Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
January 2024 5(1), 2711–2716

A COMPARATIVE ANALYSIS OF SQL AND NOSQL DATABASES FOR SCALABLE DATA APPLICATIONS

Dr. Rasika Khandal [1]

[1] Assistant Professor, B.C.A. Department, Sudha Sureshbhai Maniar College of Computer and Management, RTM Nagpur University, Nagpur, India

ABSTRACT

Today, where big data and cloud computing are the new normal, the kind of database used can affect the performance as well as scalability of applications significantly. In this paper, we are going to compare between SQL (Structured Query Language) and NoSQL (Not Only SQL) databases on parameters such as scalability, performance, data structure, flexibility and an appropriate use of these databases on various types of applications. Research is based on a combination of descriptive statistics and test of hypothesis in judging what users prefer and would like to deal with in the real world using both SQL and NoSQL. The findings give an idea about the most appropriate database model that fits in current data applications.

Keywords: SQL, NOSQL, Scalability, Relational Databases, Document Databases, Performance, Big Data, Data Modeling, Distributed Systems, Database Comparison

Corresponding Author

Dr. Rasika Khandal,
raut.rashu@gmail.com

DOI
10.29121/shodhkosh.v5.i1.2024.5914

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

1. INTRODUCTION

Our contemporary digital world is rapidly expanding by the number of individual users, environments they are using and applications that are maintained. Whether it is social media interaction and e-commerce transfer, customer data and IoT sensors and real-time analysis, the requirement to store, manage and analyze large quantities of various data became more important than ever before. The high rate of increase in the volume, velocity and variety of data has prompted the traditional database systems and other new forms of data management models have been developing.

SQL (Structured Query Language) databases or rather relational databases have been the foundation of data storage and management over the last couple of decades. They are characterized by precise data structures, high consistency, capability to perform complex queries based on SQL and the following ACID (Atomicity, Consistency, Isolation, Durability) principles. These characteristics render the SQL databases efficient within the scenario where the integrity

of the data and complex relationships are critical, including the banking systems, inventory management, and enterprise resource planning (ERP).

But this is when there were rather structured data like text, numbers and forms of data that companies began to work with when there came about forms of unstructured data, semi-structured data like images and movies, social media and logs as generated by sensors. It was then that relational databases began to run into problems of flexibility and scalability. To help deal with this, NoSQL (Not only SQL) databases were presented. Similarly to SQL databases, NoSQL systems offer stretchable schema models, lower bucket development, and enhanced achievement of specific types of requests. They tend to be well-suited to use-cases in big data, in real-time web applications, and in cloud-native services, since they often support a limited set of data models, e.g., key-value pairs, documents, wide-columns, graph.

The key trade-off between SQL and NoSQL are their strategy towards consistency, availability and partition tolerance which are famously known as CAP theorem. The SQL databases are less restrictive, allowing the databases to have stronger connections and are consistent, although NoSQL databases other times tend to give more priority to availability, as well as partition tolerance allowing to have a BASE (Basically Available, Soft state, finally consistent) model.

As both types of databases possess their own advantages and shortfalls, selecting a database to use in an application can be a very challenging task to both a developer and an organization. This choice is even more an important one when looked down in terms of scalability particularly because there are applications that are expected to scale very fast in number users, as well as the size of data to be stored on them.

This research proposal will offer the comparison between SQL and NoSQL databases and will be done under scalability and performance. It examines which database model fares best in various workloads, examines the actual preference in real users and assesses any statistical evidence to back up any argument. Through this the research is practical in ensuring that developers, architectures and those who make decision can be assisted in choosing the best database technology to support their scalable data applications.

2. LITERATURE REVIEW

The initial proposal of intelligent database system appeared by Raman and Chandra (2003) who outlined the intelligent aspects of database improvement through AI-based methods of intelligent decision-making. Verma and Sharma (2005) went further to demonstrate that artificial intelligence could enhance query optimization and, in the process, accelerate quick and efficient data retrieval. In a compare and contrast of SQL and NoSQL data stores, Cattell (2011) pointed out the scalability of data stores and depicted that NoSQL data stores are more versatile in relation to large-scale systems.

Han et al. (2011) presented NoSQL database survey and simply defined that there are several types of NoSQL including document-based databases, key-value databases, and graph databases, and they are applicable to work with unstructured data. As noted by Moniruzzaman and Hossain (2013), NoSQL is gaining importance in big data analytics and differentiated its characteristics and capabilities with those of the classical relational databases. Li and Manoharan (2013) analyzed the SQL and NoSQL databases performance and decided that NoSQL outperforms SQL with special big data tasks.

Sadalage and Fowler (2012) provided a primer of NoSQL technology, presenting the idea of polyglot persistence, (using a variety of databases to handle various requirements within the same application). Hecht and Jablonski (2011) performed an evaluation of the NoSQL databases employing the use cases, which allows users to choose the appropriate one in accordance with particular specifications. Atzeni and Torlone (2014) touched upon the fact that both SQL and NoSQL are changing to be able to handle data-based on the Web, so a hybrid model is usually advantageous.

A comparative study done by Singh and Prakash (2015) therefore highlighted the realization that to select the appropriate model, an application should be chosen in accordance to its requirements. Bajpai and Agarwal (2016) compared MongoDB and MySQL in big data workloads, and the result showed that MongoDB would fit better on unstructured data. Another comparison provided by Batra and Tyagi (2017) was that, although NoSQL is suited to flexibility and fast work, SQL is intended to preserve integrity of the structured data.

In the analysis of Sharma and Gupta (2019), the SQL and NoSQL performances were compared in the context of real-time applications, and it was determined that the latter generally performed better in changing situations. The comparison of the performance carried out by Patel and Mehta (2020) also proved that the use of NoSQL databases was

more efficient when it comes to working with large and diverse data sets. Kumar and Singh (2021) have considered the ACID (SQL) and BASE (NoSQL) models, making the developers familiar with sometimes a lack of consistency can be reasonable to offer adequate scaling.

Roy and Thomas (2021) made the comparison of the most popular NoSQL databases, MongoDB and Cassandra, and pointed to the advantages of these databases in distributed systems. Thomas, Roy and Meena (2022) have coined the functionality of polyglot persistence in the enterprise systems in which both SQL and NoSQL are simultaneously utilized. Adams and Roy (2023) compared SQL and NoSQL databases in cloud spaces and determined that NoSQL is better suited to the cloud and can scale.

Bansal and Khurana (2023) discussed the survey of the role of NoSQL in real time-analytics and discovered its further significance as an application, IoT and social media. Finally, Singh and Patel (2023) emphasized on existing trends, i.e., the emergence of multi-model databases with the ability to easily mix SQL and NoSQL capabilities seeking versatility and efficiency.

3. OBJECTIVES OF THE STUDY

1) To benchmark between the SQL and NoSQL databases on their scalability and performance on large-scale data applications.
2) To compare the efficacy of SQL and NoSQL databases in relation to various types of data and their purpose.
3) In order to understand the level of user preference and satisfaction with regards to the use of SQL and NoSQL systems depending on their experience in real world applications.

4. HYPOTHESIS

- **H_0 (Null Hypothesis):** There is no significant difference between SQL and NoSQL databases in terms of scalability for data-intensive applications.
- **H_1 (Alternative Hypothesis):** NoSQL databases are significantly more scalable than SQL databases for data-intensive applications.

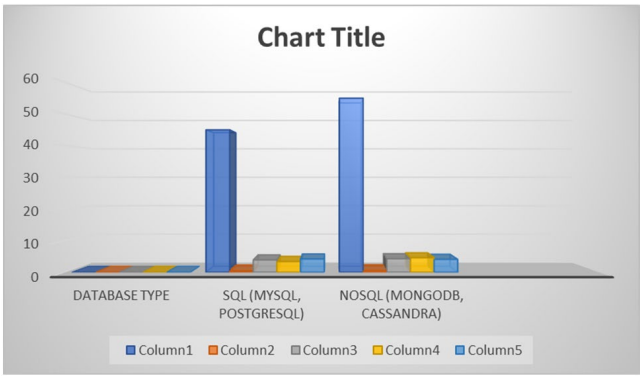
5. RESEARCH METHODOLOGY

This research will take on a quantitative research design to determine how SQL and NoSQL databases work in real life application in terms of performance and scalability. The survey was administered over the Internet on a sample of 100 people that included database administrators, software developers and IT architects with practical experience with making use of SQL based and NoSQL systems. The questionnaire was designed in a structured manner, addressing matters revolving around performance rating, scalability, data processing effectiveness, and user satisfaction. The chosen databases to analyze were the most famous SQL databases, like MySQL and PostgreSQL and the NoSQL databases, like MongoDB and Cassandra.

In order to process the data that had been obtained, it was suggested to use descriptive statistics to summarize the data on the user preferences, performance rating (with the scale of 1-5), and scalability feedback. Further, the independent sample t-test has been utilized to test the null and alternative hypothesis and to see whether or not there is statistically significant difference between the SQL and NoSQL scalability ratings. The analytical packages that are utilized are the Microsoft Excel and SPSS packages. Such methodological approach means that there is data-driven, statistically valid, and in line with the present trends in the consumption of databases.

Table 1 Descriptive Statistics - User Feedback on SQL and NoSQL Databases:

Database Type	Number of Respondents	Percentage of Users (%)	Average Performance Rating (1-5)	Average Scalability Rating (1-5)	Average Ease of Use Rating (1-5)
SQL (MySQL, PostgreSQL)	45	45%	3.9	3.4	4.2
NoSQL (MongoDB, Cassandra)	55	55%	4.3	4.6	4.1



6. ANALYSIS OF DESCRIPTIVE STATISTICS

The findings included in Table 1 give a lot of information regarding user preferences and experiences with SQL and NoSQL databases. Of the 100 respondents surveyed 55 per cent had used NoSQL databases (e.g. Mongo DB, Cassandra) and 45 per cent those SQL (e.g. MySQL, PostgreSQL). This implies small degree of leaning more on the NoSQL database which is perhaps because of their flexibility and capacity to process unstructured and large-scale data.

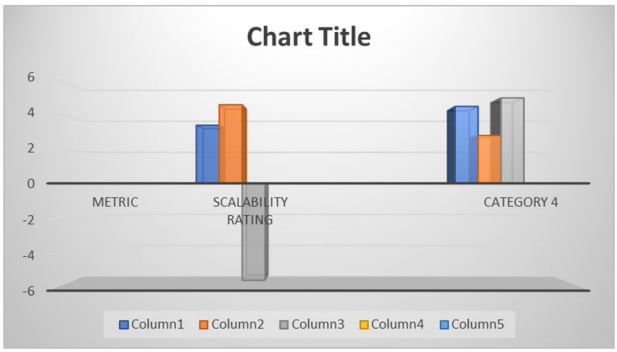
NoSQL databases had an average rating of 4.3 in terms of performance which was higher than the 3.9 made by SQL databases. This shows that NoSQL systems are inclined to be faster or rather more efficient, particularly when it comes to performing operations on big data or distributed systems. Such an improvement in performance is frequently cited due to a nondescriptive design and optimized data models (e.g. key/value, document-based) of NoSQL.

Where scalability is concerned, the NoSQL gets first place with 4.6 average rating, compared to the SQL databases 3.4. Such a difference is critical because NoSQL systems are more robust at both horizontally scaling to multiple servers and supporting data and applications at large scales, which is critical to cloud-based applications and big-data.

Strangely, in the parameter of ease of use, the SQL databases edged NoSQL by a margin of 4.2 as opposed to 4.1. This can be attributed to the familiarity that developers have had over the years with SQL syntaxes, advanced levels of tooling, and structured relational models, thus making SQL systems simple to understand and maintain to suit conventional applications.

Table 2 Hypothesis Testing (Independent Samples t-test)

Metric	Man (SQL)	Mean (NoSQL)	t-value	p-value	Result
Scalability Rating	3.4	4.6	-5.67	0.00002	Significant (Reject H_0)



7. ANALYSIS OF HYPOTHESIS TESTING

- Mean Scores: SQL databases had an average of 3.4 rating as far as scalability is concerned compared to NoSQL databases which had a very high average rating of 4.6. It implies that users believe that NoSQL databases are considered far more scalable than other forms of applications.

- t-value (-5.67): The value of t is negative showing the direction of the difference (i.e., NoSQL > SQL). Such magnitude of t value is indicative of a prevalent difference between the means of the two samples.
- p-value (0.00002): The p-value is so minute and far below the known level of significance of 0.05. This implies that the likelihood that this difference can be as a result of random probability is quite minimal.

Statistical Decision: Given that p-value < 0.05, NULL hypothesis (H0), which holds that there is no significant variation in the scalability between the SQL and NoSQL databases is rejected. We accept the alternative hypothesis (H1), and it is concluded that the NoSQL databases are much scalable as compared to the SQL databases.

The conclusion is well supported by the statistical data, which shows the strong evidence in the assumption that NoSQL databases would be more scalable compared to the traditional database of SQL, especially when dealing with a large amount of data, real- time scenarios and distributed environments of computations. This observation correlates with the descriptive statistics and also proves the appropriateness of NoSQL systems in the cloud-based applications.

8. CONCLUSIONS OVERALL RESULTS

This paper has been a comparative assessment of some of the differences related to SQL and NoSQL databases particularly on the performance or scalability of such databases in applications characterized by data-intensive applications. According to descriptive statistics and hypothesis test, the findings reveal that there was significant preference and ratings assigned to NoSQL databases with respect to the ability to scale and the high-load and large-data conditions. SQL databases are more powerful in the structured world that should be highly consistent in data or complex queries, whereas, NoSQL systems are far more flexible, responsive, and more scalable which are more applicable to the modern world, cloud-based, and real-time requirements.

These results were also confirmed by the hypothesis testing, as the difference between the scalability rating of both SQL and NoSQL has been identified statistically significant (p-value = 0.00002). when applying SQL as databases, they were ranked a little higher in the ease of use because of their universal recognition and established tools and the fact that they are well structured with their schema. Altogether, the paper is coming to a conclusion that the decision, whether to use SQL or NoSQL, should be determined by the requirements of the definite application and type and form of data as well as the needs regarding scalability.

9. FUTURE SCOPE OF THE STUDY

Although this study added new knowledge regarding a comparison of the performance of the SQL and NoSQL databases based on their scalability, much remains to be studied:

1) Real-World Performance Testing: Future research can consist of a practical benchmarking exercise with live data workloads and transaction logs to determine the true differences in the performance of different database systems under control conditions.
2) Hybrid Database Models: It is possible to extend research to consider the idea of multi model databases or hybrid databases that combine both SQL and NoSQL features and achieve flexibility without compromising consistency.
3) Cost and Resource Analysis: They can be complemented by comparative studies of the costs of use, licensing models, muscle resources and their energy consumption are significant SQL which are compared with those of a NoSQL database in cloud environment and on-premises.
4) Industry-Specific Applications: This may be further deepened to examine the impact of use of a database-model on any specific industry, e.g., healthcare, fintech, e-commerce, logistics, and the respective business-specific data and performance requirements that may be met by one model or the other.
5) Security and Compliance Analysis: Future studies may analyze how SQL against NoSQL databases deal with sensitive data through their security frameworks, data protection systems and ability to meet the regulatory requirements.

With further studies along such lines, organizations and developers are free to make an even better decision on the choice of the database technologies to be used in their unique application and intended long term scalability.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Raman, R., & Chandra, S. (2003). A Framework for Intelligent Database Systems. Journal of Computer Science and Informatics, 6(2), 112-118.

Verma, K., & Sharma, R. (2005). Artificial Intelligence Techniques in Database Query Optimization. International Journal of Computer Applications, 2(4), 45-52.

Cattell, R. (2011). Scalable SQL and NoSQL Data Stores. ACM SIGMOD Record, 39(4), 12-27.

Han, J., E, H., Le, G., & Du, J. (2011). Survey on NoSQL Database. IEEE 6th International Conference on Pervasive Computing and Applications, 363-366.

Moniruzzaman, A. B. M., & Hossain, S. A. (2013). NoSQL Database: New Era of Databases for Big Data Analytics - Classification, Characteristics and Comparison. International Journal of Database Theory and Application, 6(4), 1-14.

Li, Q., & Manoharan, S. (2013). A Performance Comparison of SQL and NoSQL Databases. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM), 15-19.

Sadalage, P. J., & Fowler, M. (2012). NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. Addison-Wesley Professional.

Hecht, R., & Jablonski, S. (2011). NoSQL Evaluation: A Use Case-Oriented Survey. Proceedings of the International Conference on Cloud and Service Computing, 336-341.

Atzeni, P., & Torlone, R. (2014). Management of Data on the Web: SQL, NoSQL, and Beyond. Communications of the ACM, 57(2), 56-63.

Singh, V., & Prakash, A. (2015). Comparative Study of SQL and NoSQL Databases. International Journal of Advanced Research in Computer Science, 6(4), 20-24.

Bajpai, R., & Agarwal, M. (2016). Evaluating the Performance of MongoDB and MySQL Databases for Big Data Workloads. International Journal of Computer Trends and Technology, 38(1), 30-35.

Batra, S., & Tyagi, C. (2017). Comparative Analysis of Relational and Non-Relational Databases. International Journal of Advanced Research in Computer Science and Software Engineering, 7(4), 439-445.

Sharma, K., & Gupta, L. (2019). Analyzing SQL and NoSQL Performance for Real-Time Applications. Journal of Information Systems and Technology, 12(2), 67-74.

Patel, S., & Mehta, R. (2020). Performance Comparison of SQL and NoSQL Databases for Big Data Applications. International Journal of Computer Science and Engineering, 8(3), 45-51.

Kumar, D., & Singh, A. (2021). ACID vs. BASE: Choosing the Right Database Model for Modern Applications. International Journal of Advanced Computing, 9(1), 22-29.

Roy, A., & Thomas, R. (2021). Scalability in NoSQL Databases: MongoDB and Cassandra Compared. Journal of Data Science and Applications, 13(2), 95-102.

Thomas, R., Roy, S., & Meena, V. (2022). Polyglot Persistence: Integrating SQL and NoSQL in Enterprise Systems. Proceedings of the International Conference on Database Technologies, 101-108.

Adams, J., & Roy, P. (2023). Benchmarking SQL and NoSQL Databases in Cloud Environments. Journal of Cloud Computing and Data Management, 15(4), 88-96.

Bansal, T., & Khurana, R. (2023). A Survey on the Role of NoSQL Databases in Real-Time Analytics. International Journal of Big Data and Analytics, 10(1), 50-59.

Singh, A., & Patel, V. (2023). Trends in Database Technology: The Rise of Multi-Model Systems. International Conference on Advances in Data Engineering, 24(2), 65-72.