ANALYTICAL EVALUATION OF STRUCTUREDTO UNSTRUCTURED DATABASE FOR LARGE VOLUME OF E-COMMERCE DATASET

Jitanshu Sekhar Patra Asst. Prof. Einstein Academy of Technology and Management, Bhubaneswar, India
Neelamani Samal Asst. Prof. Einstein Academy of Technology and Management, Bhubaneswar, India
Swatisubhra Nayak Student, Einstein Academy of Technology and Management, Bhubaneswar, India

Abstract: The purpose of the study is to estimate document-oriented MongoDB database with SQL in terms of Performance. For improved dependability and competence of parallel computing, NoSQL techniques distributes the data indiscriminately on data nodes. This approach provides phenomenal benefits for general data accessing methods. Hence the migration as well as conversion of RDBMS to NoSQL is now a best choice or attention of many researchers in recent times.
In software development, migration from a RDBMS to NoSQL is a challenge for database administrators as well as programmers. However, due to explosive growth of big data, the well-structured format of Relational Databases may limit the scalability to maintain horizontal scaling. So to gain flexibility as well feasibility of parallel processing, the data storage methods take NoSQL databases into this consideration.

KEY WORDS: RDBMS, NoSQL, RDBMS, Big Data

INTRODUCTION

As we all are aware that in today, to handle a database is a challenging task for the user. To solve the problem of maintaining big data, a new kind of data store has been introduced called "NoSQL" database. There are lots of NoSQL databases available today and it is the most challenging task for the users to choose one of them for their system.

Our focus is to evaluate the performance of switch over to RDBMS to NoSQL platform. We came into consideration that, ACID transaction properties of RDBMS databases receive less attention in NoSQL data base. As a result, this will give higher impact in performance. In NoSQL engine, the ACID properties are replaces by BASE Architecture. More description of this Architecture can be found in [1]. Secondly, Joins as well as Complex Transactions are not supported inNoSQL databases due to their specific architecture.

Research Methodology

Organizations now adopting MongoDB because it enables them to build database system which are faster, handle huge big data, and manage applications more efficiently and conveniently. MongoDB documents are designed for easy accessing of data.

It provides new level of scalability and availability which are not provided by other traditional database. Because of high data volume and throughput, MongoDB can be easily implemented without much downtime, and without changing the application strategies.

An example would be the booking of tickets behind a travel reservation system, which involves complex transactions. However the core booking system might run on MySQL, those system parts with users including serving booking, integrating with social networks, managing

sessions – would be better when we placed in MongoDB. MongoDB came with the consideration of giving the new way of data storage.

NoSQL structure can be used to store data in a dynamic schema, instead of a tabular representation l ike SQL(Relational database). NOSQL structure like MongoDB modified JSON format into its BSON format, which store the data as a binary format. So due to its binary format provide more reliable and efficient structure in the area of storage space and faster speed.

Research objectives

The proposed system taken into considerations with NoSQL to provide data processing, migrating and retrieving in a reliable and flexible way.

Nowadays there are number of challenges connected with the migrating data from structured to unstructured databases.We have to use such database models which helpful to reduce data redundancy issues. Most internet companies migrate their database from relational to NoSQL databases because of the high volume of stored data is massive and so relational databases strategies are failed to satisfy the flexibility, good performance expectations. So to aim of such proposed work is to evaluate the performance issues, scalability, high availability and flexible database model[13].

Tools & Techniques of Research Software and Hardware:

- Microsoft Windows7 64-bit
- PHP 5.3.5
- MySQL 5.1.62
- 10 gen MongoDB 2.0.5
- Lenovo system, CPU 2.30 GHz (I used only 40% from CPU)
- 4 GB RAM

Hypothesis

our research is to compare key-value store implementations on NoSQL and SQL databases. so we simply focused on why NoSQL databases perform better than the SQL databases in e-commerce. There may be a wide variation in performance based on the type of operation (like read or write) so choosing perfect NoSQL system, we definitely improve such things to some extent.

Review of Literature

A. Babu et al. (2017) present a comparison of different NoSQL databases like HBase, Casandra, and MongoDB based on their structure, performance, consistency, scalability, and transactional features. Also, discuss various methodologies for migrating the data from SQL relational database to NoSQL database [2].

F. Zhu et al. (2012) present scalable, fast, and high- performance system to implement read SQL query (no insert or update) by exploiting a number of NoSQL database features. HBase package was used as a data storage and design a partition joined table to perform joinoperations that are not supported by the NoSQL. The defect for this approach that it has time cost that can be reduced using several approaches like pre-joining tables [3].

S. Khan et al. (2013) proposed a method which provides SQL Query language support to the NoSQL database MongoDB by adding an interface between the application layer and database

layer. This interface contains all the routing information as the conversion rules, this will help to communicate with the database layer and to convert from one format to another.This model will achieve the scalability of the big data without affecting the logical implementation [4].

R. Lawrence (2014) proposed a new system that allows NoSQL data to be accessed using SQL queries and simply deals with any software packages supporting Java Database Connectivity package. The proposed system uses MongoDB package as a NoSQL data storage. The results show that the joins operations can be implemented efficiently but it adds a minimal overhead in translating process [5].

M. Hanine et al. (2014) introduce the new processes for migration data from SQL database to NoSQL database. This process consists of two steps: Loading the logical structure of the source database and then mapping between the RDBMS and MongoDB model. This process tries to solve the joins between tables in RDBMS. If the table to be migrated is join to another table, it has to migrate both tables to a single table in NoSQL [6].

M. Potey et al. (2015) developed a system for converting structured database, SQL to the unstructured database, NoSQL BigData. The results of this study show that the database system becomes more reliable and efficient when classi cal dat abase syst em s are developed an d complemented by using specific features and proprieties for the NoSQL database [7].

Y. Zheng et al. (2015) proposed a method that follows NoSQLs De-normalization, Duplication and Intelligent (DDI) principles keys. The final results show that the suggested method improve access performance about forty-seven percentage [8].

M. Claudino et al. (2016) proposed a new technique which is used to improve the data conversion process between SQL and NoSQL databases. In this work, the authors present the conceptual model in the relational database system and implemented it in NoSQL database systems. The results show that, by using this model, the obtained NoSQL database is completely related to the source relational one [9].

G. Akansha et al. (2016) proposed a fast and space efficient algorithm to validate data between cross platform databases (RDBMS) and bloom filters (NoSQL) using de-normalized schema structures. The experimental results show that the proposed algorithm has the ability to validate huge datasets and pinpoint the exact corrupted records in constant space and linear time complexity up to the desired error probability. The main limitation of this method is the small probability of false positives which can be eliminated by various optimizations made to the bloom filter.[10].

K. Jeremy et al. (2016) presents the SQL relational model in terms of associative arrays and identifies the key mathematical properties that are preserved within SQL. The experimental results show that the associative arrays can provide a model for polystores to enhance the exchange of data and execution queries.[11]
L. Changqing et al. (2017) proposed efficient techniques for on SQL and NoSQL data transformation based on Espresso heuristic algorithm and depends on four related transformation steps.The final results show that this technique can be reduced the amount of memory usage and transformation execution time [12].

Statement of the Problem/Problem Definition Our focus in this paper is on the comparison between a well-known document-oriented type of NoSQL databases i.e. MongoDB with the Microsoft SQL Server as a relational database. We investigate the fundamental differences between these two database management systems in terms of performance of processing the queries. Our selected queries are run on the same dataset with the same number of records

There are lots of benchmarking reports available in internet and in research papers. Most of the benchmarking reports measure the overall database performance only by throughput and latency. This is an adequate performance analysis but need not to be the end. In this thesis we define some of the new perspectives which also need to be considered during NoSQL performance analysis.

CONCLUSION

The main significance is we will investigate the performance helps in determining how the data should be stored and managed. Our focus on this paper is to evaluate the performance between a well-known document-oriented type of NoSQL databases i.e., MongoDB with the Microsoft SQL server as a relational Database.

We will also compares the basic concepts and expressions used by two databases, as well as, performs some tests for evaluating the speed of these two databases in terms of insert operation, simple and complex queries.

REFERENCES:

1. D.Pritchett,"Base: An acid alternative," Queue 6,no.3 (2008): pp.48-55.
2. Dadapeer, N. M.Indravasan, and G. Adarsh“A Survey on Security of NoSQL Databases”. International Journal of Innovative Research in Computer and Communication Engineering. ISSN (Online): 2320-980, Vol. 4, Issue 4, April 2016.
3. A B M Moniruzzaman and S. A. Hossain. “NoSQL Database: New Era of Databases for Big data Analytics, Classification, Characteristics and Comparison”.International Journal of Database Theory and Application, ISSN: 2005-4270,Vol.6,No.4,August,2013.
4. F. Zhu, J. Liu, and L. Xu:“A Fast and High Throughput SQL Query System for Big Data”. International Conference on Web Information Systems Engineering (WISE), pp 783-788-2012.
5. S. Khan and Prof.Vanita Mane.“SQL Support over MongoDB using Metadata”, International Journal of Scientific and Research Publications,Volume 3, Issue 10, October 2013.
6. R. Lawrence. “Integration and virtualization of relational SQL and NoSQL systems including MySQL and MongoDB”. International Conference in Computational Science and Computational Intelligence (CSCI),volume 1, pp 285-290, March 2014.
7. M. Hanine, A. Bendarag and O. Boutkhoum,”Data Migration Methodology from relational to NoSQL Databases”, International journal of Computer, Electrical,Automation, Control and Information Engineering, Vol: 9, No: 12, 2014
8. M. Potey, M. Digrase, G. Deshmukh, and M. Nerkar.“Database Migration from Structured Database to Non-Structured Database”. International Conference on Recent Trends &Advancements in Engineering Technology (ICRTAET) 2015.
9. A. Abdullah and Q. Zhuge.“From Relational Databases to NoSQL Databases: Performance Evaluation”. Research Journal of Applied Sciences, Engineering and Technology,Vol: 11 No: 4 pp:434-439, 2015

10. W.L. Low,J.Lee,and P.Teoh.,"Conceptual Mappings to Convert Relational into NoSQL Databases". 18th International Conference on Enterprise Information Systems, January 2016.
11. A. Goyal, A. Swaminathan, R. Pande and V. Attar, "Crosplatform (RDBMS to NoSQL) database validation tool usingbloom filter," 2016 International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, 2016, pp.1-5.
12. J. Kepner et al., "Associative array model of SQL, NoSQL, and NewSQL databases," 2016 IEEE High Performance Extreme Computing Conference (HPEC),Waltham, MA, 2016, pp.1-9.
13. P. Sareen and P. Kuma, "NoSQL Database and its Comparison with SQL Database". International Journal of Computer Science & Communication Networks,Vol: 5 No: 5,pp 293-298, 2011.

