NO SQL (NOT ONLY SQL) DataBases: A Review

Kiran Kumar B G [1*], Prakash H.Unki [2], Suvarna L.Kattimani [3]

[1]Dept. CSE, BLDEA'S V.P. Dr. P. G. Halakatti College of Engineering and Technology, Vijayapur – 586103, India.

Corresponding Author: bgkiran16@gmail.com, Tel.: +00-99804-05065

Abstract— This Review portrays the significance of NoSQL databases in the field of Big Data and distributed computing. Because a number of clients added to cloud computing and since they produce the huge volume of organized, unstructured and semi-structured information, to deal with these kinds of information NoSQL databases are preferred. This survey helps to analysis the data models of NoSql and its advantages and limitations of NoSql and relational databases. In the end, the Review concludes with various reasons to adopt NoSQL in Big Data and Cloud computing.

Keywords— Not Only SQL, BigData,GridFS,JSON

I. INTRODUCTION

Relational databases are working based on the relational model it was developed 30 years ago to provide the data storage for business applications. Since it has the common database model for storing a wide range of data and information of different types. However, the software and hardware technologies have evolved in various manners in terms of data storage, data mining, data processing, analysis, and application.

The invent of web 2.0 has created a various advanced application based on the storing and processing large of amount of data and it requires high scalability and availability which increases challenges to relational database management systems. The Relational DBMS are having limitations related to performance and scalability when a large amount of data increases continually.

This leads to the invention of a new concept called Not Only SQL(NoSql).but it has become more popular since the adoption of the Database as a Service in cloud computing. Because of large volume, high scalability, high-performance schema-less architecture provided by NoSql.

NoSql (Not Only SQL) is designed and developed to adopt the data scalability when the data grow infinitely. Most of the NoSQL Systems have removed multiplatform and unnecessary features of relational database management software, making them more compact, lightweight and efficient than traditional relational database management system. Section 2 discusses the literature review. The final conclusion is discussed in section 3.

II. RELATED WORK

In [1], the authors have demonstrated of data storage system in Mongo DB. In this detailed analysis, the authors exhibited features of Mongo DB for storing the document-oriented databases and adopted Grid file system framework which is utilized to enhance the data access efficiency in Mongo DB. Mongo DB uses BSON to store the embedded documents.

In [2], the authors have evaluated performance of NoSql using Mongo DB, and compares the query response of both MySQL and NoSQL databases, concluded that results of NoSQL databases are faster than MySQL, and document-oriented storage is suitable to hold the health records and finally suggested the NoSql is the best fit for large-scale applications like EHR typically hold a large amount of patient data.

In [3], the authors have presented the importance of schema design for NoSQL Databases because they became popular for high performance and high Scalability and availability they can provide. But it's not easy to design schema for large volume databases because there is no standard data type format available for heterogeneous data which was generated by a huge number of internet users. and also developed a standard model for Cassandra data storage. They proposed a new concept called "NoSQL Schema Evaluator (NoSE) "is able to provide automatic schema design process model for NoSQL databases.

In [4], the authors have proposed a new concept called Reconfiguration in sharded NoSQL Systems.in a morphous system mainly shard the primary keys and stores in multiple

sites. This improves the rate of availability of data among the servers. This tool provides configurations of NoSql Systems data storage in a real-time manner. The author has suggested usage of morphus, it supports concurrent read and writes operations among the multiple servers. Morphus follows Master Slaver replication architecture, range partitioning, and dynamic data transfer.

In [5], the authors have discussed how SQL injection attacks disable the web applications, usually these application data transferred from client to server. during the transmission, the intruders insert the additional query it is called query injection.it modifies the original query of the client application, at server side, it assumes that request coming from client application but it is injected with SQL query it is going to damage the web application or hack the information or crash the entire system. Authors also discussed the various injection detections mechanisms like penetration testing.

In [6], the authors present a new Idea that implementing NoSql in cloud computing. as we know the cloud computing involves a large number of internet users and various cloud service provides and usage of cloud services generates huge volume of data, using traditional RDBMS handling these data is very difficult and less efficient, so authors suggest the adopting new concept called NoSql which provides the storage of large volume of data .and it has great advantages of high performance and high scalability. The data models of NoSQL are well suited to cloud computing and it meets the requirements of cloud users and service providers. Finally proposed to adopt the document-oriented storage, key-value, column-oriented storage in cloud computing.

In [7], the authors discussed storage and efficiency of NoSQL databases in cloud computing. Cloud data and cloud services are commonly delivered through the network. Data is more sensitive and important in the network it should be protected from unauthorized access. And storing the data in appropriate mode is essential .so most of the corporate companies and organization recommends to adopt NoSQL because it meets a requirement of high availability and high scalability of a large volume of data. Authors proposed the implementation of Hadoop and Map Reduce and Mongo DB in cloud computing.

In [8], authors have presented the concept of the capacity planning of NoSQL databases. this study helps to understand the limitations of NoSQL storage and its implementation.it is very essential to analyze the capacity of a database before going to the implementation of a real-time application. For better efficiency and availability of the data and high scalability. They took a survey on linked in website, LinkedIn serves online connectivity for data exchange of millions of users by using stateful services like RESTful, HTTP, Soap etc. these techniques have various limitations in

terms of no of users, amount of data to be transfer, authors have cleanly noted down results of capacity of these individual services.

In [9], The authors have presented the idea of improvising the storage Space efficiency using compression, and de-duplication methods, De-duplication is a concept of maintaining the single copy of duplicate records on multiple sites. De-duplication creates the problem of disk bottleneck.[9],For compression, multiple hash methods were used. but the solution for the disk bottleneck problem was unaddressed.

In [10], authors have conducted a survey on Big Data Analytics, they mainly highlighted the working of NoSQL in the field of Big Data Analytics, for this analysis they used Cassandra NoSQL Database to perform analytics ,Cassandra widely used in many social network companies to track the usage of data, engagement of particular services ,sessions, average session, average visits particular website. And so on. This analytics helps to calculate the sales and helps to make business forecasting, strategy building, and market analysis. Also discussed many challenges of analytics they are collection and Integration, storage and management, modeling, scalability, cost efficiency.

In [11], this paper author proposed the importance of optimization of partition technique in the NoSQL database.in cloud computing data availability is one of the major challenges.to increase the data availability various techniques have been used. like Data replication, sharding, Duplication etc. in these techniques data is divided and stored in multiple sites. Sometimes an improper division of the data creates the problem of data unavailability .to solve this kind of problem effective optimization is required. Authors proposed De-normalization of the table and ignoring ACID properties. and adopting CAP theorem is helpful in optimization of partitions in NoSql.

In [12], authors have explored the advantages of NoSQL Database Systems and ignore the limitations, since the integration of cloud and Big Data has created different dimension for data w.r.t analytics, business intelligence .these two factors have the latest advancement in every business done through online or over a internet.to manage multiple data dimension RDBMS is not capable, thus NoSql is adopted to manage the Big Data. the main Advantage of NoSql is its Schema-less, multiple ways of storing data is available, like a key-value pair, document-oriented, graph-based, etc. and NoSql is managed by System itself, thus no need of database administrator. provide working model Of NoSQL Database.

In [13], authors have analyzed the detailed Workflow of NoSQL databases. NoSql database adopted following

properties BASE, CAP[13]. and provide a working model of NoSQL Databases which are explained in detail they are as follows "Key-Value stores, Document databases, Column stores, Graph databases".

In [14] , authors have discussed the need of document orient storage for Global positioning systems, because in RDBMS database have limited data type format to hold the GPS location information, since GPS is based on longitude and latitude along with location tag are continuously changing, so it very difficult to handle dynamic contents generated from Global positioning system.to solve these problems authors have proposed the NoSql database Document-oriented storage.in this Document-oriented storage, data is stored in the format of Embedded "JSON or XML" format.it is very easy for users to perform search and it's a schema-less architecture.

In [15], authors have discussed the benefits of NoSQL they are Scalability and Global Availability Performance and Flexible Data Modeling each of which explained with appropriate data models like document-oriented, key-value, column-oriented storage.

III.   CONCLUSION

This Review describes the importance of NoSql in the field of Big Data and Cloud Computing. And this study focuses on the detailed benefits of NoSQL they are global availability and scalability. And also covers limitations they are immature and still there is no standard interface was available but these limitations are negligible. This review compares the MySQL and NoSQL data models along with their characteristics benefits, and limitations. Finally, it concludes that NoSql efficient and preferred by most of the users and corporate companies.

REFERENCES

[1]. Y. Gu, X. Wang, S. Shen, J. Wang and J. Kim, "Analysis of data storage mechanism in NoSQL database MongoDB," *2015 IEEE International Conference on Consumer Electronics - Taiwan*, Taipei, 2015, pp. 70-71.

[2]. J. Klein, I. Gorton, N. Ernst, P. Donohoe, K. Pham and C. Matser, "Application-Specific Evaluation of No SQL Databases," *2015 IEEE International Congress on Big Data*, New York, NY, 2015, pp. 526-534.

[3]. M. J. Mior, K. Salem, A. Aboulnaga and R. Liu, "NoSE: Schema Design for NoSQL Applications," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 10, pp. 2275-2289, 1 Oct. 2017.

[4]. M. Ghosh, W. Wang, G. Holla and I. Gupta, "Morphus: Supporting Online Reconfigurations in Sharded NoSQL Systems," in IEEE Transactions on Emerging Topics in Computing, vol. 5, no. 4, pp. 466-479, 2017.

[5]. A. Maraj, E. Rogova, G. Jakupi and X. Grajqevci, "Testing techniques and analysis of SQL injection attacks," *2017 2nd International Conference on Knowledge Engineering and Applications (ICKEA)*, London, 2017, pp. 55-59.

[6]. A. Antoniadis, Y. Gerbessiotis, M. Roussopoulos and A. Delis, "Tossing NoSQL-Databases Out to Public Clouds," *2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing*, London, 2014, pp. 223-232.

[7]. Dayne Hammes ,Hiram Medero,"Comparison of NoSQL and SQL Databases in the Cloud", Georgia Southern University Hiram_Medero@georgiasouthern.edu, Hammes, Medero, et al. Proceedings of the Southern Association for Information Systems Conference, Macon, GA, USA March 21st –22nd , 2014

[8]. R. Pasumarti, R. Barot, S. Xia, A. Xu and H. Ramachandra, "Capacity Measurement and Planning for NoSQL Databases," *2017 IEEE 11th International Conference on Semantic Computing (ICSC)*, San Diego, CA, 2017, pp. 390-394.

[9]. V. Bhatia and A. Jangra, "SETiNS: Storage efficiency techniques in No-SQL database for Cloud based design," *2014 International Conference on Advances in Engineering & Technology Research (ICAETR - 2014)*, Unnao, 2014, pp. 1-5.

[10]. S. Malhotra, M. N. Doja, B. Alam and M. Alam, "Bigdata analysis and comparison of bigdata analytic approches," *2017 International Conference on Computing, Communication and Automation (ICCCA)*, Greater Noida, 2017, pp. 309-314.

[11]. L. Ho, M. Hsieh, J. Wu and P. Liu, "Data Partition Optimization for Column-Family NoSQL Databases," *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, Chengdu, 2015, pp. 668-675.

[12]. B. Jose and S. Abraham, "Exploring the merits of nosql: A study based on mongodb," *2017 International Conference on Networks & Advances in Computational Technologies (NetACT)*, Thiruvanthapuram, 2017, pp. 266-271

[13]. Raju Sharma1, Yatendra kashyap,"A Study Of Nosql Databases And Working Overviews", International Journal of Recent Trends in Engineering & Research (IJRTER) Volume 02, Issue 02; February – 2016[ISSN:2455-1457].

[14]. D. C. M. Maia, B. D. C. Camargos and M. Holanda, "Voluntary geographic information systems with document-based NoSQL databases," *2016 11th Iberian Conference on Information Systems and Technologies (CISTI)*, Las Palmas, 2016, pp. 1-6.

[15]. Nikhil Dasharath Karande, "A Survey Paper on NoSQL Databases: Key-Value Data Stores and Document Stores", International Journal of Research in Advent Technology, Vol.6, No.2, February 2018 E-ISSN: 2321-9637.

Authors Profile

Dr. Prakash H. Unki pursued Bachelor of Engineering from Karnataka University Dharwad, Master of Technology and Ph.D from Visvesvaraya Technological University Belguam, India. He is currently working as Professor in Department of Computer Science & Engg in BLDEA's College of Engineering and Technology, Vijaypur, Karnataka, India. He is Member of ISTE (The Indian Society for Technical Education), IEI (The Institute of Engineers), CSI (The Computer Society of India), He has published more than 14 research papers in reputed international journals and conferences including IEEE. His main research work focuses on Digital Image Processing. He has 21 years of teaching experience.

Mrs Suvarna L.Kattimani pursued Bachelor of Engineering from karnataka University Dharwad, and Master of Technology from   Indian Institute of Engineering Science and Technology Shibpur,West Bengal,India.. She is currently pursuing Ph.D. from Visvesvaraya Technological University Belguam, India and currently working as Assistant Professor in Department of Computer Science & Engineering
in BLDEA's College of Engineering and Technology, Vijaypur, Karnataka, India,She has published more than 18 research papers in reputed international journals and conferences including IEEE. Her main research work focuses on Digital Image Processing. She has 17 years of teaching experience.

Mr.Kiran Kumar BG pursued Bachelor of Engineering in Information Science &Engg and M.Tech in (Master of Engg Management) from Visvesvaraya Technological University Belguam, India He is currently pursuing M.Tech in Computer Science &Engg in BLDEA's College of Engineering and Technology, Vijaypur, Karnataka.