

A Critical Overview of NoSQL Databases

Manpreet Chopra

*Department of Computer Science
Punjabi University,
Patiala, Punjab, India*

Rajesh Kumar Bawa

*Department of Computer Science
Punjabi University,
Patiala, Punjab, India*

ABSTRACT

With the growth of Digital world, complexity in terms of volume, variety and velocity is being observed directly referring to 'Big Data'. This depicts the data collection which cannot be effectively managed using classical relational model like RDBMS because of its large growing size. To overcome such problem of handling, processing and managing large databases, NoSQL databases play a vital role in today's world and modern organizations. There are various types of NoSQL databases available which can be opted on specific requirements like scalability, application type etc. This paper is intended to make the users familiarize with three mostly used NoSQL databases : Cassandra, MongoDB and Couchbase and their comparisons on certain attributes like scalability, Performance and Reliability etc.

Key Words - NoSql Databases, Mongoddb, Cassandra, Couchbase.

I. Introduction

Databases are considered an important part of the organizations. With the constant growth of keep and analyzed knowledge, relative databases exhibit a spread of limitations, e.g. the constraints of measurability and storage, and potency losing of query because of the big volumes of data which makes storage and management of large databases difficult. NoSQL databases emerged as a problem solver with collection of latest options[1]. As compared with relational Databases, NoSQL DBs are versatile and capable of taking advantage of latest clusters. With the growing trend in computer architectures and modern databases which require features like eventual consistency, open source, effective scalability (vertical and horizontal), schema free, replication and storage of non-relational data; NoSQL databases are capable at addressing these requirements[2].

NoSQL databases are software products which store data without using SQL. These use Keys for storing data known as Key-value stores. These tend to be faster than RDBMS as these are non-relational and lack in SQL. NoSQL databases use BASE principle: Basically Available, Soft State, and Eventually Consistent. This principle is characterized by highly available data while sacrificing its consistency [3]. Few open source NoSQL databases are Cassandra used by Facebook, CouchDB in LinkedIn and HBase in Apache. With respect to technology, three popular NoSQL databases are available:

1. **Key Value Stores:** In this type of NoSQL databases, the data is stored in index values and retrieval of data is achieved by Key. Data can be unstructured or structured. For example: Amazon's Simple DB.

2. **Column Oriented Database:** For each record, instead of storing data in structured tables that consists of rows and columns like in traditional databases, these databases have a closely related data column. For example : Cassandra Databases used by Facebook.

3. **Document based stores:** The data is stored as documents for records instead of structured tables with uniform fields. User has a flexibility to use any length for any number of fields.

Types of NoSQL databases :

1) **Cassandra:** A Decentralized, highly scalable, eventually consistent DB

Cassandra is second-generation highly scalable distributed database. In 2008, It was open sourced by Facebook [4]. Currently, it is being developed under project named Apache Incubator . The system offers a high tolerance to fault , highly available and decentralized storage for data which gives flexibility of scaling up the system by adding hardware nodes to it. It is implemented on an "eventually consistent" model which is capable of maintaining consistency in data stores of the system. To offer more fault tolerance , at multiple nodes data is automatically replicated. This database supports the replication at multiple data centers and replacement of failed nodes . Cassandra is used at Facebook, Twitter, Cisco and more companies which have abundant active data sets. The largest production cluster consists of 100 TB of data in more than 150 machines.

2) **MongoDB**

MongoDB (“humongous”) is a document database which is designed to work easy and fast and is very scalable. It is highly suitable and ideal for website infrastructure. It is perfect for all types of Web content like wikis , blogs , messages and comments etc. And additionally ideal for product information ,sessions and user profiles. It is not good for transactional systems like banking where perfect durability is required. It is well fitted for real-time applications which includes requirements like real time reporting and agile projects where changes in the underlying database schema is observed often. It is not a good practice to use software having a complex (multiple objects) transactional system. With in MongoDB, there is an fascinating combination of modern Web used semantics with well proved database techniques. In certain ways, MongoDB is similar to MySQL than to other NoSQL” databases: It consists of a query optimizer with ad-hoc queries and a customized network layer. To achieve good speed and efficiency , it also offers a user to organize document into collections which are similar to SQL tables. However, to get optimum performance and horizontal scalability , MongoDB is incapable to support transactions that span across multiple collections. It gives privilege to perform atomic operations on one object, but does not provide flexibility to modify objects atomically from two collections. MongoDB stores data in BSON with a good binary representation having more data types which is a efficient JSON document. BSON documents are very fast at persisting large number of data structures which including maps, associative arrays, structs and also supports any dynamic language objects. By Using both MongoDB and BSON, data can be stored directly in database. It is also schema-less database in which any fields can be added without making much in database. Fields can be added very quickly, which is demand of agile environments.

3) **Couchbase**

NoSQL document oriented database which is based on distributed multi-model suitable for interactive applications. It uses JSON format for storage of data with leverage of SQL to make applications easier. Couchbase can use for vertical and horizontal scaling. It is helpful in providing consistency and partial tolerance. High redundancy and availability is supported by two types of replication in Couchbase. One type of replication is done for same cluster and another type is used for testing between the clusters which is called Cross Data Center Replication(XDCR).

II. Literature Review

Vatika Sharma et al. [1] had done survey about NoSQL, its background, fundamentals like ACID, BASE and CAP theorem. The main aim of this paper is to give an summary of NoSQL databases, regarding however it's declined the dominance of SQL, with its history and features. It also describes its fundamentals that type the bottom of the NoSQL databases like ACID, BASE and CAP theorem. ACID property isn't utilized in the NoSQL database thanks to data consistency therefore we get to understand however SQL lags data consistency.

A B M Moniruzzaman et al. [2] give - classification, characteristics and analysis of NoSQL databases in massive data analytics. This report is meant to assist users, particularly to the organizations to get an independent understanding of the strengths and weaknesses of assorted NoSQL information approaches to supporting applications those method Brobdingnagian volumes of information. This study report's motivation is to offer an absolute understanding of the strengths and weaknesses of varied NoSQL database approaches to support applications that method vast volumes of data; similarly on provide a worldwide summary of non-relational NoSQL databases.

João Ricardo Lourenço et al. [3] had done the performance comparison of various NoSQL databases. In this paper, author had gathered a brief and up-to-date comparison of NoSQL engines, their most useful use case situations from the programmer viewpoint, their benefits and disadvantages by measuring the presently on the market literature. In this paper, author concluded that though there are a spread of studies and evaluations of NoSQL technology, there's still not enough data to verify however suited every non-relational database is during a specific state of affairs or system. Moreover, every operating system differs from another and every one the required functionalities and mechanisms extremely have an effect on the database selection. typically there's no chance of clearly stating the simplest information answer.

K.Chitra et al. [4] had done Study on primarily on the market, scalable and Eventually Consistent NOSQL Databases. This paper analyses the requirement of following generation data storage that is that the need of

the present large-scale social networking or cloud applications additionally author analyze the capabilities of assorted NOSQL models like BigTable, Cassandra, CouchDB, generator and MongoDB. During this paper author conclude that NOSQL databases usually method data faster than relative databases. Developers typically don't have their NOSQL databases supporting ACID properties, so as to extend performance, however this will cause issues once used for applications that need nice exactness.

Veronika Abramova et al. [5] had reviewed the performance of various NoSQL databases. In this paper, five most well liked NoSQL databases square measure evaluated: Cassandra, HBase, MongoDB, OrientDB and Redis. Author compare those databases in terms of question performance, supported reads and updates, taking into thought the everyday workloads, as pictured by the Yahoo! Cloud Serving Benchmark. This comparison permits users to choose the most appropriate database in step with the precise mechanisms and application desires. Author concludes that MongoDB, Redis, and OrientDB square measure best databases to perform scan operations, whereas Column Family databases, Cassandra and HBase, have a far better performance throughout execution of updates.

III. Comparison Study of various quality attributes for Different NoSQL databases

Attributes	MongoDB	Cassandra	Couchbase
Availability	-	+	+
Scalability	-	+	+
Read Performance	++	-	-
Write Performance	-	++	+
Durability	+	+	-
Reliability	++	+	-

Where - = poor, + = average and ++ = good

IV. Conclusion

It can be concluded that NoSQL databases process the data faster than relational databases because of their simpler model. There have been a variety of studies and evaluations can be made on basis of various quality factors for each of NoSQL databases. It would help the users to opt for right database as per requirements. There are large number of NoSQL databases available, so, the choice of well suited NoSQL databases becomes difficult. The choice can be based on certain parameters like scalability which will help the user and organizations to choose right database from the available ones. Also, it can be concluded that necessary functionalities and mechanisms effect the choice of database.

V. References

- [1] V. Sharma and M. Dave, "SQL and NoSQL Databases," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 2, no. 8, pp. 20–27, 2012.
- [2] A. B. M. Moniruzzaman and S. A. Hossain, "NoSQL Database : New Era of Databases for Big data Analytics - Classification , Characteristics and Comparison," *Int. J. database theory Appl.*, vol. 6, no. 4, pp. 1–14, 2013.
- [3] J. R. . Lourenço, V. . Abramova, M. . Vieira, B. . Cabral, and J. . b Bernardino, "NOSQL databases: A software engineering perspective," *Springer Adv. Intell. Syst. Comput.*, vol. 353, no. 6, pp. 741–750, 2015.
- [4] K. Chitra and B. Jeevarani, "International Journal of Advanced Research in Computer Science and Software Engineering," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 5, pp. 991–996, 2013.
- [5] V. A. A, J. B. A, and P. F. B, "Which NoSQL Database? A Performance Overview," *Open J. Databases*, vol. 1, no. 2, pp. 17–24, 2014.