A Systematic Literature Review on Migrating Relational Databases to NoSQL Databases

Alberto da Silva Gideão[1,2], **Ronaldo dos Santos Mello**[1]

[1]Departamento de Informática e Estatística – Programa de Pós-Graduação em Ciência da Computação (PPGCC) – Universidade Federal de Santa Catarina (UFSC) Florianópolis – SC – Brasil

[2]Instituto Superior Politécnico – Caála – Huambo – Angola

albertogideao20@gmail.com,r.mello@ufsc.br

Abstract. *The growing need to manage large volumes of heterogeneous data has driven the adoption of NoSQL databases, which offer greater flexibility and scalability compared to traditional relational databases. However, data migration between these models still is a challenge, as there is not a robust solution. This work presents a Systematic Literature Review (SLR) on recent approaches for migrating relational databases to NoSQL databases, as well as an analysis of the selected studies. The results highlight trends and gaps in the literature, emphasizing the need for more efficient approaches to support relational data migration to different NoSQL technologies. This work contributes to the systematization of knowledge in the field and provides insights for future researches.*

Keywords*: Migration, Relational Database, NoSQL Databases, Workload.*

Resumo. *A crescente necessidade de gerenciar grandes volumes de dados heterogêneos tem impulsionado a adoção de bancos de dados NoSQL, que oferecem maior flexibilidade e escalabilidade em comparação com os tradicionais bancos de dados relacionais. No entanto, a migração de dados entre esses modelos ainda é um desafio, uma vez que a literatura carece de uma solução robusta. Este trabalho apresenta uma Revisão Sistemática da Literatura (RSL) sobre abordagens de migração de bancos de dados relacionais para bancos de dados NoSQL, assim como uma análise dos estudos selecionados. Os resultados apontam tendências e lacunas na literatura, evidenciando a necessidade de abordagens mais eficientes para apoiar a migração de dados relacionais para diferentes tecnologias NoSQL. Este trabalho contribui para a sistematização do conhecimento na área e oferece insights para pesquisas futuras.*

Palavras-chave*: Migração. BD Relacional. BDs NoSQl. Carga de trabalho.*

1. Introduction

The migration of relational databases to NoSQL databases has become an increasingly relevant topic in recent years, driven by the need to handle *Big Data*, i.e., large volumes

of heterogeneous data, as well as the speed at which these data are generated. While relational databases have historically been predominant in organizing and storing structured information, NoSQL databases have emerged as a more flexible and scalable alternative, allowing adaptation to heterogeneous, complex, and large-scale data models, which are common in the Big Data era [Sadalage and Fowler 2013, de Lima and Mello 2016, Kleppmann 2016].

However, the migration from relational databases to NoSQL still represents a challenge. Fundamental differences between data models, the way transactions are processed, and the need to preserve data integrity during the transition make the process complex and require well-structured methodologies [Han et al. 2013]. Despite the growing adoption of NoSQL databases, there is still no widely accepted method for efficiently and reliably performing this conversion, especially considering the complexity of operations carried out in relational databases.

Given this scenario, this study aims to conduct a Systematic Literature Review (SLR) to identify recent methods, strategies, and challenges in migrating relational databases to NoSQL databases, as the literature focuses on the one-to-one mapping between relational model and a single NoSQL data model. To achieve this, we follow Kitchenham's methodological guidelines [Kitchenham 2004] to ensure a careful selection of academic publications.

This review aims to provide valuable insights for both the academic community and IT professionals, assisting in the development of more effective solutions for data migration between different database paradigms. Additionally, the study seeks to contribute to the systematization of knowledge in the field, identifying gaps and opportunities for future research on the subject.

The rest of this paper is organized as follows. Section 2 presents some related concepts. Section 3 details the application of the SLR, Section 4 gives an overview of the selected studies and Section 5 presents a comparative analysis of them. At last, Section 6 is dedicated to the conclusion.

2. Fundamentals

This section details some fundamental concepts underlying the topic at hand, starting by the notion of Big Data. Big Data refers to the enormous volume of data generated daily, which stands out not only for its quantity but also for the variety and speed at which it is produced and processed. These data are widely used across different sectors, from scientific research to the business and social worlds, directly influencing decision-making processes. Nevertheless, the analysis of large volumes of data can significantly transform various fields, enabling the discovery of complex patterns and strategic insights [Sun 2024].

Relational databases face challenges in handling Big Data, as their rigid model does not adapt well to semistructured and unstructured Big Data, and their horizontal scalability is limited. To address these challenges, NoSQL databases emerged, offering greater flexibility and distributed processing, making them an effective alternative for large volumes of data [Sadalage and Fowler 2013, Shah et al. 2021].

NoSQL databases are classified into four main categories: *(i) document-oriented*,

such as MongoDB, which stores data mainly in JSON or BSON format; *(ii) column-based*, such as Cassandra, optimized for large volumes of analytical data; *(iii) graph-based*, such as Neo4j, focused on modeling complex relationships; and *(iv) key-value stores*, such as Redis, ideal for fast read and write operations [Sadalage and Fowler 2013]. Additionally, unlike relational databases that follow the ACID model, NoSQL databases adopt the BASE model, prioritizing availability and scalability in distributed systems, even if immediate consistency is sacrificed.

Although NoSQL databases bring significant benefits in large-scale and dynamic data scenarios, they do not completely replace relational databases but rather complement them. While relational databases are essential for applications requiring strong consistency and transactional integrity, such as financial systems, NoSQL databases offer greater flexibility and performance for distributed and high-volume data applications. Thus, the choice between one model or the other depends on the specific needs of each application, balancing scalability, structure, and reliability [Sadalage and Fowler 2013]. Our research is motivated by the need of an application to migrate its relational schema and data to one or more NoSQL databases in order to better manage heterogeneous data and leverage Big Data processing for some kinds of operations.

3. Systematic Literature Review

[Wazlawick 2009] emphasizes the importance of methodological procedures in scientific research, highlighting that these processes are essential to ensure the validity and reliability of the results obtained. The SLR is a methodology that enables the identification, evaluation, and critical interpretation of existing research on a specific topic, phenomenon, or research question. The SLR allows researchers to gain a comprehensive and detailed view of advancements in the field of study, which is crucial for developing new solutions or improving existing methods.

To conduct this SLR, the guide proposed by Kitchenham (2004) was used, recognized for its widespread acceptance and application in scientific literature. The primary objective of this SLR is to collect, analyze, and synthesize existing research on data migration between relational and NoSQL databases, identifying the methods, tools, and criteria used for this transition.

Based on Kitchenham's guide, three research questions (RQ) were formulated to guide this review:

- **RQ-1**: What are the approaches for migrating relational databases to NoSQL databases?
- **RQ-2**: What are the considered NoSQL data models by the migration approach?
- **RQ-3**: The approach presents any evaluation regarding performance and efficiency of the migration, and what metrics can be used to evaluate these impacts?

In order to search for works related to the migration of relational databases to NoSQL databases, a search was conducted in four renowned academic repositories: A*CM-DL, IEEE Xplore, SCOPUS,* and *Springer* Link. These repositories were selected due to their breadth and high quality in publishing scientific articles on technology and computer science, especially in the field of Databases. The search was automated using specific search strings developed to cover the main topics of the work, including titles, abstracts, and full texts of the articles. The search expressions used were:

Text: *("relational database" AND "NoSQL migration" AND modeling) OR ("database migration" AND "NoSQL modeling") OR ("relational database" AND "NoSQL" AND performance)*;

Title: *"relational to NoSQL migration" OR "NoSQL database models" OR "relational to NoSQL mapping" OR "relational to NoSQL conversion"*;

Abstract: *migration OR "NoSQL databases"*.

These strings were adjusted for each repository's interface to ensure that the studies found were directly relevant to the topic of data migration. Additionally, inclusion and exclusion criteria were defined to further refine the search results. The inclusion criteria considered primary research articles published in high-impact journals written in English with focus on migration from relational to NoSQL databases, and published between 2010 and 2024. In order to ensure relevance and quality, we also defined the following exclusion criteria: *(i)* secondary articles, such as literature reviews not focused on data migration; *(ii)* duplicate studies; and *(iii)* studies with inaccessible content.

Table 1 presents the repositories searched and the results obtained. Initially, 3,895 works were found using the search strings. Next, the inclusion and exclusion criteria were applied to the titles and abstracts, reducing the number to 818. After applying the same criteria to the full text, the final number obtained was 9.

Tabela 1. Searched Repositories and Results

Repository	Initial Selection	Final Selection
IEEE Xplore	32	8
ACM-DL	495	0
SCOPUS	37	1
SpringerLink	254	0
Total	**818**	**9**

The 9 selected works are detailed in the next section.

4. Selected Studies

The studies related to the migration from relational databases to NoSQL databases address different methods and data models aimed at meeting the growing demands of Big Data management. They are summarized in Table 2.

A migration model was proposed by [Namdeo and Suman 2021] to transform historical and real-time data in parallel, implemented in MySQL and MongoDB, demonstrating superior performance compared to other approaches. On the other hand, [Hamouda 2023] presents a strategy based on transformation rules, evaluating database properties to migrate from relational to document-oriented databases, proving its effectiveness in reducing migration time.

[Tandya and Azizah 2023] designed algorithms that use relational database metadata, focusing on embedding and referencing based on workload. Despite performance improvements, challenges remain in scenarios with high-frequency reads of embedded documents. [Siregar and Azizah 2023] deals with migration to column-oriented NoSQL

Tabela 2. Selected Studies

#	Reference	Title	Vehicle
1	[Alami et al. 2024]	Migrating Relational Databases to NoSQL-Oriented Documents Using Object-Oriented Concepts	IJIES
2	[Hamouda 2023]	Seamless Transition: Migrating from Relational Databases to Document-Oriented Databases	IDAACS
3	[Tandya and Azizah 2023]	Migration of Relational Database to NoSQL Document-Oriented Database	ICoDSE
4	[Siregar and Azizah 2023]	Migration From Relational Database to Column-Oriented NoSQL Database	ICoDSE
5	[Eldrrat and Maatuk 2023a]	Data Migration from Conventional Databases into NoSQL: Methods and Techniques	MI-STA
6	[Eldrrat and Maatuk 2023b]	A Novel Approach for Migrating Relational Databases into Graph Databases	eSmarTA
7	[Namdeo and Suman 2021]	A Model for Relational to NoSQL Database Migration: Snapshot-Live Stream Db Migration Model	ICACCS
8	[Yassine and Awad 2018]	Migrating from SQL to NOSQL Database: Practices and Analysis	IIT
9	[Karnitis and Arnicans 2015]	Migration of Relational Database to Document-Oriented Database: Structure Denormalization and Data Transformation	ICCICSN

database Cassandra. It proposes a set of direct and indirect methods, with the latter showing better performance due to the extraction of the ER diagram and conversion to a suitable NoSQL logic schema.

Approaches using common mappings between relational and NoSQL databases were analyzed, suggesting combinations of nested documents with referenced relationships to optimize retrieval time [Yassine and Awad 2018]. Additionally, semi-automatic migration methods were implemented in database browsers, enabling the rapid creation of proof-of-concept for new NoSQL solutions [Karnitis and Arnicans 2015].

Another relevant study is [Eldrrat and Maatuk 2023a], which discusses strategies and techniques for migrating data from relational databases to NoSQL graph databases. The authors analyze the main challenges of this transition, including data model adaptation, information integrity, and performance in the new environment. They developed algorithms based on conceptual models to convert relational data into graphs, ensuring a better equivalence in data representation.

5. Comparative Analysis

Table 3 presents a comparison of the related works, highlighting: *(i)* Supported data models; *(ii)* Consideration of workload; *(iii)* Post-migration scalability. The first criteria is

interesting in order to see how comprehensive is the approach in terms of NoSQL data models it is able to migrate relational schema and/or data. The second criteria investigates if the approach considers an expected workload besides the relational schema. Workload information is usually a good input to guide the migration to a suitable data representation in the NoSQL database that performs better for the main database operations. The third one checks if the migration provides good results in terms of performance.

Tabela 3. Comparison of Related Works

Reference	**Supported Data Models**	**Workload Consideration**	**Post-Migration Scalability**
[Namdeo and Suman 2021]	Document-oriented (MongoDB)	-	High for historical data; Medium for real-time data
[Hamouda 2023]	Document-oriented	Data volume and complexity	Moderate
[Tandya and Azizah 2023]	Document-oriented (MongoDB)	Table usage frequency and accessed relationships	High
[Siregar and Azizah 2023]	Column-oriented (Cassandra)	-	High
[Eldrrat and Maatuk 2023a]	Graph	-	Moderate
[Karnitis and Arnicans 2015]	Document-oriented (Clusterpoint)	-	Moderate
[Yassine and Awad 2018]	Document-oriented (MongoDB)	-	High
[Alami et al. 2024]	Document-oriented (MongoDB)	-	High

As we see in Table 3, the migration from relational databases to NoSQL databases has been a hot topic in the literature in the last decade, with different strategies proposed for data conversion. However, the existing works present limitations that may compromise the effectiveness and applicability of the solutions in real-world Big Data contexts.

[Namdeo and Suman 2021] focus exclusively on converting MySQL to MongoDB, without generalization to other databases or more complex schemas. [Hamouda 2023] prioritizes migration time but does not evaluate performance under high loads or in a distributed scenario. It makes its post-performance migration not so good. Meanwhile, [Tandya and Azizah 2023] uses an algorithm based on foreign keys for data nesting or data referencing in MongoDB, being also limited to NoSQL document-oriented databases.

[Siregar and Azizah 2023] focus on obtaining a good post-migration performance, but it is also limited to a single NoSQL data model. [Eldrrat and Maatuk 2023a] is the

only approach that investigates the conversion of relational schemas to graphs but its performance post-migration is not so high. [Karnitis and Arnicans 2015] proposes a semi-automatic conversion method also focused on NoSQL document-oriented databases with moderate post-migration scalability in complex data representation scenarios.

Finally, the works of [Yassine and Awad 2018] and [Alami et al. 2024] also restricts their work to mapping MySQL to MongoDB, without evaluating other NoSQL data representation in terms of performance.

Given this scenario, we see that there is space for more robust approaches. First of all, we see a strong focus on migration strategies for document-oriented databases, while other ones are less considered, and no proposals for the NoSQL key-value database exist. We also noticed a lack of approaches that consider more than one NoSQL data model and database for hosting relational data. We argue that this is a very relevant issue in the context of polyglot persistence [Kiehn et al. 2022], where application conceptual data should be designed to be stored into multiple database models and technologies in order to maximize performance in terms of data storage and manipulation. In such a scenario, we have additional problems to deal with, like the potential of a given data entity or relationship be represented into more than one NoSQL data model for improving data availability, and how to establish relationships for data represented in heterogeneous data models and database technologies.

Another point is that the approaches emphasize conversion rules from the relational model to a NoSQL model based solely on the structure of the relational schema, i.e., they do not considered the expected application workload. As stated before, workload estimation could be a good guide to specify logical structures in the NoSQL data models to also maximize data manipulation performance for the most expected operations over the data. The conversion process could consider workload information to decide which conversion rules are the most suitable to be applied.

At last, we see a concern about post-migration performance evaluation in terms of scalability, with some good results mainly in terms of document-oriented databases. Unfortunately, there is a lack of similar evaluations for the other NoSQL data models.

6. Conclusion

This paper presents a SLR and an analysis of approaches for migrating relational databases to NoSQL databases. We argue that this is a relevant issue when an application intends to move these data (or part of them) to NoSQL technologies in order to improve some Big Data manipulation and analytics operations performance.

We see a continuous research on this subject in this decade, but, as stated in the previous section, there are not so many solutions for the migration of all NoSQL data models in the literature, except document-oriented databases. Workload consideration in the migration process is also a topic few explored.

Given this scenario, future studies should address the proposal of more robust approaches that consider multiple NoSQL data models at the same time as well as a deep analysis of workload information to efficiently distribute application data among the most suitable NoSQL technologies. Despite that, we realized no discussion about data integrity preservation in the selected studies, e.g., relational integrity constraints analysis

and consideration by the migration processes. This is also relevant in order to better guarantee data consistency.

Referências

Alami, A. E., Khourdifi, Y., Mouden, Z. A. E., Lahmer, M., and Hasnaoui, M. L. (2024). Migrating Relational Databases to NoSQL-Oriented Documents Using Object-Oriented Concepts. *International Journal of Intelligent Engineering and Systems*, 17(4):634–648.

de Lima, C. and Mello, R. S. (2016). On Proposing and Evaluating a NoSQL Document Database Logical Approach. *Int. J. Web Inf. Syst.*, 12(4):398–417.

Eldrrat, H. and Maatuk, A. (2023a). Data Migration from Conventional Databases into NoSQL: Methods and Techniques. In *III International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)*, pages 370–375. IEEE.

Eldrrat, H. A. and Maatuk, A. M. (2023b). A Novel Approach for Migrating Relational Databases into Graph Databases. In *International Conference on Emerging Smart Technologies and Applications (eSmarTA)*. IEEE.

Hamouda, S. (2023). Seamless Transition: Migrating from Relational Databases to Document-Oriented Databases. In *XII International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDA-ACS)*, pages 883–888. IEEE.

Han, J., Kamber, M., and Pei, J. (2013). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.

Karnitis, G. and Arnicans, G. (2015). Migration of Relational Database to Document-Oriented Database: Structure Denormalization and Data Transformation. In *VII International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN)*, pages 113–118. IEEE.

Kiehn, F., Schmidt, M., Glake, D., Panse, F., Wingerath, W., Wollmer, B., Poppinga, M., and Ritter, N. (2022). Polyglot Data Management: State of the Art & Open Challenges. *Proc. VLDB Endow.*, 15(12):3750–3753.

Kitchenham, B. (2004). Procedures for Performing Systematic Reviews. *Keele, UK, Keele University*, 33(2004):1–26.

Kleppmann, M. (2016). *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*. O'Reilly.

Namdeo, B. and Suman, U. (2021). A Model for Relational to NoSQL database Migration: Snapshot-Live Stream Db Migration Model. In *VII International Conference on Advanced Computing and Communication Systems (ICACCS)*, pages 199–2004. IEEE.

Sadalage, P. J. and Fowler, M. (2013). *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence*. Pearson Education.

Shah, S. I. H., Peristeras, V., and Magnisalis, I. (2021). Government Big Data Ecosystem: Definitions, Types of Data, Actors, and Roles and the Impact in Public Administrations. *ACM J. Data Inf. Qual.*, 13(2):8:1–8:25.

Siregar, R. A. S. and Azizah, F. N. (2023). Migration From Relational Database to Column-Oriented NoSQL Database. In *International Conference on Data and Software Engineering (ICoDSE)*, pages 174–179. IEEE.

Sun, Z. (2024). Big Data 4.0 = Meta4 (Big Data) = The Era of Big Intelligence. In *VII International Conference on Software Engineering and Information Management (ICSIM)*, pages 14–22. ACM.

Tandya, W. and Azizah, F. N. (2023). Migration of Relational Database to NoSQL Document-Oriented Database. In *International Conference on Data and Software Engineering (ICoDSE)*, pages 180–185. IEEE.

Wazlawick, R. S. (2009). *Metodologia de Pesquisa em Ciência da Computação*. Elsevier.

Yassine, F. and Awad, M. A. (2018). Migrating from SQL to NoSQL Database: Practices and Analysis. In *International Conference on Innovations in Information Technology (IIT)*, pages 58–62. IEEE.