NoSQL Databases Deployment in Various Application Domains: A Systematic Literature Review

Jelena Petrovački, Dejana Gladić, Sara Havzi, Darko Stefanović, Sonja Ristić
Faculty of Technical Sciences, Department of Industrial Engineering and Engineering Management
University of Novi Sad
Trg Dositeja Obradovića 6, 21000 Novi Sad
{jelenapetrovacki, dejanagladic, havzisara, darko.stefanovic, sdristic}@uns.ac.rs

Abstract. *The continuous growth of data volume to be stored and processed in the shortest possible time brings new requirements in terms of data storage, processing, analysis and visualization. Relational databases are not capable to store and process big data efficiently and NoSQL databases have emerged as an appropriate alternative. The range of available NoSQL systems is wide and the suitability of a given NoSQL system depends on the problem domain. In this paper, a systematic literature review of various NoSQL systems by application domains is performed in order to ease the selection of appropriate NoSQL system for a specific purpose.*

Keywords. NoSQL, Big Data, application domain, usage domain, document database, graph database, key-value database, column database, data analysis, Systematic Literature Review

1 Introduction

Big data as a concept is increasingly used in the business and educational spheres of modern life. The term big data encompasses a huge set of data collected from a variety of sources, from business and educational processes, through administration and governance to the social aspects of human life, such as social networks and various electronic communication channels.

Several different authors describe the big data concept through five essential characteristics designated as 5V:

- Volume – represents the amount of stored and processed data that can reach up to several terabytes and petabytes;
- Variety – means the format of data that is stored and divided into three main categories: unstructured, semi-structured and structured data;
- Velocity – represents the speed of data generation and processing and big data is suitable for processes that must be processed in limited real-time;
- Veracity – represents the quality and accuracy of stored data; and
- Value – the value of data that is stored and analyzed because it makes no sense to create and store data that will later serve no purpose (Faridoon & Imran, 2021).

Adequately processed and analysed data should influence more efficient and effective decision-making that leads to progress.

Big data sets raise various challenges, including methods to collect, store and process different types and structures of data. Huge volumes of data recorded in heterogeneous formats that have to be processed in limited real-time, are difficult to be handled by traditional relational database management systems (RDBMSs). Structured Query Language (SQL) is standardized data manipulation and query language of RDBMSs. Therefore, relational databases managed by RDBMSs are also known as SQL databases. For decades, SQL databases were considered as standard technology for persisting and managing large volumes of data. Big data sets bring the need for levels of availability beyond those supported by SQL databases and the challenges of scaling such databases horizontally led to the emergence of a new generation of databases grouped under the umbrella term NoSQL databases (Strauch & Kriha, 2011). NoSQL databases are a current approach to collecting and storing such large and distributed datasets (Philip Chen & Zhang, 2014), that allow the volume, variety and velocity characteristics of big data sets (Naik & Joshi, 2017) to be adequately managed.

There is a plethora of various NoSQL systems. They are based on different paradigms and aimed at storage of various data formats. Consequently, the particular suitability of a given NoSQL database system depends on the problem it is aimed to solve. In the paper we perform systematic literature review (SLR) on the deployment of NoSQL systems in various application domains. The main goals of presented SLR are: i) to understand the applicability of various NoSQL systems in different application domains, in order to ease the selection of appropriate NoSQL system for a specific purpose; ii) to identify challenges of various NoSQL systems and issues of their deployment in different domains; and iii) to discuss the potential research gaps. In order to accurately answer the given topic, the application of NoSQL databases in different domains, the research of the concept of big data and its prevalence in different sectors of the economy was first performed due to the close connection between these two concepts. Different authors mention the same economic activities as the domains of application, primarily big data, and in this paper, the ones that are most often mentioned in most papers are singled out in order to make a search query adequately. Economic activities that are analyzed are as follows: communications, media and entertainment, healthcare, education, industry, government, insurance, trade (retail and wholesale), transportation, and agriculture. Based on the listed economic activities, the search query described in Section 3 will be generated.

The reviewed papers most often mention the application of big data sets within the Internet of Things (IoT) system, which is implemented in various sectors, and it is very important to search for the application of NoSQL databases within such a created system.

In the second chapter, a theoretical overview of the concepts on which SLR on a given topic is based is described in detail. The third chapter includes a review of SLR planning, implementation, as well as analysis of selected papers through descriptive statistics analysis. The fourth chapter provides answers to research questions that are defined in the initial phase of the literature review. In the fifth chapter, the concluding opinions on this topic are given, as well as the directions for further research.

2 Theoretical foundations

The fact that most people think of tables and SQL as the first thought of the term "*database*" is not surprising. The reason for this is the relatively late emergence and beginning of the use of NoSQL databases that began to be used in the late 2000s. Since then, they have become more and more popular, and it seems that their usage will become more and more important in the future. The most significant difference with well-known relational databases is in the format of the stored data.

The creators of NoSQL databases emphasize their flexibility as a major advantage over traditional databases. Flexibility is reflected in a database schema that is not strictly defined as in a relational database, and data models may differ from each other. In addition to this advantage of NoSQL databases, many others stand out, among others:

- structured, semi-structured and unstructured data can be stored,
- the costs are much lower,
- ability to distribute large amounts of data stored in the cloud, via multiple servers,
- fast and efficient horizontal scaling,
- faster query execution,
- ease of integration of the persistence layer with the business logic layer of application and
- ease of usage.

Relational databases offer a very strict model of transaction control called ACID (Atomicity, Consistency, Isolation and Durability) that is not needed for many types of applications. A strict transactional model can reduce the performance of many systems (Stonebraker, 2010). Just as ACID is core feature of SQL databases, BASE (Basically Available, Soft state and Eventual consistency) is feature of NoSQL databases. NoSQL databases are designed with horizontal scalability and partition tolerance delivering increased availability and fault tolerance at a cost of temporary data inconsistency.

Four groups of NoSQL databases stand out: key-value databases, document oriented databases, column oriented databases and graph databases. They are quite different from each other and in the Table 1. are given an overview and general characteristics of each type of NoSQL databases.

3 Methodology

In the process of conducting the SLR, presented here, the approach of Barbara Kitchenham is followed, according to which all activities can be divided into three phases: planning the review, conducting the review and reporting the review (Kitchenham, 2004).

3.1 Planning the literature review

At the beginning of this phase of the research, a search of already existing literature reviews dealing with this topic is approached (Kitchenham, 2004). An SLR on a similar topic was found, but the way to review the literature is different. The query set in (Faridoon & Imran, 2021) differs from the query set in this paper and different index databases are searched. The focus of the search in (Faridoon & Imran, 2021) was on big data data storage tools based on NoSQL databases, their advantages and disadvantages.

Table 1. Overview of NoSQL database types

DOCUMENT DATABASE	Data model	JSON, BSON or XML document
	About DB	The data is stored in the form of documents. Each document contains fields and their values. Documents can be nested with each other.
	Use cases	Commerce platforms, trading platforms, content management, web and mobile application development across industries
	Advantages	Flexibility - documents within a collection do not have to be structured the same (consistency is not required). The format in which the data within the database document is written is similar to object notation, so the use of such structured data is easier to create applications.
	Disadvantages	Nesting documents can lead to a complex system and compromise performance.
	Popular databases	MongoDB, SimpleDB, CouchDB, Couchbase
KEY-VALUE DATABASE	Data model	Table with two columns: key and value
	About DB	The simplest form of NoSQL database that resembles a relational table with two columns - key and value. The keys are unique identifiers for the values. The values can be another key-value pair in which case the structure of the database grows more complex.
	Use cases	Application logs, shopping carts, user preferences, user profiles
	Advantages	Easy to use.
	Disadvantages	There is no defined query language, so search is only possible based on the key.
	Popular databases	Aerospike, Redis, Riak
WIDE-COLUMN DATABASE	Data model	A set of columns
	About DB	The table is organized as a set of columns - emphasis on columns, unlike relational databases where the emphasis is on rows. Multiple columns form a column family.
	Use cases	Analytics, internet search, large-scale web applications
	Advantages	They are suitable for applying aggregate functions to data stored within the same column. Also, they are very useful for data warehouses, or when there is a need for handling intensive querying.
	Disadvantages	These databases are not very efficient with online transactional processing.
	Popular databases	BigTable, HBase, Cassandra, Hypertable
GRAPH DATABASE	Data model	Graph structure (nodes and egdes)
	About DB	Data is stored in nodes and edges. Elements are represented as nodes, and connections between elements as edges. They are often used in addition to existing relational databases.
	Use cases	Fraud detection, social networks, knowledge graphs, geospatial application, recommendation engines
	Advantages	Suitable for highlighting the relationship between elements (without the need to join tables as in SQL)
	Disadvantages	Scalability -they are hard to be scaled across a number of servers.
	Popular databases	Neo4j, InfiniteGraph, HyperGraphDB

The aim of presented SLR is precisely determined, and it includes the identification of a data storage model that will in the best way that is in terms of the highest number of performances, meet the needs of the application for which the database is being created.

In order to meet the aim of this research, adequate research questions have been defined. Research questions include the following:

RQ1: Which NoSQL systems are suitable for storing data for a particular application domain?

RQ1.1: What are the properties and data types of data stored in these NoSQL databases?

RQ2: What are the advantages of NoSQL systems?

RQ2.1: What are the advantages of a particular NoSQL system compared to other NoSQL systems?

RQ2.2: What are the advantages of NoSQL systems compared to RDBMSs?

RQ3: What are the challenges of using the NoSQL systems?

In order to answer the previously defined questions, the index databases Scopus and Web of Science (WoS) were searched.

When searching for the listed index databases, the keywords *NoSQL database*, *usage*, *utilization*, *implementation* and *domain* were used firstly. A query that represents a combination of these keywords was generated as follows:

NoSQL AND domain AND (usage OR utilization OR implementation).

After analyzing the papers obtained by applying the presented query, it was concluded that the found papers will not provide an answer to the defined research questions, and that the research goal will not be fulfilled.

The main shortcoming of the previously defined query was noticed, which are imprecisely defined desired domains for the usage of NoSQL databases. Therefore, it is decided to explicitly enumerate the most important economic activities, as it is stated in Section 1. The modified query is:

NoSQL AND (communication OR media OR entertainment OR healthcare OR education OR industry OR insurance OR trade OR transportation OR IoT OR agriculture OR government)
AND PUBYEAR > 2016
AND (LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ar"))
AND (LIMIT-TO (LANGUAGE , "English")).

Some of the defined exclusion criteria were applied through the presented query. Here is an overview of all the identified exclusion criteria:

1. Non-English papers are excluded.

2. Papers other than "conference paper" or "journal article" should be removed.
3. It is necessary to remove papers that appear in the search results of both index databases.
4. If one author has more than one paper regarding the same approach, only one paper should be included in the review.
5. Papers that deal only with the analysis and implementation of some of the SQL databases or some auxiliary frameworks without the analysis of NoSQL databases and tools are not considered.
6. Papers that theoretically describe different NoSQL systems without analyzing system performance over specific data from any of the identified usage domains are excluded.

Identified inclusion criteria are:

1. Papers that are accessible electronically.
2. Papers that are published since 2017.
3. Papers that test the performance of various NoSQL system on data collected from a real system to propose the best NoSQL database.
4. Papers describing the implementation of a software solution from a real system, with a focus on NoSQL databases and tools.
5. Papers comparing NoSQL and SQL databases over some specific data from different domains.

For further analysis and presentation of descriptive statistics, data such as: publication year, publication type, authors, keywords, NoSQL system and usage domain mentioned in the paper, advantages and challenges that are solved by using NoSQL systems and shortcomings and challenges that NoSQL systems cannot respond to were extracted from each paper.

3.2 Conducting the literature review

The initial search of both selected index databases resulted in a total of 845 papers that meet the defined criteria. Then the papers were selected based on the title of the paper, as well as based on the abstract, and the numbers of selected papers are shown in Table 2.

Table 2. Number of selected papers by search phases

	WoS	SCOPUS
Initial search	317	528
Number of selected papers based on title	76	42
Number of selected papers based on abstract	59	29
Number of selected works based on their content	32	11
In total:	43	

The papers accepted in the previous phases were analyzed in detail and 43 papers were selected on which descriptive statistics will be based and on the basis of which the research questions will be answered.

Through all phases, the previously mentioned inclusion and exclusion criteria were applied.

3.3 Results of review

In this section descriptive statistics based on relevant information extracted from 43 selected papers is presented. For a more compact presentation of the contents of Table 3. and Table 4., ordinal numbers of papers are used, and a list of associated ordinal numbers is given in the table in the Appendix.

The origin of selected papers is shown in Table 3. It can be seen that 69.8% of papers were published at various conferences, while the rest of the papers (30.2%) were published in scientific journals.

Table 3. Source type of selected papers

Source type	Papers	Percentage
Conference proceedings	1, 2, 3, 8, 9, 10, 11, 12, 13, 15, 16, 17, 19, 20, 21, 22, 23, 26, 27, 29, 31, 32, 34, 35, 36, 37, 39, 40, 41, 42	69,8 %
Journal article	4, 5, 6, 7, 14, 18, 24, 25, 28, 30, 33, 38, 43	30,2 %

In Fig 1. is given the number of published works by year. Based on the graph shown in Fig 1, we cannot observe a trend of publishing papers on this topic in a certain period of time.

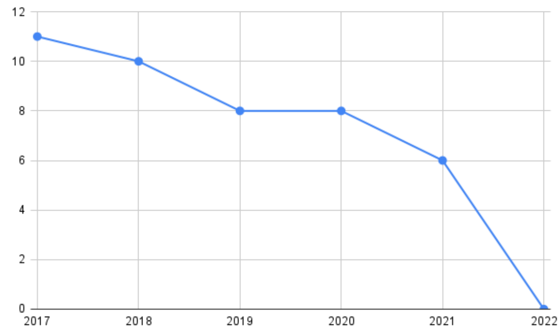

Figure 1. Number of selected studies per year

A large part of the selected papers (44.2%) refers to the comparative analysis of different NoSQL databases and testing their performance, while in a smaller number of papers (16.3%) the authors compare the performance of NoSQL and relational databases. As can be seen in Table 4, other papers (39.5%) deal with the implementation of a specific NoSQL database that they consider to be the most suitable for the needs of the application for which the database is being created.

Table 4. Division of papers based on content

	Papers	Percentage
Comparison of SQL and NoSQL databases	5, 8, 19, 20, 24, 33, 36	16,3 %
Concrete application of a particular NoSQL database	2, 9, 13, 15, 18, 23, 27, 29, 32, 34, 37, 38, 39, 40, 41, 42, 43	39,5 %
Comparison of different NoSQL databases	1, 3, 4, 6, 7, 10, 11, 12, 14, 16, 17, 21, 22, 25, 26, 28, 30, 31, 38	44,2 %

The analysis of selected papers also recorded the key words that most often appear in papers that correspond to the chosen topic. As is shown in Fig 2., the largest share of papers is based on the key words: *NoSQL Database* and *Big Data*. In addition to them, various databases (*MongoDB*, *Cassandra* and *Couchbase*) appear in the identified keywords. When it comes to application domains, the key word "*healthcare*" stands out, which leads to the assumption that most of the papers will be based on the use of NoSQL databases within medicine. The appearance of keywords "*IoT*" is evident, which we can relate to the moment when most of the data to be stored within the database is actually generated in this way, and that the term Internet of Things can be associated with all identified application domains.

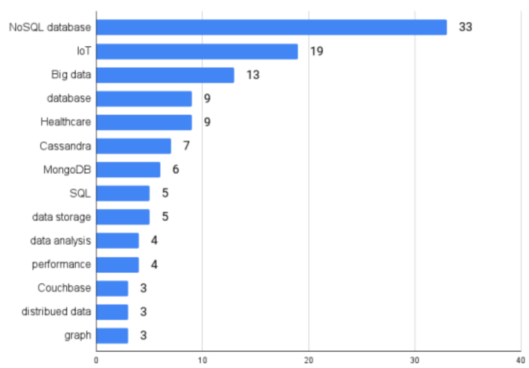

Figure 2. Keywords from selected papers

Different data structures will give different results when it comes to the performance of the database within which the data is stored. It can be concluded that for appropriate data structures some databases are suitable to a greater extent and some to a lesser extent. Assuming that relatively similar data formats are used within one domain of the application, it is justified to generalize the use of the database to the categories shown in the graph in Fig 3. As could be assumed when analyzing keywords, the number of papers based on the use of NoSQL databases for health purposes is the largest. Also, it cannot be claimed that NoSQL databases are not used or suitable for storing data from certain categories (education, insurance), but only that the search results did not find adequate work related to these application domains.

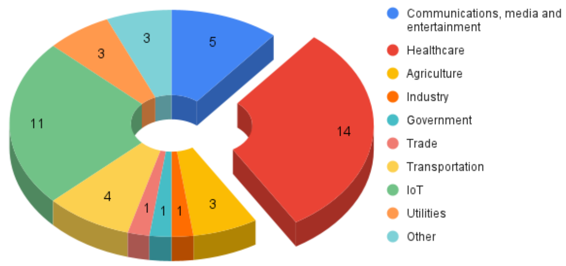

Figure 3. The number of usage domain appearances in selected papers

A detailed analysis of the papers recorded the NoSQL databases that were discussed in the largest number of papers. From the graph in Fig. 4 it can be seen that in the papers the most analyzed document databases, given that MongoDB is mentioned in 21 selected papers, which is almost half of the total number of selected papers. Cassandra, a representative of column databases, and Neo4j, which belongs to graph databases, also appear in most papers.

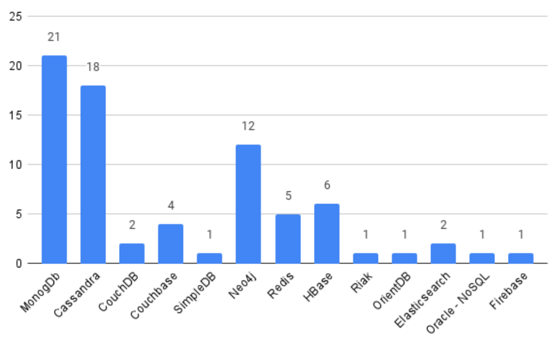

Figure 4. The number of NoSQL databases appearances in selected papers

The latest descriptive statistics shown in the graph in Fig. 5 represent in which application domain the most common type of NoSQL database is. In most cases, there is no significant difference for any of the types of NoSQL databases, which calls into question whether the analysis of types of NoSQL databases by application domains makes sense. The answer to this question will be presented through the analysis of selected papers, where the content of the papers will be described in more detail.

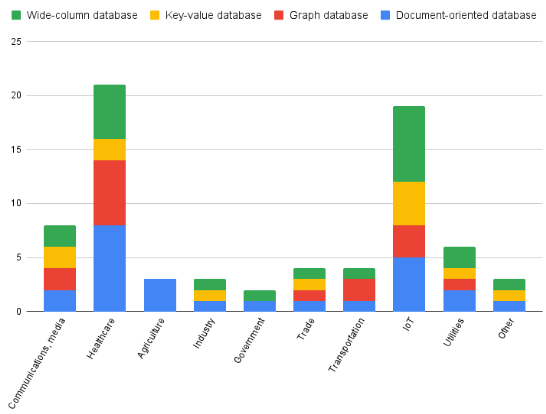

Figure 5. NoSQL types per usage domains

4 Discussion

In this chapter, previously asked research questions will be answered.

In order to give an answer to RQ1, the graphical representation from Fig. 5 is important. Given that in most cases it cannot be observed that some type of NoSQL database is predominantly used within some of the categories, it cannot be said with certainty that the domain of application is crucial when choosing a database.

The answer to RQ1.1 would be more precise, because at the level of granularity on the basis of which this question is defined, it is possible to draw conclusions – for which types of data which database is suitable. Then it could be researched which types of data are most represented in which domain of application.

In paper named “*Which NoSQL database for IoT applications?*” (Amghar et al., 2018), the authors review the literature based on which they conclude that most IoT applications use NoSQL databases for their data storage needs. They ask which of the many available NoSQL databases is the best choice and come to the conclusion that it depends on the needs of the domain. They identify 7 needs of IoT applications: data heterogeneity, semantic interoperability, scalability, real time processing, security, spatial data handling and data aggregation. The characteristics of the selected NoSQL databases are analyzed and it is concluded which database would correspond to which type of data. The results of the research mostly emphasize that in the domains of application that involve a large amount of spatial data, there is an obvious need to use NoSQL graph databases, given the importance of entity connectivity.

The domain of application described in most of the papers is health. Observing the breadth of this domain, it is expected that data formats will differ depending on the specific needs of the application for which the database is being created. It is not uncommon for different applications to need to store different data in different databases, in order to ensure the best performance of the application.

This situation has been described in (Celesti, et al., 2020b) and (Cerbulescu, C. & Cerbulescu, C. M., 2017) and (Kundu et al., 2021) describes the different needs of an application belonging to the same domain as applications on which the first two papers are based.

(Celesti, 2020b) state that in the tele-rehabilitation system in which patients send data via devices from their home, and doctors review it and return therapy, two groups of data are separated. The first group includes personal data on doctors and patients, and since this is data that is rarely changed and structured, the best option for storing this group of data is a relational database. The second group consists of patient-generated data. The volume of this data is much larger and most often this data is not structured, so in this situation the use of NoSQL database is recommended.

(Cerbulescu, C. & Cerbulescu, C. M., 2017) in their study emphasize the importance of rapid processing of critical data. Their approach is based on extracting a small amount of key data that is considered critical and this data is recorded within a relational database. This database has good performance when it comes to small amounts of data. In addition, in order for the SQL database to be as fast as possible, they set the expiration date on the data entered in it, and unnecessary data is deleted after the expiration of that period. All other data, which is not identified as critical, is stored within the NoSQL database. The previous analysis also answered research question RQ2.2 because there is a clear difference in the usage of relational and NoSQL databases. The authors cite these differences as the main reason for choosing one or another type of database.

Through the study (Kundu et al., 2021) we can confirm the statement that within one domain of application there may be needs for different databases. In order to build a system that shows "trust" to the doctor, the Neo4J database graph was used. For each new registered doctor, a new node is added that contains information about the doctor. A new node is also created for each patient who schedules an examination, and a new edge is created for each scheduled appointment with a doctor. Each scheduled appointment with a doctor increases the doctor's score (if a patient returns to the same doctor several times, the number of points added to the previous score increases). Applications based on the principle of "trust" are inherent in other domains of application, not only health.

In order to answer the research question RQ2.1, the selected papers must be analyzed in detail. The choice of database also depends on the domain in which it is applied, but mostly on the data structure that is stored there and the needs of applications for certain performance tools. In the paper (Chen & Lee, 2019), the authors describe the method of selecting NoSQL systems depending on the purpose and needs of applications. The first case study is a shopping site that is slow because it uses a relational database to store large amounts of data. The solution to the problem is to use the NoSQL database, which will enable the merging of various relational database tables in advance and shorten the time required to execute select queries. A wide column NoSQL database was chosen because access to the database often requires searching for data by a specific feature. The choice was reduced to Apache Cassandra and Apache HBase. HBase spent much less time accessing data than Apache Cassandra so HBase was chosen as adequate for implementation. Another study chose the document store NoSQL database, MongoDB, for collecting and storing files with multimedia material for the needs of the news agency, because it is the most used in that category. In (Pramukantoro et al., 2017), the authors also select the MongoDB database for storing data such as surveillance camera images, the degree of CO gas presence, or temperature and fluidity data from different IoT sensors. The main reason is that it contains a GridFS format that allows you to store large amounts of multimedia content. The authors of the paper (Sharma & Kaur, 2019) want to determine the areas in the city where there is heavy traffic, at what time of day and how much these parts violate traffic safety. The data to be stored are geolocation points (latitude and longitude) and time data. They choose the Apache Cassandra Cluster tool whose interface is suitable for data modeling and query execution. Also, using that tool can be created a cluster of many nodes to which data can be distributed. The node is easy to add and remove because they are all independent of each other. If one of the nodes fails, the others can run smoothly, but the malfunction may increase the time required to execute the query. In each selected paper, the advantages and reasons for choosing a certain NoSQL tool are defined, but in response to the question RQ2.1. some of them are described.

The response to RQ3 is not entirely satisfactory. In order to obtain more precise information about the challenges when using NoSQL databases, it is necessary to define a different search query for index databases. From the accepted papers, it was concluded that the biggest challenge is data security. As an example, the authors of the paper (J. Aqel et al., 2019) describe weak mechanisms for data encryption. All data in MongoDB is stored as plain text and there is no encryption mechanism to encrypt data files and all passwords in Cassandra are encrypted using of MD5 hash function so that the passwords are very weak.

5 Conclusion

Based on the conducted systematic review of the literature, it can be concluded that there is no rule according to which an adequate NoSQL database would be selected based on the domain of the application for which the database is being created.

The categorization of the usage of NoSQL databases by application domains is not the most precise, but on the basis of selected papers, it is possible to obtain an image of which data features most often occur in which application domains.

In most cases, it is necessary to store data in several different types of databases for the needs of applications, because different data structures are used within one application. Although the focus of this paper is based on NoSQL databases, the usage of relational databases, which are still number one when it comes to transactions and static data structures, cannot be completely ignored.

Further research directions would include a more detailed search of the challenges that arise during the implementation of NoSQL databases, then the classification of identified challenges into those whose negative effects can be mitigated, as well as those whose negative effects cannot be avoided.

References

Ai-Sakran, A., Qattous, H., & Hijjawi, M. (2018). A proposed performance evaluation of NoSQL databases in the field of IoT. *2018 8th International Conference on Computer Science and Information Technology, CSIT 2018*, 32–37. https://doi.org/10.1109/CSIT.2018.8486199

Al Sadi, M. B., Wimmer, H., Chen, L., & Wang, K. (2017). Improving the efficiency of big forensic data analysis using NoSQL. *International Conference on Mobile Multimedia Communications (MobiMedia), 2017-July*, 240–248. https://doi.org/10.475/eai.13-7-2017.2270344

Almeida, A., Oliveira, F., Lebre, R., & Costa, C. (2020). NoSQL distributed database for DICOM objects. *Proceedings - 2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2020*, 1882–1885. https://doi.org/10.1109/BIBM49941.2020.9313430

Amghar, S., Cherdal, S., & Mouline, S. (2018). Which NoSQL database for IoT applications? *2018 International Conference on Selected Topics in Mobile and Wireless Networking, MoWNeT 2018*, 131–137. https://doi.org/10.1109/MoWNet.2018.8428922

Asiminidis, C., Kokkonis, G., & Kontogiannis, S. (2019). Database Systems Performance Evaluation for IoT Applications. *SSRN Electronic Journal*, *November 2019*. https://doi.org/10.2139/ssrn.3360886

Braulio, L. D. C., Moreno, E. D., De MacEdo, D. D. J., Kreutz, D., & Dantas, M. A. R. (2018). Towards a Hybrid Storage Architecture for IoT. *Proceedings - IEEE Symposium on Computers and Communications*, *2018-June*(5), 470–473. https://doi.org/10.1109/ISCC.2018.8538474

Celesti, A., Buzachis, A., Galletta, A., Fazio, M., & Villari, M. (2018). A NoSQL Graph Approach to Manage IoTaaS in Cloud/Edge Environments. *Proceedings - 2018 IEEE 6th International Conference on Future Internet of Things and Cloud, FiCloud 2018*, 407–412. https://doi.org/10.1109/FiCloud.2018.00065

Celesti, A., Celesti, F., Galletta, A., Fazio, M., & Villari, M. (2020a). Improving Machine Learning Algorithm Processing Time in Tele-Rehabilization through a NoSQL Graph Database Approach: A Preliminary Study. *Proceedings - IEEE Symposium on Computers and Communications*, *2020-July*. https://doi.org/10.1109/ISCC50000.2020.9219700

Celesti, A., Lay-Ekuakille, A., Wan, J., Fazio, M., Celesti, F., Romano, A., Bramanti, P., & Villari, M. (2020b). Information management in IoT cloud-based tele-rehabilitation as a service for smart cities: Comparison of NoSQL approaches. *Measurement: Journal of the International Measurement Confederation*, *151*, 107218. https://doi.org/10.1016/j.measurement.2019.107218

Cerbulescu, C., & Cerbulescu, C. M. (2017). Proposed Architecture to manage critical states predictions in IoT applications. *ACM International Conference Proceeding Series*, *Part F1309*, 1–4. https://doi.org/10.1145/3136273.3136279

Chauhan, N., & Malik, M. (2021). Home Automation With NoSQL and Node-RED Through Message Queuing Telemetry Transport. *Advances in Intelligent Systems and Computing*, *1172*(April), 427–434. https://doi.org/10.1007/978-981-15-5566-4_37

Chen, J. K., & Lee, W. Z. (2019). An introduction of NoSQL databases based on their categories and application industries. *Algorithms*, *12*(5), 1–16. https://doi.org/10.3390/a12050106

Cordeiro, J. R., & Postolache, O. (2018). Big Data Storage for a Health Predictive System. *2018 International Symposium in Sensing and Instrumentation in IoT Era, ISSI 2018*, 1–6. https://doi.org/10.1109/ISSI.2018.8538285

Dias, L. B., Holanda, M., Huacarpuma, R. C., & De Sousa, R. T. (2018). NoSQL database performance tuning for IoT data cassandra case study. *IoTBDS 2018 - Proceedings of the 3rd International Conference on Internet of Things, Big Data and Security*, *2018-March*(IoTBDS 2018), 277–284. https://doi.org/10.5220/0006782702770284

Faridoon A., & Imran A. (2021). Big data storage tools using NoSQL databases and their applications in various domains: a systematic review. *Computing and Informatics*, *40*, 489–521. https://doi.org/10.31577/cai_2021_3_489

Ferencz, K., & Domokos, J. (2019). IoT Sensor data acquisition and storage system using raspberry Pi and apache cassandra. *CANDO-EPE 2018 - Proceedings IEEE International Conference and Workshop in Obuda on Electrical and Power Engineering*, 143–146. https://doi.org/10.1109/CANDO-EPE.2018.8601139

Gamal, A., Barakat, S., & Rezk, A. (2021). Integrated Document-based Electronic Health Records Persistence Framework. *International Journal of Advanced Computer Science and Applications*, *12*(10), 147–155. https://doi.org/10.14569/IJACSA.2021.0121017

Ghiwari, S., Sambrekar, K., & Rajpurohit, V. S. (2018). Hierarchical Storage for Agro Informatics System Using NoSQL Technology. *2017 International Conference on Computing, Communication, Control and Automation, ICCUBEA 2017*, 1–5. https://doi.org/10.1109/ICCUBEA.2017.8463693

Hak, F., Guimarães, T., Abelha, A., & Santos, M. (2020). An Exploratory Study of a NoSQL Database for a Clinical Data Repository. *Advances in Intelligent Systems and Computing*, *1161 AISC*, 476–483. https://doi.org/10.1007/978-3-030-45697-9_46

Hendawi, A., Gupta, J., Liu, J., Teredesai, A., Ramakrishnan, N., Shah, M., El-Sappagh, S., Kwak, K. S., & Ali, M. (2019). Benchmarking large-scale data management for Internet of Things. *Journal of Supercomputing*, *75*(12), 8207–8230. https://doi.org/10.1007/s11227-019-02984-6

J. Aqel, M., Al-Sakran, A., & Hunaity, M. (2019). A Comparative Study Of NoSQL Databases. *Bioscience Biotechnology Research Communications*, *12*(1), 17–26. https://doi.org/10.21786/bbrc/12.1/3

Khan, W., Ahmad, W., Luo, B., & Ahmed, E. (2019). SQL database with physical database tuning technique and NoSQL graph database comparisons. *Proceedings of 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2019*, *Itnec*, 110–116. https://doi.org/10.1109/ITNEC.2019.8729264

Khan, W., ahmed, E., & Shahzad, W. (2017).

Predictive Performance Comparison Analysis of Relational & NoSQL Graph Databases. *International Journal of Advanced Computer Science and Applications*, *8*(5). https://doi.org/10.14569/ijacsa.2017.080564

Kitchenham, B. (2004). Procedures for Performing Systematic Reviews. *Keele University Technical Report*, 33(2004), 1-26. https://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf

Klein, J., Gorton, I., Ernst, N., Donohoe, P., Pham, K., & Matser, C. (2015). Performance evaluation of NoSQL databases: A case study. *PABS 2015 - Proceedings of the 1st ACM/SPEC International Workshop on Performance Analysis of Big Data Systems*, *February*, 5–10. https://doi.org/10.1145/2694730.2694731

Küçükkeçeci, C., & Yazıcı, A. (2018). Big Data Model Simulation on a Graph Database for Surveillance in Wireless Multimedia Sensor Networks. *Big Data Research*, *11*(September), 33–43. https://doi.org/10.1016/j.bdr.2017.09.003

Kundu, G., Mukherjee, N., & Mondal, S. (2021). Building a Graph Database for Storing Heterogeneous Healthcare Data. *Smart Innovation, Systems and Technologies*, *196*, 193–201. https://doi.org/10.1007/978-981-15-7062-9_19

Magdum, J., & Barhate, R. (2018). Performance Analysis of DML Operations on NoSQL Databases for Streaming Data. *Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018*, 1–6. https://doi.org/10.1109/ICCUBEA.2018.8697482

Mahmood, K., Orsborn, K., & Risch, T. (2019). Comparison of NoSQL datastores for large scale data stream log analytics. *Proceedings - 2019 IEEE International Conference on Smart Computing, SMARTCOMP 2019*, 478–480. https://doi.org/10.1109/SMARTCOMP.2019.00093

Naik, K., & Joshi, A. (2017). Role of Big Data in various sectors. *Proceedings of the International Conference on IoT in Social, Mobile, Analytics and Cloud, I-SMAC 2017*, 117–122. https://doi.org/10.1109/I-SMAC.2017.8058321

Nassif, E. H., Hicham, H., Yaagoubi, R., & Badir, H. (2020). Assessing NoSql Approaches for Spatial Big Data Management. *Lecture Notes in Networks and Systems*, *92*, 49–58. https://doi.org/10.1007/978-3-030-33103-0_6

Philip Chen, C. L., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, *275*, 314–347. https://doi.org/10.1016/j.ins.2014.01.015

Pramukantoro, E. S., Yahya, W., Arganata, G., Bhawiyuga, A., & Basuki, A. (2017). Topic Based IoT Data Storage Framework For Heterogeneous Sensor Data. *2017 11th International Conference on Telecommunication Systems Services and Applications (TSSA)*, 1-4. https://doi.org/ 10.1109/TSSA.2017.8272895

Pramukantoro, E. S., Primanita Kartikasari, D., & Siregar, R. A. (2019). Performance evaluation of MongoDB, cassandra, and HBase for heterogenous IoT data storage. *Proceedings of ICAITI 2019 - 2nd International Conference on Applied Information Technology and Innovation: Exploring the Future Technology of Applied Information Technology and Innovation*, 203–206. https://doi.org/10.1109/ICAITI48442.2019.8982159

Rasheed, Y., Qutqut, M. H., & Almasalha, F. (2019). Overview of the Current Status of NoSQL Database. *IJCSNS International Journal of Computer Science and Network Security*, *19*(4), 47-53. http://paper.ijcsns.org/07_book/201904/20190406.pdf

Reichardt, M., Gundall, M., & Schotten, H. D. (2021). Benchmarking the Operation Times of NoSQL and MySQL Databases for Python Clients. *IECON Proceedings (Industrial Electronics Conference)*, *2021-Octob*(October). https://doi.org/10.1109/IECON48115.2021.9589382

Sharma, M., & Kaur, G. (2019). Providing Low Cost IoT Enabled Security Solutions for High Intensity Regions. *2019 12th International Conference on Contemporary Computing, IC3 2019*, 1–7. https://doi.org/10.1109/IC3.2019.8844949

Shen, Y. (2021). Construction of a Wireless Sensing Network System for Leisure Agriculture for Cloud-Based Agricultural Internet of Things. *Journal of Sensors*, *2021*. https://doi.org/10.1155/2021/3021771

Silver, N., Ettinger, K., Kolden, J., Malgeri, M., & Weaver, E. (2016). "The suitcase": Using C4, semantics and NoSQL for managing motion picture data. *SMPTE 2016 - Annual Technical Conference and Exhibition*, *2016-Janua*(June), 1–15. https://doi.org/10.5594/m001715

Stonebraker, M. (2010). SQL databases v. NoSQL databases. *Communications of the ACM*, *53*(4), 10–11. https://doi.org/10.1145/1721654.1721659

Strauch, C. (2011). NOSQL databases. In *Advanced Information and Knowledge Processing*. Lecture Notes, Stuttgart Media University, vol. 20. https://doi.org/10.1007/978-3-030-04603-3_8

Subahi, A. F. (2019). Edge-Based IoT Medical Record System: Requirements, Recommendations and Conceptual Design. *IEEE Access*, *7*, 94150–94159. https://doi.org/10.1109/ACCESS.2019.2927958

Tokognon, A. C., Gao, B., Tian, G. Y., & Yan, Y. (2017). Structural Health Monitoring Framework Based on Internet of Things: A Survey. *IEEE Internet of Things Journal*, *4*(3), 619–635. https://doi.org/10.1109/JIOT.2017.2664072

Vanathi, R., & Shaik Abdul Khadir, A. (2017). A Robust Architectural Framework for Big Data Stream Computing in Personal Healthcare Real Time Analytics. *Proceedings - 2nd World Congress on Computing and Communication Technologies, WCCCT 2017*, 97–104. https://doi.org/10.1109/WCCCT.2016.32

Vela, B., Cavero, J. M., Caceres, P., Sierra, A., & Cuesta, C. E. (2018). Defining a NoSQL Document Database of Accessible Transport Routes. *Proceedings - 2017 IEEE International Conference on Internet of Things, IEEE Green Computing and Communications, IEEE Cyber, Physical and Social Computing, IEEE Smart Data, IThings-GreenCom-CPSCom-SmartData 2017*, *2018-Janua*, 1125–1129. https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2017.172

Venkatraman, S., Overmars, A., Fahd, K., Parvin, S., & Kaspi, S. (2020). Security Challenges for Big Data and IoT. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3378904.3378907

Wang, S., Zhu, Y., Wang, F., & Qian, Z. (2017). The selection of NoSQL database in E-government system of China meteorological administration. *2017 IEEE 2nd International Conference on Big Data Analysis, ICBDA 2017*, 105–109. https://doi.org/10.1109/ICBDA.2017.8078786

Wu, S., Bao, L., Zhu, Z., Yi, F., & Chen, W. (2018). Storage and retrieval of massive heterogeneous IoT data based on hybrid storage. *ICNC-FSKD 2017 - 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, 2982–2987. https://doi.org/10.1109/FSKD.2017.8393258

Appendix

No	Reference	No	Reference
1	(Amghar et al., 2018)	25	(J. Aqel et al., 2019)
2	(Ghiwari et al., 2018)	26	(Nassif et al., 2020)
3	(Wang et al., 2017)	27	(Almeida et al., 2020)
4	(Celesti, et al., 2020b)	28	(Küçükkeçeci & Yazıcı, 2018)
5	(Subahi, 2019)	29	(Vela et al., 2018)
6	(Hendawi et al., 2019)	30	(Gamal et al., 2021)
7	(Rasheed et al., 2019)	31	(Magdum & Barhate, 2018)
8	(Celesti et al., 2018)	32	(Ferencz & Domokos, 2019)
9	(Vanathi & Shaik Abdul Khadir, 2017)	33	(Asiminidis et al., 2019)
10	(Braulio et al., 2018)	34	(Dias et al., 2018)
11	(Pramukantoro et al., 2019)	35	(Venkatraman et al., 2020)
12	(Mahmood et al., 2019)	36	(Cerbulescu, C. & Cerbulescu, C. M., 2017)
13	(Celesti, et al., 2020a)	37	(Brown, 2020)
14	(Chen & Lee, 2019)	38	(Tokognon et al., 2017)
15	(Al Sadi et al., 2017)	39	(Sharma & Kaur, 2019)
16	(Ai-Sakran et al., 2018)	40	(Hak et al., 2020)
17	(Klein et al., 2015)	41	(Silver et al., 2016)
18	(Shen, 2021)	42	(Kundu et al., 2021)
19	(Reichardt et al., 2021)	43	(Chauhan & Malik, 2021)
20	(Khan et al., 2019)	44	(Faridoon & Imran, 2021)
21	(Cordeiro & Postolache, 2018)	45	(Philip Chen & Zhang, 2014)
22	(Wu et al., 2018)	46	(Naik & Joshi, 2017)
23	(Pramukantoro et al., 2017)	47	(Stonebraker, 2010)
24	(Khan et al., 2017)	48	(Kitchenham, 2004)