

A systematic mapping study of comparative review of Relational and NoSQL database models

Mohammad Idris Attal
Free University of Bozen-Bolzano
Bolzano, South Tyrol, ITALY
mohammadidris.attal@stud-inf.unibz.it
i.attal.npa@gmail.com

ABSTRACT

Context: Prior research compared NoSQL and relational databases using a single domain focus approach, and this overview left the technical debt of gaining a general comparative overview of the topic.

Objective: The purpose of this study is to find primary studies that compare NoSQL database model with the relational database model and help in giving an overview of which factors affecting in choosing one model over the other for a given criteria.

Method: A systematic mapping study (SMS) was carried out which applied manual searches to discover relevant primary studies.

Results: 26 primary studies were identified which presents the relevant information about the comparisons, affecting factors and applied experiments on each model, and such studies present the evidence of how each comparison reached its final findings.

Conclusion: It has been found from the extensive analysis of the comparisons that NoSQL database models are introduced as an alternative solutions and they are not a replacement for relational databases and in fact both of them can coexist together. Choosing of one over the other depends on the certain criteria of the business needs and each model has advantages and disadvantages.

CCS CONCEPTS

• **Information systems** → *Data management systems, Query languages, Information retrieval*; • **Theory of computation** → *Database theory, Database query languages (principles), Theory and algorithms for application domains*.

KEYWORDS

Relational Database, SQL, NoSQL, Non Relational Database

1 INTRODUCTION

The databases have been one of the most common ways to store data for the past few decades, and we have seen them as a useful tool for managing and storing data. Relational databases have been used for managing data over the years, and they have shown prominent results in managing data[8]. However, with the rapid change of technologies and the huge amount generated data from different resources in structured, semi-structured, or unstructured forms, the questions of scalability and availability of data concerns many businesses when it comes to choosing the proper choice of the database model for their data management. In order to handle such problems and concerns, NoSQL databases were suggested as an alternative option to bring further improvements[8].

Companies and businesses are dealing with tons of data every day, and such data are generating from different sources and in

different shapes. Relational databases are mostly used in such kinds of applications, but they show a degrade in performance when the amount of data is increasing and handling of such amounts of data becomes an issue[8][3]. NoSQL databases were an alternative model for addressing such problems and they were suggested as another solution to the existing problem but not as a complete replacement for relational databases, and this emphasizes the possibility of co-existence of both technologies together[8]. Data storing and schema definition are features of relational databases which NoSQL databases do not use or somehow use in a loosely coupled fashion. So, this exploration gives us the details that there is no definite solution but there are some pros and cons of each database model.

In this paper we study the overview of factor affecting in choosing between relational and NoSQL database models, we use systematic mapping study (SMS) approach to collect data and show our findings based on the research. The results of this SMS indicates that there is no definite or solid solution for choosing a database model for a business[8][3], and there is little correlation between performance and the data model each database uses[13].

The rest of the paper is structured as follows: section 2 discusses related work of the current study on relational and NoSQL databases; section 3 discusses mapping study process conducted to select and classify papers; section 4 discusses the results of mapping; section 5 discusses conclusion and future work and outlines possible suggestions.

2 BACKGROUND AND RELATED WORK

This section presents the theoretical background and related research which has been done related to the comparison of relational and NoSQL databases and the affecting factors in choosing one over the other.

In the [8] the author indicates his findings by supporting the notion that NoSQL was introduced as a supporting solution to the existing relational databases, and further explains NoSQL as a solution for those concerns which relational databases could not fulfill. These findings also express the coexistence of both technologies together and explain that NoSQL is not a replacement for relational databases. Relational databases support structured data and NoSQL databases support semi-structured or unstructured data in a loosely coupled fashion. This explains there are advantages and disadvantages of both databases and there is no definitive or perfect solution.

In the [3] the author explains that massive amounts of unstructured data is produced from different sources such as real-time applications and social media sites and it is really hard to be manage

through relational databases because they only work with structured data. Furthermore, when the volume of data increases in relational databases, performance decreases rapidly. NoSQL databases were introduced to meet the requirements of the new class applications. NoSQL databases handle unstructured data efficiently and also offer scalability of application in a promising manner.

In [5] the author expresses the notion that in unstructured data NoSQL databases perform better than Relational databases, but there is a need for testing the performance of both databases in structured data as well and characteristics of each database model are different such as Relational databases offer ACID (Atomicity, Consistency, Isolation, Durability) for transactions, which assures proper completion of transactions and data consistency. However, NoSQL database model uses BASE (Basically, Available, Soft-State, Eventually Consistent) for high scalability, availability, and performance. Since Relational and NoSQL databases show different characteristics, one can not replace the other and their scope of domain is different. For instance, relational databases are used mostly in applications where there is a need for financial rules, internal control, complex queries and aggregate functions, whereas NoSQL is used in applications that do not need transactions and consistency.

The paper [6] express the notion of that not all NoSQL databases perform better in compare to relational ones, some are even worse. Furthermore, the paper emphasizes that performance varies for each database model based on its perform operations, for instance some are slow to instantiate but fast enough to write, read and delete. On the other hand, some are fast instantiate but slow to other operations. Finally, the paper expresses there is no correlation between performance and data model each database uses.

The paper [10] expresses the notion of NoSQL databases performing better than relational databases when they are handling a huge amount of data. The author express in the paper this notion by further explaining it in a comparison of two data models, such as Oracle as relational and MongoDB as non relational one. He gives the evidence from the experiment and shares the results of the experiment by describing that MongoDB performs better for insert, delete and select operation, whereas for update operation Oracle performs better than MongoDB in the given environment.

The paper [12] expresses the notion that NoSQL databases are replacing Relational databases is misleading, and NoSQL databases are introduced to be used as an additional option in applications when the use of relational databases is less appropriate due to modern applications requirements. NoSQL and relational databases are not comparable because they are designed for different purposes. Therefore, it is recommended to use NoSQL or relational databases based on the needs of the software or project.

The paper [1] expresses the notion that relational databases face many challenges when it comes to handle huge amounts of data or big data applications, and this requires a high-rise in read and write performance. Since the use of relational databases has proved to be inadequate when it comes to store and query dynamic user data, especially for bulky and highly parallel applications such as social networks. Therefore, NoSQL was created to handle such cases, and the author in the paper investigates NoSQL databases performance and suggests companies to choose NoSQL databases by describing its mainstream benefits over the relational databases.

In [11] the author expresses his notion by explaining that NoSQL database is schema less and allow horizontal scalability and it has advantages over the relational databases while dealing with unstructured or huge amount data. Furthermore, he expresses his findings on the comparison of SQL and MongoDB(NoSQL) databases by stating that NoSQL queries are executed faster as compared to SQL queries and also NoSQL databases can handle unstructured data very easily.

3 METHODOLOGY

A systematic mapping study (SMS) was conducted on the factor affecting in choosing between relational and NoSQL database models. This SMS will have the following steps: definition of research question, conducted research, screening of papers and data extraction.

3.1 Objective

The objective of this study is to identify primary studies on comparison of relational and NoSQL databases and how this comparison helps in finding the important factor affecting in choosing between relational and NoSQL database models.

3.2 Research Question

In order to fulfill the purpose of this study, we seek to provide answers to the following research questions:

- **RQ:** What is the distribution of research topics about relational and NoSQL models?
- **RQ:** What is the type of studies performed in comparison of relational and NoSQL models?
- **RQ:** What is the ratio of study type classification versus topics exploration?
- **RQ:** What is the ratio of study type classification versus database models used in the experiments?

From the research question, it is expected to answer what are the key factors involved in choosing a database model for your targeted system or software, and how can the current existing research in this area helps in the selection of the targeted database model.

3.3 Search Strategies

To find studies for this SMS, automatic search were made in three prestigious digital libraries of computer science domain. The three selected libraries were IEEE Xplore, Scopus and ACM Digital Library. The studies were found by using the search strings adapted to each database format and the search was a combination of keywords related to the comparison of relational and NoSQL databases. The following search term was used to find results for this SMS:

- (1) ("relational database" OR SQL) AND ("NoSQL database" OR NoSQL) AND ("comparison" OR "comparing" OR "comparative")

The only difference in the search strings was about the ACM library where in addition to our main search strings we also added the abstract and keyword combination in order to get the optimal result-set in this digital library. The added section for the abstract

and keyword was the addition of this string "(\"relational database\" OR SQL) AND (\"NoSQL database\" OR NoSQL)".

Another important point in our search strategy is about Inclusion and exclusion criteria, which set the boundaries for the systematic mapping study (SMS), so, here we presents the criteria that will be used to determine which research studies will be included and which should be excluded due to the predefined criteria. Search results were collected with the consideration of the below inclusion and exclusion criteria:

3.3.1 Inclusion Criteria.

- The study is about the comparison of relational and NoSQL databases
- The study is about the comparison of a specific model of relational and NoSQL databases

3.3.2 Exclusion Criteria.

- All the publication which is not in English
- Not related to the software industry
- Not related to the searched key words
- The publication which is in search query but not focused on the main topic (topics related to IOT)
- All the publication which is published before 1999
- All the publication which is not in type of Journal or Conference

The main reason for including publication starting form the year 1999 is because of the introduction of NoSQL databases as a technology on the year 1998.

3.4 Study Selection Process

The study selection process started from the application of query string in each digital library with the consideration of the application of exclusion criteria. The method by which we arrived at the intended outcome will be thoroughly explained in the following paragraphs:

The process started with the application of query string in each digital library and with the application of query string we have reach to the total number of 494 papers which were in two main types of academic publications, journal and conferences. In order to narrow down the amount of papers which were returned with the application of the query string, we have have applied the inclusion criteria by starting to read the title of each paper and decide to include it for further process or not. This process was fast and rapid and we have shortlisted 183 papers out of 494 which were related to the topic.

The next step in the process was to exclude all the duplicate papers in the total amount of related papers which were 183 papers and after our examination we have reach to 132 papers which were shortlisted to be included for our next step in the exploration process.

Consequently, the next step in the process was a manual process of initial review of 132 papers by reading the abstract of each paper, which results in the selection of 49 papers out of 132 as relevant papers for further exploration. After the further exploration of 49 papers and further reading of each paper some as full-text and some just the introduction and conclusion, we have come up with the 26 papers as finally selected to be included in the SMS.

The main reasons of selecting 26 papers out of 49 in the further exploration phase was, first, the focus of some papers was not mainly on the comparison of relational and NoSQL databases but something else such as a specific tool or a specific project and secondly, some papers were talking about only one database model or mixing this knowledge area with other concepts such as the combination of big data with database models and so on.

The final selected papers are in line with answering our research question in full manner and they are related to the context. An example of the listing of considered papers for our SMS study can be found at this link: <https://docs.google.com/spreadsheets/d/1WaSknlTQiBnC2Z7TdeF-7AbfsA2IB-O46r-zNALdkm8/edit#gid=0>.

3.4.1 Overview of the Studies. A visual representation of the process study selection and overviews of studies are shown in the Figure 1 of this SMS.

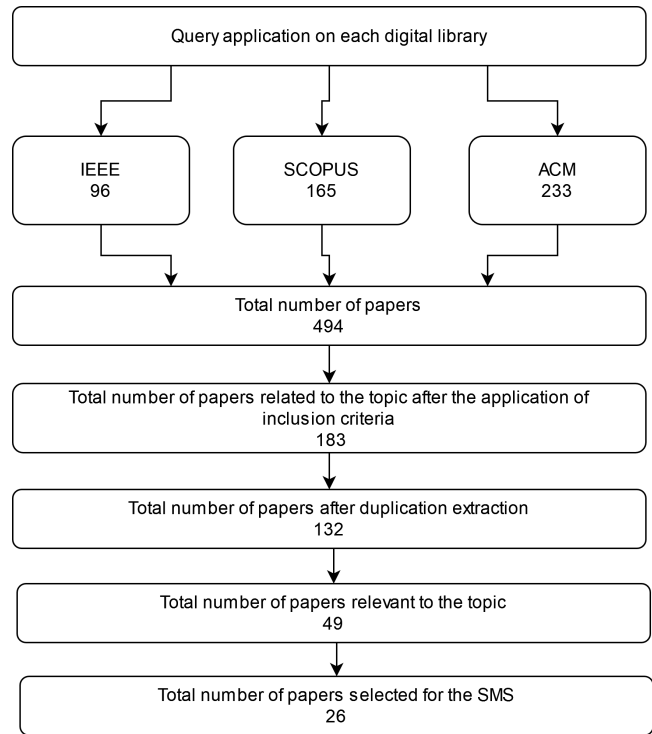


Figure 1: Overview of papers selection process for the SMS

4 RESULTS OF THE MAPPING

This section gives a clearer picture of the studies and the comprehension of their overall scope. It also explains how the quality of the mapping results was assessed by addressing the research questions.

4.1 Scope and Trend of Research Area

The process of analyzing the scope and trend of research in this research area will be really hard to examine and such examination will lead to a subjective result regarding the scope and trends of this research area. With the consideration of all the limitations in this overview and with the help of our final 26 papers which were explored in this SMS, we have come up with such an analysis.

Results of our examination show there has been a balanced research coverage in this research area since 2011 and our findings show there was a tempting increase of research over the years in this area and it is expected that such kind of interest will continue in the future as well. A visual representation of the trends over the years has been shown in the Figure 2 of this SMS.

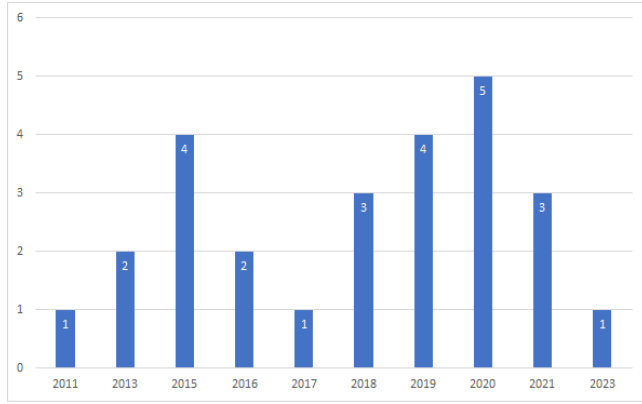


Figure 2: Overview of numbers of papers published annually

4.2 Topics Exploration in the SMS

The topics exploration section answer to the research question of "What is the distribution of research topics about relational and NoSQL models" by providing the following information.

The process of topics exploration in this SMS has been executed manually and the results of such exploration have been categorized into three main areas. The first main area is about the findings related to the notion which introduces NoSQL as an alternative solution to the current existing relational databases and further elaborates the idea that NoSQL is not a replacement for the current relational databases and both models can coexist together[3, 6, 8, 10]. The other important aspect of topics exploration discussion is about the notion which expresses NoSQL as a better choice over relational databases for handling huge amounts of data. Furthermore, the selection of NoSQL over the relational database model is preferred because of the efficiency of such models in scalability and performance[1–5, 7]. Finally, the last part of topics exploration findings shows that each database model has advantages and disadvantages and it is highly recommended for businesses to choose each model based on their business needs[2, 5, 9, 11]. A visual representation of the topics exploration has been shown in the Table 2 of this SMS.

Table 1: An Overview of Topics Exploration in The SMS

Key exploration outcome	Studies
Introduce NoSQL as an alternative solution; Support the notion NoSQL is not a replacement for Relational databases;	[S01][S02][S04][S05] [S06][S07][S11][S25]
NoSQL is better choice for scalability; NoSQL handle huge data efficiently; NoSQL Performs better than Relational databases	[S02][S03][S09][S10] [S12][S13][S14][S15] [S16][S17][S18][S19] [S20][S21][S26]
Each database model have advantages and disadvantages and businesses should choose it based on its needs;	[S03][S08][S17][S18] [S20][S23][S24][25]

4.3 Study Type Exploration in the SMS

The study type exploration or classification section answer to the research question of "What is the type of studies performed in comparison of relational and NoSQL models" by providing the following information.

The study-type exploration in the SMS demonstrates the findings of such classification through a manual process and such findings are subject to the specific body of knowledge of this SMS only. Therefore, the classification of study-type is dependable on several internal factors involved in each SMS which will result in subjective classification in each SMS study. With the consideration of all involved factors and aspects in this SMS, we have come up with five common categories of the study-type such as Comparison study, Evaluation Study, General Comparison Study, Comparative Survey and Research Discussions.

Comparison study is the type of studies which is focused on the comparison of relational and NoSQL database models by comparing a specific model of each version for instance the comparison of MongoDB with MySQL and so on. Evaluation study is the type of studies which is focused on the evaluating of each database model by bench-marking each model on the provided experiment environment. General comparison study is the type of studies which is comparing relational and NoSQL database model in general without of considering a specific database model version such as MongoDB, MySQL, Cassandra and so on. Comparative survey is the type of studies in which the researcher tried to performed surveys to conclude his findings related to relational and NoSQL database models. Finally, research discussion is the type of study which was presenting its research finding in the form of a research discussion which ties everything to its research question.

A visual representation of the study-type has been shown in the Figure 3 of this SMS.

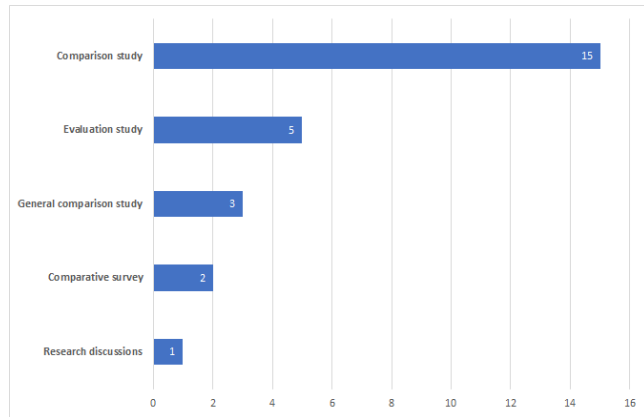


Figure 3: Overview of study-type classification in the SMS

The two most prestigious and important academic publication type journal and conferences have been considered as the main source of conducting this SMS. A visual representation of the publication type has been shown in the Figure 4 of this SMS.

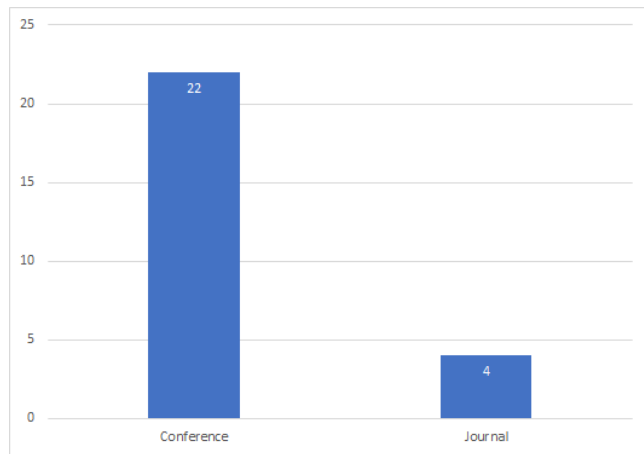


Figure 4: Overview of number of papers by publication type

The overview of study-type classification versus topics-exploration expresses answer to the research question of "What is the ratio of study type classification versus topics exploration" by providing the following information.

A thorough examination of study-type with topics-exploration has been reviewed in this part of the SMS. The main take away from such exploration was to explore the relationship between these two areas and further examine which type of studies are mostly in favor of which type of findings. Indication of such kinds of examination will help the reader to get a clear picture of the correlation between these two territories. A visual representation of such findings has been shown in the Figure 5 of this SMS.

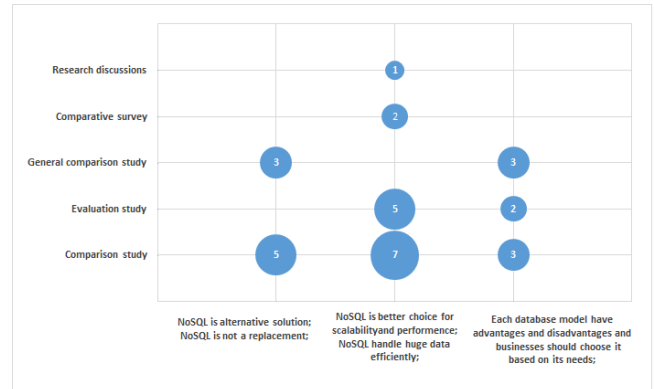


Figure 5: Overview of study-type classification vs topics-exploration

In order to examine the correlation between study-type and database-model-type-experiments, we need to first examine the database-model-type-experiment exploration in this SMS.

The database-model-type-experiment exploration has been achieved after a thorough examination of each paper involved in the SMS. Such exploration was categorized based on the parameters involved in each comparison and based on such categorization we have come up with two initial outcomes. First, there was no specific model comparison in the studies and they were suggesting the notion of using each database model based on the certain needs of the application and business model. Second, they were comparing specific types of database models and based on each comparison final findings were presented in each paper which mostly expressing the idea that NoSQL database models perform better than relational ones with the huge amount of data[3–5, 7]. A visual representation of the database-model-type-experiment has been shown in the Figure 6 of this SMS.

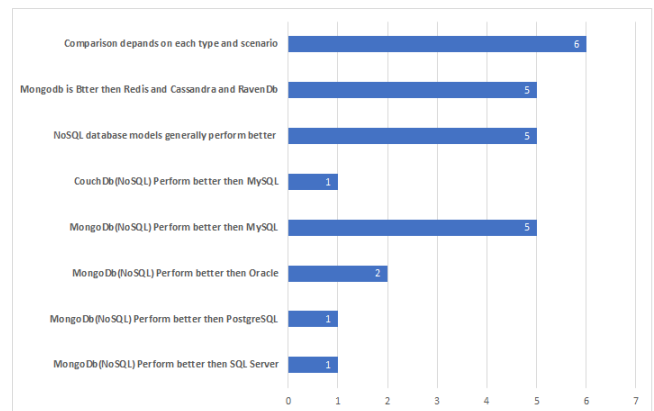


Figure 6: Overview of numbers of database-model-based-experiment in the SMS

The Overview of study-type classification vs database-model-based-experiment in the SMS expresses answer to the research question of "What is the ratio of study type classification versus

database models used in the experiments" by providing the following information.

After the thorough examination of the database-model-based-experiment on the above mentioned paragraph, we took further our SMS findings by exploring the examination of study-type with database-model-based-experiment. The main outcome of such exploration was to explore the relationship between these two areas and further examine which type of studies are mostly in favor of database-model-based-experiments. Indication of such kinds of examination will help with the reader to get a clear picture of the correlation between these two territories. A visual representation of such kinds of findings has been shown in the Figure 7 of this SMS.

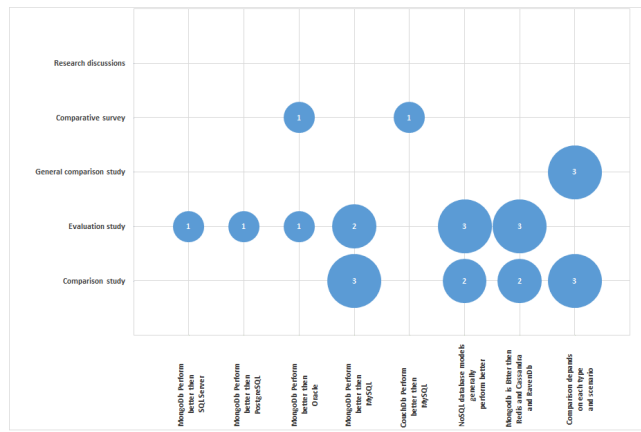


Figure 7: Overview of study-type classification vs database-model-based-experiment in the SMS

5 CONCLUSION AND FUTURE WORK

This paper presented a systematic mapping study on the comparison of NoSQL and relational database models and the factors affecting on choosing each model. The aim of this study was to identify primary studies that examine the affecting factors involved in the comparison of NosQL and relational database models and based on which criteria such models are compared. The database search returned 494 studies. After applying inclusion criteria, 183 were identified as related studies and the exclusion of duplicate studies limit such results to 132 studies, after the application of the review, we have come up to 49 studies which were potentially relevant, but in the further review and extraction phase, it was found that only 26 could answer our research question thoroughly.

6 out of those studies were stressing out the notion that there is no specific or better database model solution and the choosing of each model depends on the type and scenario of each business. However, the rest of the papers taking another approach towards the comparison and they have done a specific model comparison on each types of NoSQL and relational database models. 5 of the studies are expressing the notion that generally NoSQL databases perform better with huge or unstructured amounts of data. The other 5 of these studies compare MongoDB and MySQL database models and present their results by expressing that MongoDB performs better than MySQL. 4 of these studies are comparing MongoDB with other

relational database models such as PostgreSQL, Oracle and SQL Server and the results express the notion that overall MongoDB is performing better with unstructured data. Finally, 5 out of 6 remaining studies are ranking comparisons between the NoSQL database models by expressing that Mongodo performs better than Redis and Cassandra and RavenDb and the last remaining paper is talking about the comparison of CouchDb and MySQL and expresses the notion that CouchDb performs better in general.

5.1 Limitation and threats of validity

This study is limited to the results of three digital libraries which were selected for this SMS study. This indicates that there are limitations in performing this mapping study and there is no single study existing in the form of mapping study in this area. In addition to this, other important factors involved would be selection bias or even bias in the extraction of data or the problem of misclassification, so by considering all the above mentioned points, this SMS could update in future in order to answer all questions in this domain.

The systematic mapping process was carried out by just one researcher, which raises concerns about potential distortions in study selection, data extraction, and quality assessment. To mitigate this threat, after the conclusion of each step, a review phase was conducted by the researcher to follow up with the criteria defined for the evaluation.

REFERENCES

- [1] J. Chaudhary, V. Vyas, and C. K. Jha. 2023. *Qualitative Analysis of SQL and NoSQL Database with an Emphasis on Performance*. Smart Innovation, Systems and Technologies, Vol. 312. 155–165 pages. www.scopus.com
- [2] C. A. Györfi, D. V. Dumşeu-Burescu, D. R. Zmaranda, R. Ş. Györfi, G. A. Gabor, and G. D. Pecherle. 2020. Performance analysis of nosql and relational databases with couchdb and mysql for application's data storage. *Applied Sciences (Switzerland)* 10, 23 (2020), 1–21. www.scopus.com Cited By :7.
- [3] Mohammad A. Hassan. 2021. Relational and NoSQL Databases: The Appropriate Database Model Choice. In *2021 22nd International Arab Conference on Information Technology (ACIT)*. 1–6. <https://doi.org/10.1109/ACIT53391.2021.9677042>
- [4] Min-Gyue Jung, Seon-A Youn, Jayon Bae, and Yong-Lak Choi. 2015. A Study on Data Input and Output Performance Comparison of MongoDB and PostgreSQL in the Big Data Environment. In *2015 8th International Conference on Database Theory and Application (DTA)*. 14–17. <https://doi.org/10.1109/DTA.2015.14>
- [5] S. Kanchan, P. Kaur, and P. Apoorva. 2021. Empirical Evaluation of NoSQL and Relational Database Systems. *Recent Advances in Computer Science and Communications* 14, 8 (2021), 2637–2650. www.scopus.com Cited By :1.
- [6] Yishan Li and Sathiamoorthy Manoharan. 2013. A performance comparison of SQL and NoSQL databases. In *2013 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*. 15–19. <https://doi.org/10.1109/PACRIM.2013.6625441>
- [7] Sowndarya Palanisamy and P. SuvithaVani. 2020. A survey on RDBMS and NoSQL Databases MySQL vs MongoDB. In *2020 International Conference on Computer Communication and Informatics (ICCCI)*. 1–7. <https://doi.org/10.1109/ICCCI48352.2020.9104047>
- [8] Kosovare Sahatqija, Jaumin Ajdari, Xhemal Zenuni, Bujar Raufi, and Florije Ismaili. 2018. Comparison between relational and NoSQL databases. In *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. 0216–0221. <https://doi.org/10.23919/MIPRO.2018.8400041>
- [9] M. Sandeep Kumar and J. Prabhu. 2018. Comparison of nosql database and traditional database-an emphatic analysis. *International Journal on Informatics Visualization* 2, 2 (2018), 51–55. www.scopus.com Cited By :5.
- [10] B Sirish Shetty and KC Akshay. 2019. Performance Analysis of Queries in RDBMS vs NoSQL. In *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, Vol. 1. 1283–1286. <https://doi.org/10.1109/ICICICT46008.2019.8993394>
- [11] N. Srivastava. 2019. Nosql - A solution for handling big data. *International Journal of Engineering and Advanced Technology* 9, 1 (2019), 7606–7610. www.scopus.com

- [12] S. Uyanga, N. Munkhtsetseg, S. Batbayar, and S. Bat-Ulzii. 2021. *A Comparative Study of NoSQL and Relational Database*. Smart Innovation, Systems and Technologies, Vol. 212. 116–122 pages. www.scopus.com Cited By :1.
- [13] Rashid Zafar, Eiad Yafi, Megat F. Zuhairi, and Hassan Dao. 2016. Big Data: The NoSQL and RDBMS review. In *2016 International Conference on Information and Communication Technolgy(ICICTM)*. 120–126. <https://doi.org/10.1109/ICICTM.2016.7890788>