will be divided into three parts. Section one discusses related work and review research articles and journals related to a similar topic and critique the research articles and compare them. In section two, the author will discuss the main features of both NoSQL and relational databases and compare them against each other. The last section will be the conclusion, where the author provides a summary of the report and any future research related.

II.   Literature Review

An article by Lourenço, Cabral, Bernardino and Vieira, (2015) experimented on performance and space using NoSQL database and comparing it with relational database. The author has used three NoSQL databases these include, Cassandra, MongoDB, Couchbase server and one relational database Ms SQL Server to perform the comparison of performance and space when managing a large set of data for the write request.

The result from the experiment for the performance indicated that Cassandra was best performed out of the NoSQL databases. MongoDB had very low throughput and higher workload and could not complete the task compared to the other NoSQL databases whereas, for Cassandra the throughput was steady. Out of all three NoSQL databases the MongoDB performance was poor when compared to other databases. When compared with the relational database (Ms SQL server) the performance was adequate though there was degradation. Moreover, when space was experimented the relational database (Ms SQL server) performed better compared to NoSQL databases. However, NoSQL Cassandra database is better performance-wise when the write-heavy

dataset is included. (Lourenço, Cabral, Bernardino and Vieira, 2015). In the journal the author only has used one relational database to carry out the experiment. The author could have compared a few more relational databases against the NoSQL databases, for instance, Oracle database, MySQL or IBM DB2 instead of using just one relational database to compare it with three NoSQL databases. In the research journal, the author provides a depth understanding of all the NoSQL database though further explanation could have been mentioned in the journal about the relational database Ms SQL Server. This illustrates that the author has provided infinitesimal information on a relational database and Ms SQL server.

A similar test has been carried out by Parker, Poe and Vrbsky, (2013) to compare the performances of the two databases. However, in this article, the author only uses one database for NoSQL and one database for relational whereas, Lourenço, Cabral, Bernardino and Vieira, (2015) used multiple NoSQL databases to test the performance. Parker, Poe and Vrbsky, (2013) uses MongoDB and Ms SQL server to compare the performance based on runtime for both NoSQL and relational databases. Both the MongoDB and SQL Server will be tested on the performance of the DML operations such as the insert, update and select. The author indicates that the NoSQL should be considered when working with a large set of data and for a non-structured database whereas, the relational database should be considered when working with a limited amount of data which is structured. In the article, the authors state, NoSQL is better for fast and simple operations when working with a large dataset.

The author also indicates that MongoDB and other NoSQL databases show deficient in some features compared to the relational database. The aggregation function in MongoDB is also deficient as it does not provide the user with simple aggregations like the relational database, but MongoDB uses MapReduce function to use these aggregation functions. After experimenting, the authors have indicated that Mongodb had better runtime for DML statements such as insert, update and simple select queries compared to SQL. However, SQL also showed better performance

after the experiment with the update statement and querying using the aggregation function. From this research article, the author suggests that MongoDB is better compared to SQL as the performance of the runtime for DML such as insert, update and select is better even though there was a drawback to MongoDB as additional step needed to take to perform the aggregation function. Moreover, MongoDB also works better when there is a schema with restrictions and has modest data. (Parker, Poe and Vrbsky, 2013). The researchers provide a clear understanding of both NoSQL and relational databases to the readers as the author goes in-depth about the two main databases MongoDB and SQL. At the end of the article, the author provides future work for both MongoDB and SQL such as doing complex schema type or running it on a distributed database. The author has also provided few statistical graphs for DML statement experiment which has been tested for both MongoDB and SQL which provides the readers with a clear visual representation of the results.

Nayak, Poriya and Poojary, (2013) addresses different types of NoSQL databases such as the key-value store, column-oriented, document store, graph database and object-oriented. Furthermore, the author discusses and provides examples of where these database services are implemented. For example, key-value stores are used by database service providers such as Amazon Dynamodb and RIAK, column-oriented are used by Big table and Cassandra, document store is used by MongoDB and CouchDB, a graph database is used Neo4j and object-oriented is used by DB4o.

In the paper, Nayak, Poriya and Poojary, (2013) indicate that more users use a relational database as the users are more familiar with SQL language compared NoSQL as most of the NoSQL databases used different query languages. This indicates that the relational database is better language-wise as it only uses one query language which is SQL whereas, the NoSQL have multiple querying language for different databases which makes it difficult for the user to exchange from NoSQL database to another. The author could have addressed or researched about relational database

and provide examples of it being implemented by a database service provider like NoSQL and compare both database services. As in the paper, the author predominantly discusses NoSQL database side and the relational database is mentioned to a smaller extent in the paper. The paper mostly discusses NoSQL compared to the relational database. The author could have carried out further research on a relational database and also compare the advantages and disadvantages of both relational and NoSQL database.

Furthermore, most of the research articles and journals were based on experimenting with the performance of NoSQL and relational database and comparing them against each database to see which would perform better.

III.   Comparison of NoSQL and Relational Databases

The main feature of both NoSQL and Relational database will be compared below.

Transaction Reliability: Relational database provides a high transaction reliability as it supports ACID (Atomicity, Consistency, Isolation and Durability) consistency model whereas, NoSQL is supported by BASE (Basically Available: Guaranteed Availability) consistency model and can be range from BASE to ACID. (Mohamed, Altrafi and Ismail, 2014).

Data Model: For relational database, the data is represented in n-ary relation and the data inside the database is represented as tuples (rows) which is clustered into relations (tables). The data is stored in rows and column in table. The data model for relational database is well organised and detailed. NoSQL database uses various types of data modelling techniques such as key value store, graph stores and document store. NoSQL database uses more than one data model to represent the data. Moreover, NoSQL does not use schema or table for storage structure and can handle unstructured data. (Mohamed, Altrafi and Ismail, 2014).

Scalability: Relational database depends on vertical scalability meaning the user can increase the hardware component on the server such as the RAM and CPU to reduce the load which can be very expensive and difficult to handle the server. Whereas, NoSQL depends on horizontal scalability meaning the user can add additional servers. (Mohamed, Altrafi and Ismail, 2014).

Cloud: The relational database is not suitable for a cloud platform as it does not fully support the content of search and it is difficult to scale whereas, NoSQL database is suitable for a cloud platform as it work well with various data types such as unstructured, semi-structured and structured data. It improves the availability, scalability, flexibility and performance. (Mohamed, Altrafi and Ismail, 2014).

Big Data: NoSQL database can handle large datasets compared to relational database because NoSQL can improve the performance of storing and retrieving the data. In relational database handling big data can be a problem and in order to handle big data, the data needed to scalable and divided into multiple servers. The performance and complexity can be a problem when joining the data. (Mohamed, Altrafi and Ismail, 2014).

Data Warehouse: In relational database, the data is collected from various sources and the data gets expanded over a period as more data are being collected. This causes a big data issue and causes performance to slow down when performing OLAP, data mining or statistical analysis. NoSQL database are not designed for data warehouse applications as it is more focused on high performance, scalability, availability and for storing large data. (Mohamed, Altrafi and Ismail, 2014).

Complexity: In relational database the complexity can be an issue cause the data need to be converted into table first and if data does not fit into the table the database structure can complex. NoSQL can store various type of data structures such as structured, unstructured and semi-structured data. (Mohamed, Altrafi and Ismail, 2014).

Recovery from crash: To recovery from crash in relational database the recovery can be done through recovery manager and using log file and ARIES algorithm provides a transaction atomicity and durability. Whereas, in NoSQL database recovery of data can be retrieved from replications of backup. NoSQL database also uses other tools such as journal file in MongoDB to carry out crash recovery. (Mohamed, Altrafi and Ismail, 2014).

Security: Relational database provides a good security measures whereas, NoSQL database is more focused on storing big data and the performance so the security in NoSQL database provides a low security level however, NoSQL is still working on improving these security issues. When comparing security features such as authentication, data integrity, confidentiality, auditing and client communication with relational database and NoSQL database, relational database provide more of these security features. For instance, relational database provides authentication and other mechanism tools whereas, NoSQL uses external method to perform authentication. For confidentiality in relational database, it uses encryption method and the data is encrypted. Whereas, NoSQL does not provide confidentiality. The relational database does support data integrity as it uses ACID properties, so it provides a reliable database transaction. Whereas, in NoSQL it does not support data integrity. In relational database advanced level of auditing is used for example, Oracle Fine Grained allows to audit under specific circumstance, in contrast to NoSQL, most database does not support auditing. The client communication is relational database is secure as it uses encryptions and SSL protocol however, in NoSQL, it does not provide client communication. (Mohamed, Altrafi and Ismail, 2014).

IV. Conclusions

To conclude this report, most research article were using mostly MS SQL Server for relational database to run experiments and not much experiments were

carried out other relational tools such as MySQL, Oracle Database, IBM DB2 or Postgre. Using additional database tool would have helped the reader understand which tool is best suitable for specific situation. Furthermore, when comparing NoSQL databases with relational database the research articles could have provided additional information each database.

After reviewing the research articles security was a big problem NoSQL database was facing as the NoSQL database provides low security whereas, the relational database is good when implementing security as it provides high security features. Furthermore, from the research NoSQL databases can be used when handling large data sets. Additionally, some features did not use NoSQL or relational database, for example, for data warehouse feature NoSQL database was not required and relational database cloud was not suitable.

V. References

[1] Kunda, D. and Phiri, H., 2017. A Comparative Study of NoSQL and Relational Database. *ZAMBIA INFORMATION COMMUNICATION TECHNOLOGY (ICT) JOURNAL*, 1(1), pp.1-2.

[2] Lourenço, J., Cabral, B., Bernardino, J. and Vieira, M., 2015. Comparing NoSQL Databases with a Relational Database: Performance and Space. *Services Transactions on Big Data*, 2(1), pp.1-14.

[3] Mohamed, M., Altrafi, O. and Ismail, M., 2014. Relational Vs. NoSQL databases: A survey. *International Journal of Computer and Information Technology*, [online] 03(03), pp.599-601. Available at: <https://www.researchgate.net/publication/263272704_Relational_Vs_NoSQL_databases_A_survey> [Accessed 9 March 2020].

[4] Nayak, A., Poriya, A. and Poojary, D., 2013. Type of NOSQL Databases and its Comparison with Relational Databases. *International Journal of Applied Information Systems (IJAIS)*, 5(4), pp.16-19.

[5] Parker, Z., Poe, S. and Vrbsky, S., 2013. Comparing NoSQL MongoDB to an SQL DB. *Proceedings of the 51st ACM Southeast Conference on - ACMSE '13*,