The study Comparative of NoSQL Databases

Pooja Sharma
Dept. of Computer Science, MUIT University, Lucknow

Dr. B.D. K. Patro
Associate Professor, Dept. of Computer Science, MUIT University, Lucknow

DOI Number: 10.14704/nq.2022.20.9.NQ44531 **Neuro Quantology 2022; 20(9):4607-4608**

Introduction

In this research work, broad categories of NoSQL data storage i.e. Column-oriented, Document- based, Graph-based and Key-value are compared. Several research works have compared real solutions of NoSQL data storage such as Cassandra, MongoDB, Neo4J or DynamoDB. Real solutions are compared based on very narrow features. In this research work, broad NoSQL storage categories are compared based on various characteristics such as storage, application area, advantages, flexibility, performance and scalability. Readers can select broad categories based on their specific applications by exhaustive comparison in this research work. Furthermore, the most popularly used real solutions from 120 solutions are selected in this research work. The reason for selecting these real solutions is to compare based on scalability, storage and applications. It is observed that query languages for NoSQL are not standardized due to different storage structures of NoSQL data storage techniques. Different data storage techniques use its own query languages.

The conclusion for this category of research work is as follows:

- Big data storage techniques are studied and analyzed from existing literature work. Several researchers have concluded that distributed storage, NewSQL and NoSQL data storage techniques are used for Big data storage.
- NoSQL data storage techniques- Column-oriented, Document-Based, Graph-Based and Key-value are studied and compared in this research work.
- The broad categories of NoSQL data storage are compared which is different from existing research work in which only real solutions are compared with narrow features.
- The most commonly used real solution such as MongoDB, CouchDB, BigTable, HBase etc. are selected to compare these solutions based on features such as scalability, applications etc.
- Traditional data storage techniques and Big data storage techniques are also compared. It is concluded that Big data storage techniques outperforms traditional techniques in terms of scalability, Object relational mapping, schema etc.
- It is also concluded that standard query language is not available for NoSQL data storage.

There is need for standardizing query language for Big data storage.

Types of NoSQL Databases Recently NoSQL database are generated by the huge growth of data mostly in web and mobile applications. If it is to be considered that social media web pages such as Facebook, LinkedIn and Twitter, which are dealing with thousands of terabytes of data, then it must be noticed that besides handling huge data volume, those systems still have to maintain latency, meaning that reading and writing are supposed to be responded immediately [3]. As previously mentioned there are many NoSQL types which recently have appeared with different performances; therefore, they are compared in terms of performance and verified how the performance are related to the database types [5]. In this Section thefollowing

NoSQL databases: Cassandra MongoDB, CouchDB, Hbase and SimpleDB are compared in details.

Cassandra

Facebook continues to be the most popular and the largest social media site that contains thousands of users of the system simultaneously using tens of mil lions of servers that are distributed in many data centers around the whole world. There is always a probability any server and any of network components may fail at any given time as any software system needs to be constructed in a way that to deal with failures immediately and effi ciency. To meet those Requirements reliability and scalability, Facebook Company has developed new NoSQL Database in 2007 that called Cassandra.

Big data mining

In this research work, it is observed that K-means can perform well for numerical data. Big data is collection of numerical as well as categorical data. K-means is not efficient for categorical data as it is based on distance measure using geometric space. K-prototype algorithm is implemented in the research work, which uses Euclidean distance for numerical and Hamming distance for categorical data. Intelligent splitter is proposed in this research work which splits numerical and categorical data before sending data to Mapper and Reducer. Approximately linear speedup is achieved using proposed approach.

The conclusion for this category of research work is as follows:

➢ K-Prototype algorithm is implemented on MapReduce which overcomes the limitations of K-means which can work efficiently for numerical data. It is concluded that K-Prototype response time on Mapreduce reduces significantly when deployed on multiple clusters. Speedup is calculated as 1 for K-Prototype on 1 cluster, 2.8 for 3 clusters and 4.6 for 5 clusters. Linear speedup is not achieved on multiple clusters. The reason is time is devoted to communication cost amongst clusters.

➢ Traditional data mining and Big data mining are compared based on several features such as scalability, technologies, structure of data etc. It is concluded that due to unstructured, categorical and large-dimensional data generated due to social networking sites, business transactions etc., there is need for Big data technologies and techniques to extract pattern from dataset.

CONCLUSION

Basically, the comparison shows NoSQL databases would not replace relational databases, but instead it will become a better option for specifi c types of projects. And no one of these NoSQL databases is best for all use cases. A user's prioritization of features will be different depending on the application, as well as the type of scalability and availability required. This survey may help the user to choose the most appropriate data store based on the use case, and some examples of applications that fi t well with the different data store categories. And, a storage NoSQL. As overall results in terms of optimization, NoSQL databases can be divided into two categories the databases optimized for reads and the databases optimized for updates. Thus, MongoDB optimized to perform read operations and high availability in an unreliable environment, while Colum Family databases, such as Cassandra and HBase have a better performance during execution of updates, but delivering low latency. Also, CouchDB has some limitations, such as only providing an interface based on HTTP REST. Concurrent read and write performance is not ideal.

Reference

1. Okman, L., et al. Security issues in nosql databases. in 2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications. 2011. IEEE.
2. Sakr, S., et al., A survey of large scale data management approaches in cloud environments. IEEE Communications Surveys & Tutorials, 2011. 13(3): p. 311-336.
3. Lakshman, A. and P. Malik, Cassandra: a decentralized structured storage system. ACM SIGOPS Operating Systems Review, 2010. 44(2): p. 35-40.
4. Jain, R., S. Iyengar, and A. Arora. Overview of popular graph databases. in Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on. 2013. IEEE