Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 12, December-2020

Overview of NoSQL Databases and A Concise Description of MongoDB

V. Sumalatha
Research Scholar, CSE Department,
UCE, OU, Hyderabad, Telangana, India.

Dr. Suresh Pabboju
Professor and Director IQAC
CBIT, Hyderabad, Telangana, India.

***Abstract*: Larger data volume of various applications development leads to the today's IT growth, now a days it is observed, with Relational database we cannot work on large data to maintain high volume data base due to the strict constraints on data structure, data relations and so on. Various formats of different industries including unstructured data has to be stored and processed in the databases. Hence, NoSQL types are the solutions for various issues of large data base. Largely NOSQL (Not Only SQL) is sought-after due to the advantages such as horizontal scalability and flexibility. This paper, discusses about different types of No SQL database, including MongoDB, Couchdb, HBase, Cassandra, etc.**

Keywords—Relational Database, Big Data, NoSQL, MongoDB, Couchdb, HBase, Cassandra.

I. INTRODUCTION

In previous technology, many applications widely used relational databases to collect information. The relational database resembled the basic structure in the form of a table which have records along with columns. In that tables the data can be manipulated edited, evaluated using a well-defined structured queries and maintained as per requirement [1]. Due to the development in technology in now a days there is a necessity for large database, different format and online support easy use, flexibility there is a need of new technology to support all the above criteria along with the changing needs. There is a technology based on not a relational data which fulfills all the above criteria called NoSQL[2]. In information system today NoSQL database have its unique popularity. When a database designer unable to decide which database is used non-relational or relational with his collected data which is large in size, data in different formats then the designer has to choose NoSQL database[2].

The aim of this paper is to describe the main features and characteristics of MongoDB technology especially for NoSQL as a new solution over relational databases. This paper is organized as follows. The introduction is given an elementary introduce of the traditional relational databases and NoSQL databases. In the next section continues the part where mention some achieved results from different researchers. In this section also mentioned some related work on this topic. A detailed introduction for the types of NoSQL databases. In the third section explanation of Mongodb and result of some queries. Finally, in the fourth section conclusion is concluded [3].

II. RELATED WORKS

*A. SQL and NoSQL databases:*The relational data model(like as SQL databases) provides specific mechanisms in order to handle the data consistency within the system, to manage concurrency control ensuring that correct results for concurrent operations are generated and finally, to guarantee transaction ACID properties defined as follows [4]:

(1) Atomic: Everything in a transaction succeeds or the entire transaction is rolled back.
(2) Consistent: A transaction cannot leave the database in an inconsistent state.
(3) Isolated: Transactions cannot interfere with each other.
(4) Durable: Completed transactions persist, even when servers restart.

Any user can perform one or many database tables transaction operation by using SQL with a special property called ACID [4].

Non-relational data model (like as NoSQL databases) is following BASE (Basic Availability, Soft-state, Eventual consistency property. In general, it is harder to develop software in the fault-tolerant BASE world compared to the particular ACID world, but Eric Brewer's CAP theorem defines such configuration as a requirement in a scale-up approach. There is a difference between ACID and BASE. The CAP theorem introduced by Brewer in 2012 providing 3 main guaranties:

(1) Consistency: All nodes see the same data at the same time.
(2) Availability: A guarantee that every request receives a response about whether it succeeded or failed.
(3) Partition tolerance: The system continues to operate despite arbitrary partitioning due to occurring network failures [4].

NoSQL database may be good enough when not compare to the relational database because the relational database can collect data in the form of a table and can store data that is structured as well [5].NoSQL stands for "Not Only Sql." But this doesn't mean that this term opposes SQL, but that is not the case. NoSQL systems can coexist relational and non-relational databases. They are suitable in applications where there is a large amount of data is involved. Data in this case is structured, unstructured or semi-structured. There are different types of NoSQL databases. Some of these include Cassandra, Mongo DB,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 12, December-2020

Couch DB and HBase among others. All of them differ in terms of structure and usage as per requirement. Depending on these factors, such databases have been further classified into various categories. Since NoSQL databases cannot be used interchangeably, in this paper practical usages of them or explained as per these categories. They are not compatible like relational databases. In situations where there is a need certain category of NoSQL you cannot use a different one. As per need each database type is having its own specialized characteristics in that particular environment so the designer can choose correct database type to implement.

There are four NoSQL database categories:

- Key-Valued Stores
- Column Family Stores
- Document Databases
- Graph Databases.

*B. NoSQL DATABASES:*The different types of NoSQL database are:

1) Key-Value databases: This is one of the popular NoSQL databases. This type of databases allows the user to do operations on the database based on a key-value pair. As key-value stores depend mostly on primary-key access, they have extraordinary performance compared with other NoSQL databases and can be easily scaled.
CouchDB: Couchbase is some of the commonly used key value databases.It is an Apache Software Foundation Product and inspired by Lotus Notes, is also an open source document based on NoSQL database which focus mainly on easy use. It is a single noded database, working exactly like other databases [6]. It generally starts with the single node instance but can be uniformly upgraded to cluster. It has the capability that allows the user to run single database on multiple servers or VMs. A cluster based CouchDB provides high capacity and availability as compared to single node CouchDB. It uses Erlang, a general purpose language.CouchDB is lockless which means, there is no need to lock the database during writes. The documents in this database also use HTTP protocol and JSON, along with the ability to attach non-JSON files to them. So, CouchDB is fully compatible with any application or software which supports JSON format [6].

2) Document databases: In this type of databases documents are the fundamental formats. It stores and recovers XML, JSON, BSON documents. All data related with a given object as a single instance in the database are stored. Because of this document stores are used widely for web application programming. They provide for embedded documents, which are self-describing, hierarchical tree structures that often comprise of maps, collections, and scalar values. Document databases like MongoDB has very powerful and rich query language and constructs, with which a simpler migration from relational databases is possible. The popular document databases are MongoDB, CouchDB.

3) Column family stores: In these databases the data will be represented and stored in the form of column families as rows. These rows have number of columns related with the key of that particular row. The collection of data which is related and accessed together. Each column family can be contrasted to a container of rows in an RDBMS table. In RDBMS systems, the key is used to recognize each row (record). Here each row comprises of numerous columns (fields). Cassandra is one of the prominent column-family databases and other important column value databases are HBase and Cassendra.
HBase: It is a distributed column-oriented data store built on top of the HDFS.Data is logically organized into tables, rows and columns .HBase files are internally stored in HDFS.
Cassendra: Apache Cassandra is an open source, distributed and decentralized/distributed storage system (database), to manage very large amounts of structured data spread out across the world. It provides highly available service without single point of failure [8].
Characteristics of Cassandra [9]:

- It is a column-oriented database.
- It is highly consistent, fault-tolerant, and scalable.
- It was created for Facebook and was later open sourced.
- The data model is based on Google Bigtable.
- The distributed design is based on Amazon Dynamo.

4) Graph Databases: Graph databases facilitate the storage of entities and connections or relationships between them. Another name for entities is nodes, which have properties associated with it. A node is nothing but occurrence of an object in the application.. The connection between the nodes show the relations and are known as edges, which can have properties. Edges show the directional significance and nodes are organized based on these relationships. With this arrangement, the interesting analytics framework that enables knowledge discovery through information retrieval and filtering from document based NoSQL focusing on terms and topic mining [9].
Neo4j: It is the most popular open source Graph Database, developed by using Java technology. It is highly scalable and schema free (NoSQL).

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 12, December-2020

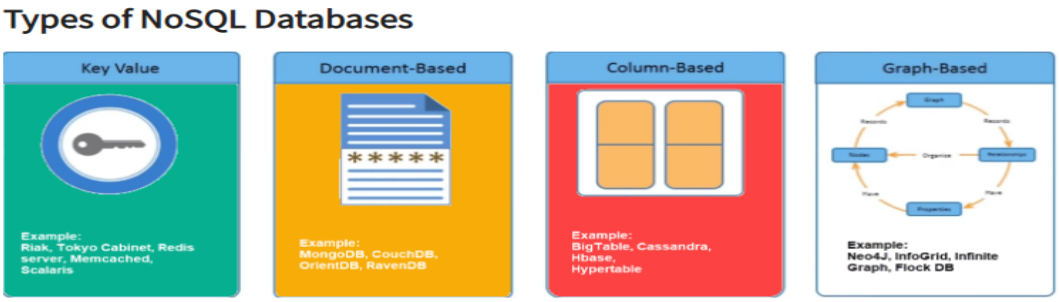

Fig 1. Types of NoSQL Databases

III. MongoDB

A. *MongoDB:* MongoDB is an open-source document database developed by MongoDB, Inc. MongoDB stores data in JSON-like documents that may vary in structure. Related information is stored together for fast query access through the MongoDB query language. MongoDB uses dynamic schemas, meaning that you can create records without first defining the structure, such as the fields or the types of their values. The record structure can be changed simply by adding new fields or deleting existing ones [7]. This data model gives you the ability to represent hierarchical relationships, to store arrays, and other more complex structures easily. Documents in a collection need not have an identical set of fields and denormalization of data is common.

B. *Features of MongoDB:*

Fig 2. Features of MongoDB

1. MongoDB provides high performance. Most of the operations in the MongoDB are faster when compared to relational databases.
2. MongoDB provides auto replication (return) feature that allows you to quickly recover data in case of a failure.
3. Horizontal scaling is possible in MongoDB because of sharing. Sharding is partitioning of data and placing it on multiple machines in such a way that the order of the data is preserved.

Horizontal scaling vs. vertical scaling:

Vertical scaling means, adding more resources to the existing machine while horizontal scaling means adding more machines to handle the data. Vertical scaling is hard to implement, on the other hand horizontal scaling is easy to implement. Horizontal scaling database examples: MongoDB, Cassandra etc.

4. Load balancing: Horizontal scaling allows MongoDB to balance and maintain the load.
5. High Availability: Auto Replication improves the availability of MongoDB database.
6. Indexing: Indexes are used to recognize and locate the data quickly locate data to avoid to search each and every document in a MongoDB database. It improves the performance of operations performed on the MongoDB database [10].

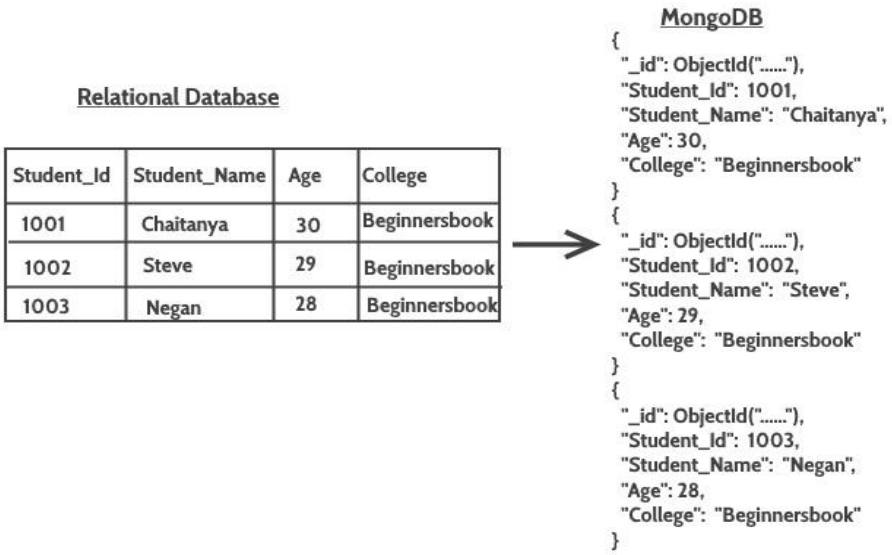

Fig 3. Difference between Relational Database and MongoDB

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 12, December-2020

C. *Data Storage in MongoDB:* Data Storage in MongoDB in JSON (Java Script Object Notation) format is based on a subset of the Java Script programming language. It is a light weight data interchange format. JSON supports all the fundamental data types, numbers, strings, and Boolean values, as well as arrays and hashes [3]. Document databases such as MongoDB use JSON records in a conventional database. Here is a document in order to store records, just as tables and rows example of JSON document: Employee example

{
"_id": ObjectId ("01"),
"Name": "Radha",
"Depart": "CSE",
}

D. *Screen Shots: Execution of Queries in MongoDB*

The execution of the queries in MongoDB to create a database, to insert a document, to search a document, to delete and to update documents are shown in the following screen shots Fig 4 and Fig

5.

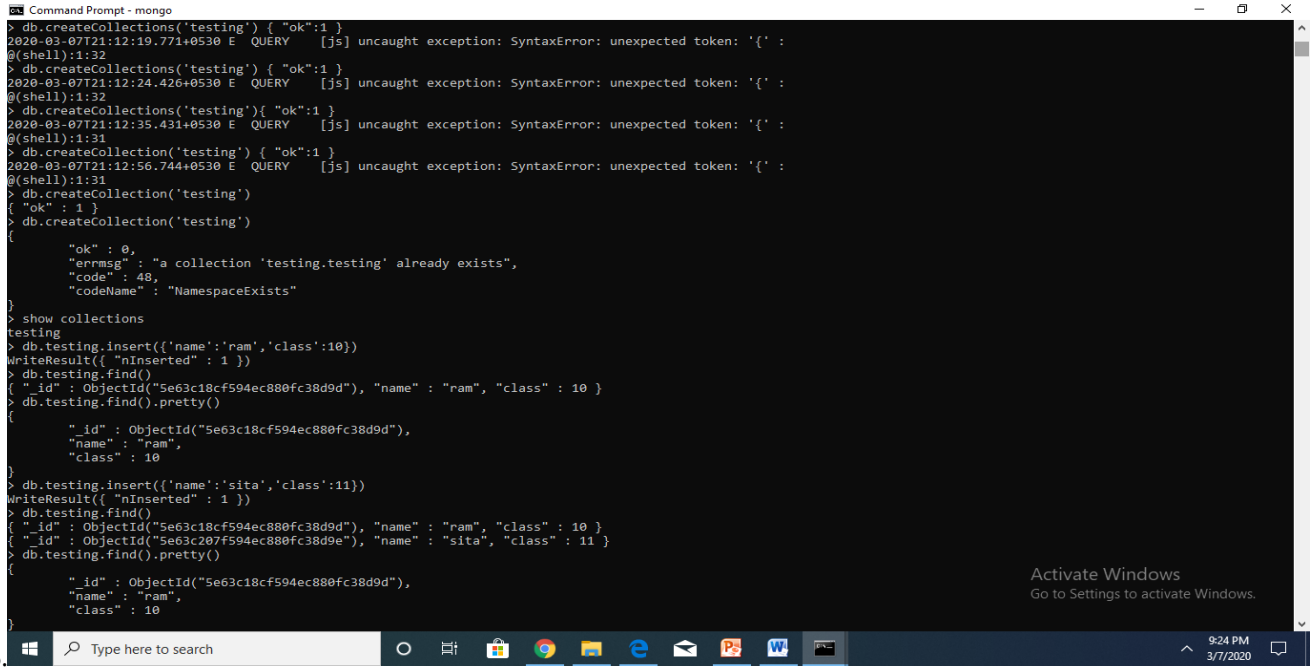

Fig 4. Insert and Find operation

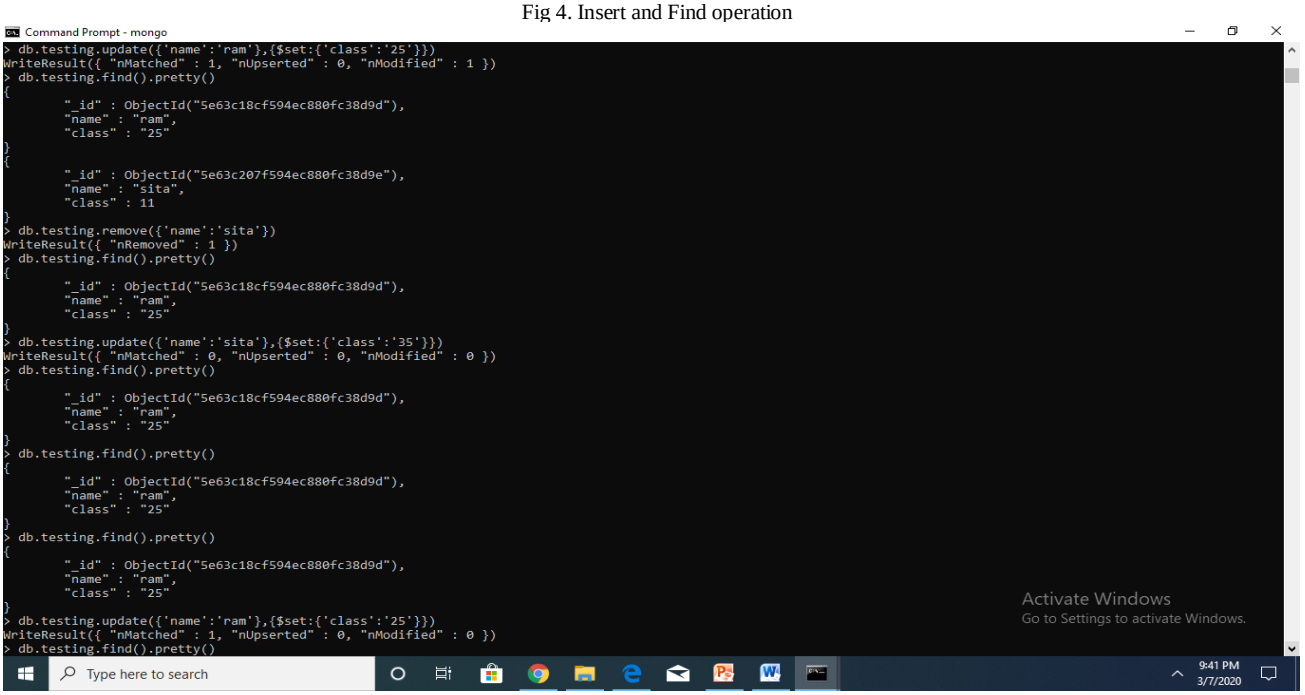

Fig 5. Remove and Update operation

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 12, December-2020

IV. CONCLUSION

NoSQL databases have schemaless in nature. This property of NoSQL databases allows including fields without making any changes in the structure. To deal with the massive growth in structured, semi structured and unstructured data, NoSQL databases are very useful. This paper highlights some features of document databases particularly MongoDB. The data representation in JSON format in MongoDB is explained along with the screen shots of the executions of few queries in MongoDB . As a future scope of this study, the query performance of MongoDB can be compared with Other NoSQL database. There may a possible way to implement a new system that combines system of both kind of technologies, where relational and NoSQL databases can be configured together within same database management and collaborate without obstacles furthermore, they can combine the benefits of each technology solution to create a more effective and more powerful system.

REFERENCES

[1] Chevalier, Max and El Malki, Mohammed and Kopliku, Arlind and Teste, Olivier and Tournier, Ronan Implementation of Multidimensional Databases with Document-Oriented NoSQL. (2015) In: 17th International Conference on Big Data Analytics and Knowledge Discovery (DaWaK 2015) in 26th DEXA Conferences and Workshops, 1 September 2015 - 4 September 2015 (Valencia, Spain).
[2] Wittawat Puangsaijai and Sutheera Puntheeranurak "A comparative study of relational database and key-value database for big data applications" .5th International Electrical Engineering Congress, Pattaya, Thailand, 8-10 March 2017.
[3] Kosovare Sahatqija, Jaumin Ajdari, Xhemal Zenuni, Bujar Raufi, Florije Ismaili "Comparison between relational and NOSQL databases" MIPRO 2018, May 21-25, 2018, Opatija Croatia.
[4] Hadi Hashem; Daniel Ranc "Evaluating NoSQL Document Oriented Data Model" Publication Year: 2016, Page(s): 51 - 56]. Published in: 2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW).
[5] Rupali Kaur, Jaspreet Kaur Sahiwal "A review of comparison between NoSQL Databases: MongoDB and CouchDB" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-6S, March 2019.
[6] Niteshwar Datt Bhardwaj and Mohali "comparative Study of CouchDB and MongoDB – NoSQL Document Oriented Databases". International Journal of Computer Applications (0975 – 8887) Volume 136 – No.3, February 2016.
[7] Sundhara Kumar K.B, Srividya, Mohanavalli.S "A Performance Comparison of Document Oriented NoSQL Databases" IEEE International Conference on Computer, Communication, and Signal Processing (ICCCSP-2017).
[8] https://www.tutorialspoint.com/cassandra/cassandra_introduction.htm.
[9] https://intellipaat.com/blog/what-is-apache-cassandra/.
[10] https://beginnersbook.com/2017/09/introduction-to-mongodb/.