WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

An Art of Handling NoSQL Databases with Respect to Big Data

Shaida Begum[a]

[a]*CSE department, PDIT, Hospet, India*

DOI: https://doi.org/10.55248/gengpi.2022.3.10.55

ABSTRACT

The cyber-physical world's computer and communication capacity is rapidly growing. To handle these tasks, a lot of data is consequently produced. The four main problems with big data are volume, diversity, velocity, and authenticity. Hadoop is one example of a storage-based data processing system that controls volume and diversity. However, an unduly complex method is necessary to process such a large number of data quickly and accurately.Structured data cannot be effectively managed using typical relational databases, which are extensively used in information management systems. NoSQL, a non-relational database management system for unstructured data, is a novel technology that was developed to handle unstructured data. Scalability is what drives NoSQL the most. One application of big data is in the field of medicine. Every person's data collection yields a significant amount of information quickly. Early detection, diagnosis, and the recommendation of efficient medications can all be achieved through the analysis of stored healthcare data through querying. The application must explicitly encrypt any sensitive data before writing it to the database in order to offer security. Additionally, joining operations are avoided while querying the NoSQL database MongoDB, resulting in effective data retrieval. NoSQL, MongoDB, Big Data, Security, Healthcare Record.

Keywords:Relational database, Non-relational database, NoSQL, MongoDB, Big Data, Security, Healthcare Record

1. Introduction

Big Data refers to data sets that are challenging for traditional technologies and tools, including relational databases and data processing software, to manage, process, or analyze due to their size (volume), complexity (variability), and rate of increase (velocity). Big data is defined as data that is between 30 and 50 terabytes (1012) and multiple petabytes (1015) in size. The unstructured nature of a large portion of the data produced by contemporary technologies, such as that from web logs, Internet transactions, Healthcare, sensors embedded in devices, Internet searches, social networks like Facebook, portable computers, smart phones and other cell phones, GPS devices, and call center records, is what primarily contributes to the complexity of big data. Big data may be successfully and efficiently collected, processed, and analyzed to get deeper insight and help with decision-making processes such as detecting patterns and significance. As a result, businesses are better equipped to comprehend their industry, market, customers, products, rivals, etc. One. NOSQL NoSQL is a non-relational database management system that is fundamentally different from conventional relational database management systems. It is intended for data stores with extremely high data storage requirements (for example Google or Facebook which collects terabits of data every day for their users). This method of data storing typically scales horizontally, avoids join procedures, and may not require a well-fixed structure. NoSQL databases do not often employ SQL for data handling and are not primarily built on tables. Numerous distinct database technologies are used in noSQL. They fall within the categories of wide-column databases, Graph Databases, Document Databases, and Key-Value Databases. Among NoSQL databases, key-value stores are the most basic. Every object in the database is kept as a name for an attribute (or "key") and its associated value. Using a key, the user can request data. Riak and Voldemort are two instances of key-value stores. Each key in a document database is paired with a complicated data structure called a document. Documents may contain nested documents or a wide variety of key-value pairs. Wide-column stores, like Cassandra and HBase, store data in columns rather than rows and are tailored for queries across big datasets. For data whose relationships can be effectively represented as a graph, graph store databases are created. Data of this type could include network topologies, social connections, and connections to public transportation. HyperGraphDB and Neo4J are two graph stores. The following is an organized list of the paper's primary

* *Corresponding author.* Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: shahida@pdit.ac.in

contribution: Using MongoDB, a NoSQL database, to handle large amounts of data To Maintain Sensitive Data's Confidentiality Analyzing a Significant Amount of Data.

The term "NoSQL" communicates two different concepts. The first suggests a data management paradigm that is non-SQL compliant. The term "Not Only SQL" can also be used to describe settings that combine traditional SQL (or SQL-like query languages) with alternative methods of querying and access. This second, more prevalent (i.e., more often offered) interpretation. Continuous Availability Using NoSQL Database accessibility is categorized by IBM into three categories: 1) High availability: During the allotted time, even if the system is briefly down for maintenance, the database and application are still accessible. 2) Constant Uptime: The system is constantly operational with no scheduled downtime. Because HA and CO work together, data is always available and system repairs may be made without having to shut down the entire system. Consider a NoSQL solution if you need apps that are always accessible. [1] [2] [3] [4] [5] [6][7]. It's important to note that this differs from simple "high availability," where unplanned downtime is still expected even if it's not intended. There is never any downtime for systems with continuous availability. In today's market, where competition is just a click away, downtime may be catastrophic to a company's brand and bottom line. The typical cost of downtime varies greatly amongst businesses, ranging from about $90,000 per hour in the media industry to about $6.48 million for sizable internet brokerages. For NoSQL Data, 1.2.3 Geographic Independence A third justification for using NoSQL is the need for true geographic independence from a database. Practically speaking, "location independence" refers to the capacity to read from and write to a database without regard to the location of those I/O operations, as well as the ability to have any write functionality propagated out from that location so that it is accessible to users and machines at other sites. For the majority of conventional databases, this functionality is difficult to build but easy to define. However, distributing data everywhere is a different matter. Location independent read operations can occasionally be supported by master/slave and typically shared systems. Location independence is necessary for a number of reasons, including the need to service clients across many different regions and keep local data at those locations for easy access. 1.2.4 Using NoSQL, a Data Model Can Be More Flexible One of the primary reasons IT professionals choose a NoSQL database over a classic RDBMS is the more flexible data model included in the majority of NoSQL systems. A NoSQL datastore can be used for the reasons listed below: Many of these use cases, as well as others, can be handled by a NoSQL data model, whereas a few of these use cases benefit from the relational model's strengths. In addition, a NoSQL datastore, as opposed to a relational database, may readily accept any types of data, including structured, semi-structured, and unstructured data. Applications that use a range of datatypes are ideal candidates for NoSQL databases.

2. Review of Literature

Understanding relational and non-relational databases:

MySQL and Oracle are the two relational databases that are most frequently used. High scalability is not supported by relational databases, because the data must fit in predefined tables or structures. There are numerous important ways that non-relational databases differ from relational systems. The data can be inserted at any moment without specifying a schema because it doesn't utilize relations (tables) as its storage structure, doesn't use SQL as its query language, and allows for no join operations. Most widely used NoSQL databases (MongoDB and Cassandra). These databases contain a lot of user-related sensitive information, which raises questions about their ability to keep user information private and secret [2]. Data files in Mongo are not encrypted. Direct information extraction from the files is possible for any attacker with access to the file system. The application must explicitly encrypt any sensitive data before writing it to the database in order to mitigate this. the creation of a MongoDB-based textbook management system [3]. A lot of time is needed for multi-table queries[8][9][10][11][12][13]. In this paper, we attempt to leverage NoSQL to tackle this challenge. One of NoSQL's hallmarks is its lack of a schema. Students' and teachers' fundamental information will be created in MongoDB. You can build up a separate operating collection for subscription textbooks, used textbooks, textbook storage, entrance textbooks, and delivery textbooks.

Existing NoSQL Process

In the context of Big Data, complicated data cannot be handled by conventional data models. One tool for handling Big Data is the NoSQL database. One of the NoSQL databases, MongoDB, is a schema-free, document-oriented database that is also open source. Documents are used to store data. As a result, the database is document-oriented. A number of collections are stored in a MongoDB database. A table's equivalent, a collection, is a list. A group of documents are kept in a collection. Fields make up a document. It might be considered a row in a collection. Each document has a unique ID. Documents with different structures can be stored without issue. The Healthcare Record (HR), one of the Big Data apps, contains important data that clinicians have input. The challenge of knowledge discovery from such HR is difficult. Data generated by body vitals including blood pressure, body temperature, and lab results are included in HR. MongoDB stores reports and prescriptions. Before putting any sensitive data into a database, the user should explicitly encrypt it using RSA Algorithms. Analysis of medical data by querying frequently results in better diagnosis, early detection, and the suggestion of effective medications. Analysis frequently results in disease modeling based on clinical documentation. For instance, a doctor may use a database query to find out how many patients are overweight and offer dietary recommendations.

This paper's numerous modules are arranged in the manner as follows. B. Recognizing Applications for Datasets Many different sources, such as social networks and the media, mobile devices, online transactions, healthcare, etc., produce big data. There are a lot of patient data in healthcare. The diversity of needed organized and unstructured data, from basic patient data and medical histories to test findings and MRI pictures, is becoming more and more difficult to preserve. Data that offers a perspective of the patient, doctor, procedure, and other forms of information are stored using MongoDB in a single data store. C. Implementing the system. Connect to the MongoDB database using a Java program to carry out a variety of operations, including data addition, deletion, updating, and analysis-based data retrieval[14][15][16][17][18]. D. Adding Information to MongoDB User authentication is required for the database system. The object-oriented approach is used by MongoDB to develop its system. Document-oriented database MongoDB was created by 10gen. It controls groups of documents written in JSON (JavaScript Object Notation). Unlike conventional relational databases like MSSQL and MySQL, NoSQL databases have their own nomenclature. Row = Document or Record Field = Column Collection = Table or View One of the NoSQL databases, MongoDB, contains one or more documents in a collection (table) (records). There may be one or more fields in each document (solumn). After creating a database and a collection, add data in the form of documents. E. Securing the Private Information Data files in MongoDB are not automatically encrypted and are not encrypted at all. Direct information extraction from the files is possible for any attacker with access to the system. The application must explicitly encrypt any sensitive data before writing it to the database in order to alleviate this. Patient biodata and problem status are examples of sensitive information in the healthcare industry. These details shouldn't be shared. The data can be entered into the HR application by the authorized user. The RSA Algorithm allows the stored data to be encrypted. F. Data Analysis Data analysis comprises reporting and querying capabilities. It is advisable to retain the data produced by body vitals, test results, and prescriptions for analysis. Real-time medical data includes information that can be generated every 2–3 seconds, such as body temperature, blood pressure, pulse/heart rate, and breathing rate. These data, which are gathered from every person, produce a large amount of data quickly. MongoDB will be used to store these data. The analysis of these data in terms of queries will result in better diagnosis, early disease identification, and the suggestion of efficient medications. A list of all diabetes patients or a list of all patients within a specific age range may be queried as part of the analysis. G. MongoDB vs MySQL Performance Comparison MongoDB uses less time when inserting data than MySQL, which also increases query efficiency. In order to add patient records to a MongoDB database, a user must first log into the HR application. The user can encrypt data using the RSA technique before inserting it to store sensitive info. Updated patient information is kept on file so that analysis can be done for disease early detection or to find information based on query retrieval.

3. Discussion

By discussing its modeling and migration procedures as well as the justifications and advantages of utilizing NoSQL, this research focused on the taxonomy of the technology. It also showed how well NoSQL is utilized to store enormous volumes of data with high scalability. This article also describes why using a NoSQL database is advantageous and how to do so effectively. Because NoSQL databases are designed to grow with their users, it is believed that this would encourage their adoption. The sort of application the system will utilize will determine which database should be used because SQL and NoSQL respond differently depending on the type of queries that are run on them. The expectations of business users must be met if a NoSQL data storage is used instead of a relational model. In light of their existing knowledge of relational models, how well will a NoSQL data store perform? As should be obvious, many NoSQL data management environments are created for two primary needs: Fast accessibility, whether that means inserting data into the model or pulling it out via some query or access method, and Scalability for volume, in order to support the gathering and management of enormous amounts of data[19][20][21][22][23][24]. The NoSQL models mentioned above can be scaled, distributed, and flexible, and distribution and parallelization can be employed to meet each of these criteria. Furthermore, these distinctive qualities are compatible with programming paradigms like MapReduce, which effectively manage the creation and operation of several parallel execution threads. The key is making use of data distribution. The good news is that when employing a distributed tabular data store or key-value store, several queries and data accesses can be performed simultaneously, especially when the hashing of the keys maps to different data storage nodes. Linear performance scaling in relation to data volume will be made possible by clever data allocation tactics. NoSQL techniques are developed for high performance reporting and analysis. Many new companies are embracing the various NoSQL concepts and launching their own specifically specialized variants on the market. It might make sense to develop a simple "pilot" project model that can be implemented in many ways in order to compare and contrast ease-of-use, space performance, and execution speed as there is no danger in experimenting with the various NoSQL approaches if you are interested in them. Although noSQL data management options have alluring performance qualities, relational database management systems remain the industry standard. It's not always easy to decide whether to use NoSQL. One must take the business requirements into account before committing to the technology, as well as the skills needed to make the transition from a conventional manner to a NoSQL approach. If you select SQL or NoSQL technologies arbitrarily or because they're in high demand, you can come to believe that you automatically made the right choice. Given that both SQL and NoSQL have benefits and drawbacks, the ideal architecture will depend on the requirements of the apps you create. Even now, the temperamental old SQL database is incredibly strong and able to adequately handle your transactional requirements. Take a look at NoSQL options when you are reaching the boundaries of relational databases and the volume of data you are processing or the scope of your operations just calls for a more distributed solution. With thoughtful decision-making, liberate your data and develop the most amazing applications ever!

4. Summary

A new generation of tools that can query massive amounts of diverse data is required for big data. The Healthcare Record (HR) contains useful data that clinicians have entered. The vast majority of healthcare records are saved as documents in the NoSQL MongoDB database. Query-based analysis of medical records stored in MongoDB enables more accurate disease diagnosis and early disease identification. When it comes to inserting data and running queries, MongoDB is more effective than MySQL. Join operations can also be avoided. Scalability, availability, and high performance are all features of MongoDB. MongoDB also protects sensitive data by encrypting it on the outside with the RSA method. Traditional relational databases, which are widely used in information management systems, cannot manage structured data effectively. NoSQL is a cutting-edge technology that was created to manage unstructured data. It is a non-relational database management system. The primary motivation for NoSQL is scalability. The use of big data in medicine is one example. Everybody's data collection swiftly produces a large amount of knowledge. Through the examination of recorded healthcare data through querying, early detection, diagnosis, and the recommendation of effective treatments can all be accomplished. Any sensitive data must be explicitly encrypted by the program before being written to the database in order to provide security. Additionally, when querying the NoSQL database MongoDB, joining procedures are avoided, leading to efficient data retrieval.

REFERENCES

[1] Abramova, V., & Bernardino, J. (2013, July). NoSQL databases: MongoDBvs Cassandra. In Proceedings of theinternational C* conference on computer scienceand software engineering (pp. 14-22).

[2] Ali, W., Shafique, M. U., Majeed, M. A., &Raza, A. (2019). Comparisonbetween SQL andNoSQL Databases andTheirRelationshipwith Big Data Analytics. Asian Journal of Research in Computer Science, 4(2), 1-10

[3] Becker, M. Y., &Sewell, P. (2004, June). Cassandra: Flexible trust management, appliedtoelectronic health records. In Proceedings. 17th IEEE Computer Security Foundations Workshop, 2004. (pp. 139-154). IEEE.

[4] Berg, K. L., Seymour, T., &Goel, R. (2013). History of databases. International Journal of Management & Information Systems (IJMIS), 17(1), 29-36.

[5] Bjeladinovic, S., Marjanovic, Z., &Babarogic, S. (2020). A proposal of architectureforintegrationand uniform use of hybrid SQL/NoSQL database components. Journal of Systems and Software, 168, 110633.

[6] Chandra, D. G. (2015). BASE analysis of NoSQL database. FutureGeneration Computer Systems, 52, 13-21.

[7] Chen, J. K., & Lee, W. Z. (2019). An introduction of NoSQL databases based on theircategoriesandapplicationindustries. Algorithms, 12(5), 106.

[8] Cuzzocrea, A., &Shahriar, H. (2017, December). Data maskingtechniquesforNoSQL database security: A systematic review. In 2017 IEEE International Conference on Big Data (Big Data) (pp. 4467-4473). IEEE.

[9] de Oliveira, V. F., Pessoa, M. A. D. O., Junqueira, F., &Miyagi, P. E. (2021). SQL andNoSQL Databases in the Context of Industry 4.0. Machines, 10(1), 20.

[10] Deka, G. C. (2013). A survey of cloud database systems. It Professional, 16(2), 50-57. IEEE.

[11] Di Martino, S., Fiadone, L., Peron, A., Riccabone, A., & Vitale, V. N. (2019, June). Industrial Internet of Things: Persistencefor Time Series withNoSQL Databases. In 2019 IEEE 28th International Conference on Enabling Technologies: InfrastructureforCollaborative Enterprises (WETICE) (pp. 340-345). IEEE.

[12] dos Santos Ferreira, G., Calil, A., & dos Santos Mello, R. (2013, December). On providing DDL support for a relationallayer over a document NoSQL database. In Proceedings of International Conference on Information Integration and Web-based Applications & Services (pp. 125-132).

[13] Gessert, F., Wingerath, W., Friedrich, S., & Ritter, N. (2017). NoSQL database systems: a survey anddecisionguidance. Computer Science-Research and Development, 32(3), 353-365.

[14] Guimaraes, V., Hondo, F., Almeida, R., Vera, H., Holanda, M., Araujo, A., ... &Lifschitz, S. (2015, November). A study of genomic data provenance in NoSQL document-oriented database systems. In 2015 IEEE International Conference on BioinformaticsandBiomedicine (BIBM) (pp. 1525-1531). IEEE.

[15] Rodriguez, K. M., Reddy, R. S., Barreiros, A. Q., &Zehtab, M. (2012, June). Optimizing Program Operations: Creating a Web-Based Application toAssignand Monitor PatientOutcomes, Educator Productivity and Service Reimbursement. In DIABETES (Vol. 61, pp. A631-A631). 1701 N BEAUREGARD ST, ALEXANDRIA, VA 22311-1717 USA: AMER DIABETES ASSOC.

[16] Kwon, D., Reddy, R., & Reis, I. M. (2021). ABCMETAapp: R shinyapplicationforsimulation-basedestimation of meanand standard deviationfor meta-analysis via approximateBayesiancomputation. Research synthesismethods, 12(6), 842–848. https://doi.org/10.1002/jrsm.1505

[17] Reddy, H. B. S., Reddy, R. R. S., Jonnalagadda, R., Singh, P., &Gogineni, A. (2022). Usability Evaluation of anUnpopular Restaurant Recommender Web Application Zomato. Asian Journal of Research in Computer Science, 13(4), 12-33.

[18] Reddy, H. B. S., Reddy, R. R. S., Jonnalagadda, R., Singh, P., &Gogineni, A. (2022). Analysis of theUnexplored Security Issues Common toAll Types of NoSQL Databases. Asian Journal of Research in Computer Science, 14(1), 1-12.

[19] Singh, P., Williams, K., Jonnalagadda, R., Gogineni, A., &; Reddy, R. R. (2022). International students: What's missing andwhatmatters. Open Journal of Social Sciences, 10(02),

[20] Jonnalagadda, R., Singh, P., Gogineni, A., Reddy, R. R., & Reddy, H. B. (2022). Developing, implementingandevaluating training for online graduate teaching assistantsbased on Addie Model. Asian Journal of EducationandSocial Studies, 1-10.

[21] Sarmiento, J. M., Gogineni, A., Bernstein, J. N., Lee, C., Lineen, E. B., Pust, G. D., &Byers, P. M. (2020).Alcohol/illicitsubstanceuse in fatalmotorcycle crashes. Journal of surgical research, 256, 243-250.

[22] Brown, M. E., Rizzuto, T., &Singh, P. (2019). Strategic compatibility, collaborationandcollective impact for community change. Leadership&Organization Development Journal.

[23] Sprague-Jones, J., Singh, P., Rousseau, M., Counts, J., &Firman, C. (2020). The Protective Factors Survey: Establishingvalidityandreliability of a self-report measure of protective factors againstchild maltreatment. ChildrenandYouth Services Review, 111, 104868

[24] Sadashiva Reddy, H. B. (2022). ExploringtheExistingandUnknown Side Effects of Privacy Preserving Data MiningAlgorithms. Doctoraldissertation. Nova Southeastern University. RetrievedfromNSUWorks, College of Computing and Engineering. (1179)