International Journal of Core Engineering & Management
Volume-7, Issue-11, 2024 ISSN No: 2348-9510

ROLE OF MONGO DB IN BIG DATA: NOSQL SOLUTION FOR REAL-TIME DATA PROCESSING

Hareesh Kumar Rapolu
hareeshkumar.rapolu@gmail.com

Abstract

The world is experiencing an exponential growth of Big Data making conventional database systems relational databases less dependable. MongoDB, one of the dominant NoSQL databases emerged to address these limitations. Its superiority to traditional database systems is explained by schema-less architecture and horizontal scalability. MongoDB is a more versatile tool that meets Big Data demands in terms of volume, velocity and variety. This article explores the role of MongoDB in the processing and management of Big Data. It demonstrates capacity to process data in real-time and has flexible data handling capabilities. The paper proposes MongoDB as an integral asset relevant in today's Big Data systems thanks to its ability to scale horizontally and process high-velocity data streams.

Keywords—Big Data, MongoDB, NoSQL, Real-Time Data Processing, Vertical Scaling, Horizontal Scalability, Actionable Insights, Artificial Intelligence, Internet of Things.

I. INTRODUCTION

Announced in 2007, MongoDB is a NoSQL database that differs from traditional relational databases such as SQL which are organized in structured tables [1]. NoSQL databases like MongoDB emerged due unprecedented challenges in data management resulting from digital content associated with advancements in Internet of Things (IoT), artificial intelligence (AI), social media and enterprise systems [1]. These advancements complicates the management of Big Data, the extremely large and varied sets of unstructured, structured and semi-structured data that is constantly growing [1]. The datasets are massive and complex in terms of volume, velocity, and variety, such that conventional data management solutions fall short of storage, processing and analysis [2]. Google Cloud states that "as data continues to expand and proliferate, new big data tools are emerging to help companies collect, process, and analyze data at the speed needed to gain the most value from it" [3].NoSQL databases are part of the prominent Big Data tools developed to address limitations of traditional data management solutions. MongoDB has reported substantial evolution and growth through continuous iterative improvements to address changing needs of contemporary data management [4]. At the core of MongoDB is its flexible design and structure for quick iterative changes and compatibility with dynamic data management needs with cost efficient schema adaptations [4]. MongoDB's lack of strict schemas coupled with its horizontal scalability enables effective management of Big Data characteristics by high data volume, velocity and variety [2]. MongoDB commands a market share of about 45.7% in the NoSQL databases, competing with Amazon DynamoDB (10.4%) and Apache Cassandra (4.5%) [5]. Most of industry areas of application for MongoDB include the service and information technology sectors [4]. Besides, MongoDB is increasingly applied in other sectors such as healthcare, higher education, finance, advertising and retail [4].

Technology	Est. Market Share (%)	Domains
MongoDB	45.7	37,883
Amazon DynamoDB	10.4	8,661
Apache Cassandra	4.5	3,692
Hbase	4.06%	3,370

Figure 1: Top four Competitors and Alternatives to NoSQL [5]

II. BIG DATA CHALLENGES

The concept of Big Data is often characterized by the "three Vs" dimensions: volume, variety and velocity [3]. It is under the three Vs that challenges of Big Data are conceptualized in relation to the processing and management of massive datasets. Volume, variety, and velocity are known to overwhelm the traditional database management systems (DBMS) as they are not sufficiently equipped to deal with rising digitalization, IoT, mobile technologies and social media data [6]. In particular, there is unprecedented sheer **volume** of data from multiple platforms, sensors and devices. A 2024 report from the International Data Corporation (IDC) showed that global data generated is projected to increase from 149 zettabytes to 175 Zettabytes by 2025. The data generated worldwide will only increase further as the Big Data continues growing ([7]).

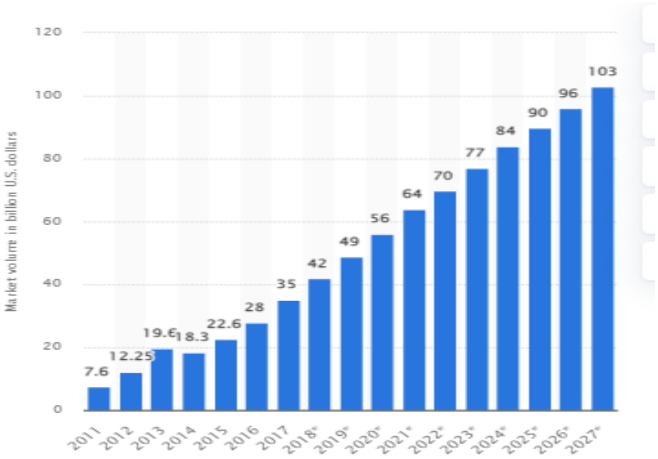

Figure 2: Big Data market size over the years and forecasts [7]

Traditional DBMSs depend on vertical scaling such that they need more RAM, CPU and storage to manage the increased load. Yet, this vertical scaling comes with limitations in terms of capacity and cost [8]. This is in contrast to horizontal scaling of modern Big Data architectures such as NoSQL's MongoDB: Data is distributed across multiple nodes or servers in a cluster for almost infinite scaling [8]. Accordingly, the problem of sphere volume makes MongoDB a more practical solution compared to TDBMSs.

The problem of Big Data extends beyond **volume** as it is also diverse as defined by the variety dimension. Traditional relational databases designed for structured data (predefined rows and columns) are less relevant in the modern world of highly unstructured data [9]. The different formats such as structured data (like transaction records), semi-structured (like XML, and JSON files) and unstructured data (like social media post, videos, images and emails) make traditional relational databases less relevant. This is a stark contrast to NoSQL databases like MongoDB as they are designed to deal with different data formats. The schema-less architecture of MongoDB makes it possible to store data in JSON-like document format (BSON) which is flexible [10]. This suggests that there is no need for schema changes or complex migrations to store structured, semi-structured and unstructured data.

Big Data **velocity** denotes the speed of generating, collecting and processing data [11]. Velocity is becoming a major challenge for many industries as the need to analyze and generate actionable insights in real-time is increasingly critical [11]. To illustrate, real-time data processing is a requirement in financial services for fraud detection, implementing high-frequency transaction and mitigating risks [12]. Uninterrupted flow of data is also required for real-time recommendations in e-commerce [10]. Healthcare services require real-time tracking of patient data from multiple sources such as wearables to detect health issues early enough [13]. These demands are beyond the capability of traditional databases given the velocity of data streams. Significant latency is observed in these systems such that analyzing high-speed data streams in real time is a challenge that MongoDB solves. It is powered by efficient indexing and flexible data models for fast read and write operations [10]. MongoDB's is also characterized by change streams that allow it to capture real-time changes for immediate reaction to frequent events.

III. MONGODB AS A NOSQL SOLUTION FOR BIG DATA

The widespread adoption of MongoDB in multiple industries is explained by its flexibility and scalability. The document-oriented NoSQL database is designed as a repository of data in Bison JSON format which translates to the capacity to accommodate large nested data structures in absence of predefined schema [14]. The fact that MongoDB lacks a rigid schema structure explains why organizations using it achieves flexibility in data storage, be it unstructured or semi-structured data. The schema-less architecture of MongoDB makes it significantly relevant for Big Data contexts where there is a rapid change in data structure. For example, sensors generate different types of data with varied frequency for IoT systems. Adopting MongoDB means seamless integration of the different data formats with limited downtimes and inexpensive schema migrations [1].

Horizontal scalability in the form of sharding is another defining feature of MongoDB. This denotes the distribution of data through different nodes and serves for effective management of datasets [9]. Data is partitioned into small manageable chunks then distributed across a cluster such that datasets are efficiently processed irrespective of traffic levels [8]. Indeed, the dominance of MongoDB in Big Data applications is largely explained by the elastic scalability [4]. This advantage is further explained by real-time data processing capabilities required in data-intensive industries like finance, healthcare, and e-commerce [10]. E-commerce platforms, for example, generate vast datasets from activities such as user interactions, purchases, and product inventories. Actionable insights are acquired from these datasets to enhance user experience, minimize rates of cart abandonment and increase sales [10]. Accordingly, it is conceivable that MongoDB stands out as a viable solution in place of traditional relational databases to process and manage Big Data in real-time.

IV. CONCLUSION

MongoDB is undoubtedly a superior database for data processing in the context of contemporary Big Data. The document-oriented model coupled with tools like flexible schema and BSON capabilities allows for its scalability and simplified management of data types. It outperforms traditional DBMs. This article contributes to the extant knowledge on the existing and potential inherent in MongoDB as the volume, velocity and variety of Big Data are expected to proliferate in the near future. Accordingly, it is conceivable that the use of MongoDB is unavoidable in the growing data-intensive environments.

REFERENCES

1. G. K. Kaur, S. Singla, and V. Khawas, "Database Management System: A Study of Increasing Impact of NoSQL Databases," in 2023 International Conference on Advanced Computing & Communication Technologies (ICACCTech), Dec. 2023, pp. 370-374.
2. A. Ali, S. Naeem, S. Anam, and M. M. Ahmed, "A state of art survey for big data processing and NoSQL database architecture," International Journal of Computing and Digital Systems, vol. 14, no. 1, p. 1, May 2023.
3. H. Yadav Structuring SQL/NoSQL databases for IoT data. International Journal of Machine Learning and Artificial Intelligence. Vol. 3, no. 5 Apr. 2024, pp. 1-2.
4. Google Cloud, "Big data defined: Examples and benefits | Google cloud," Google Cloud, [Online]. Available: https://cloud.google.com/learn/what-is-big-data. [Accessed: Feb. 2, 2024].
5. A. Vágner, "How Do NoSQL Databases Handle Variety of Big Data?," in International Congress on Information and Communication Technology, Singapore: Springer Nature Singapore, Feb. 2024, pp. 459-469.
6. Sense, "NoSQL," 6Sense, 2024. [Online]. Available: https://6sense.com/tech/nosql-databases/nosql-market-share. [Accessed: Feb. 2, 2024].
7. S. V. Oprea, A. Bâra, and N. Oprea, "Big Data Management and NoSQL Databases," Ovidius University Annals, Economic Sciences Series, vol. 23, no. 1, pp. 466-475, 2023.
8. P. Taylor, "Global big data industry market size 2011-2027," Statista, Feb. 13, 2024. [Online]. Available: https://www.statista.com/statistics/254266/global-big-data-market-forecast. [Accessed: Feb. 18, 2024].
9. A. Agrawal and Y. Mamajiwala, "Comparative Analysis of RocksDB, LMDB, and MongoDB: A Performance Evaluation," SAE Technical Paper, Jan. 2024.
10. C. Sowandi, I. Leonita, S. Arisaputra, and D. David, "Exploring the Effectiveness, Features, and Compatibility of MongoDB and MySQL: A Comprehensive Comparison of NoSQL and Relational Databases," MIND (Multimedia Artificial Intelligent Networking Database) Journal, vol. 8, no. 2, pp. 217-229, Dec. 2023.
11. H. Shwetha, D. Prajwal, and S. Sridharan, "From MongoDB to React: Unleashing the Power of MERN in E-commerce," in 2024 International Conference on Emerging Technologies in Computer Science for Interdisciplinary Applications (ICETCS), Apr. 2024, pp. 1-6.
12. K. Krishan, G. Gupta, and G. S. Bhathal, "Striking the Balance: Comprehensive Insights into Data Consistency in NoSQL Realms," in 2024 11th International Conference on Computing for Sustainable Global Development (INDIACom), Feb. 2024, pp. 715-720.
13. M. Y. Hassan, "Applications of Bigdata Technologies in the Comparison of BMTD and ARIMA Models for the Prediction of Internet Congestion," IEEE Access, Apr. 16, 2024.
14. R. Bawaneh and A. Jallad, "Big Data in Healthcare: A Review on Applications, Technologies, Benefits and Challenges," in 2024 21st International Multi-Conference on Systems, Signals & Devices (SSD), Apr. 2024, pp. 576-582.
15. S. Kleuker, "NoSQL mit MongoDB und Java," in GrundkursDatenbankentwicklung: Von der AnforderungsanalysezurkomplexenDatenbankanfrage, Wiesbaden: Springer Fachmedien Wiesbaden, Mar. 2024, pp. 383-414.