Analysis on NoSQL: MongoDB Tool

Deepa Suresh Wahane[1], Prof. Mayuri Dhondiba Dendge[2]

[1]Student, [2]Assitant Professor

Bharati Vidyapeeth's Institute of Management & Information Technology, Navi Mumbai, Maharashtra, India

ABSTRACT

The Relational Database System is basic database used from many decades. Since Mysql, Oracle are used for relational kind of databases but Nowadays structure of data has been changed. The problem of Data storage has been raised. Different form of data is available i.e. multimedia databases which is difficult to store. MongoDb can be future alternative for Relational Database. This paper gives overview of NoSQL database MongoDb. This paper is evaluation of NoSQL classification, features and benefits. This paper include Case study on MongoDb which consists of MongoDB web Shell, Architecture and Storage engines and protocols that are included in MongoDB web shell.

Keywords: NoSQL, MongoDB, Architecture of MongoDB, Non-relational Database, craigslist

I. INTRODUCTION

In NoSQL Databases, End user uses various databases to store and maintain the data. Relational databases used in most of the application, but they have disadvantage of storing and operating. To handle complex databases such as rapidly changing data and slowly changing data the relational databases are not effective. For handling unstructured data NoSQL term is used. NoSQL is the methodology which consists of several tools to handle unstructured data that can be used for updating and changing the data in real scenario. NoSQL database is mainly based on types of data, time for loading data and throughput of data. NoSQL can be termed as 'Not-Only SQL'. The advantage of NoSQL is processing unstructured type of data like social media, multimedia, emails. NoSQL can handle more amounts of data in within minutes.

NoSQL Mainly follows BASE properties i.e. Basically Available Soft State and Eventually Consistent. NoSQL don't have Tables i.e. Rows and Columns. It works on Data and Key Identification. It is based on the requirement of user or application, the correct and appropriate NoSQL database is used. Each database has its own data model structures and functions.

II. Types Of NoSQL Database:

Document Database: It has a key-value pair,

The key is related to document. In document database XML, JSON, BSON, YAML such languages are used.

Graph Database: Graph Database used for networks of data. Node contains entities and edges show the relationship. It can be directed/ non-directed of hyper edged means no edged between nodes. Facebook is efficient and scalable for this kind of database. Ex. Neo4j

BigTable Database: Simplest type of database. Every single value is stored in attribute & associated with its value. The data replication can be used. With the help of Timestamp user can obtain information about data. It based on map-reduce framework. Ex. Google, Cassendra, HBase.

Columner Database: It is familiar with RDBMS. It uses Row-Wise storage to store column data together instead of row. Column is single in every disk block gibe advantage to the faster querying.

III. BENEFITS OF NOSQL OVER RDBMS

Compare with RDBMS, the NoSQL is more scalable, more flexible & gives higher performance.

1. Dynamic Schema: In NoSQL database, data can be stored into the database without requiring any schema. It makes faster development and integration of code and takes less database administrator time.

2. Auto-sharding: In structured, relational database scales vertically, a single server is been used to stored data. The transactions are done using cross table and joins. This can be get expensive, limited to scale and can create more no. of failures. Instead of using single server, the multiple servers are for getting more capacity, this is called as sharding. NoSQL database, support for auto-sharding means they automatically spread the data across multiple servers without requiring server pool. When any servers suffer from failure, it can be replaced automatically without requiring special application.

3. Replication: Most of the NoSQL databases also support automatic replication to maintain reliability & availability of data. NoSQL database are self-healing, which gives automatic fail over & recovers as well as availability of database with geographical regions to enable data localization.

4. Integrated Caching: NoSQL has been developed with many caching capabilities, it creates the separate caching layer, same of the NoSQL database offered manageable and integrated database management for higher throughput and lowest delays.

IV MONGODB

MongoDB is open source NoSQL database which is document based database. The language used in MongoDB is c++. 10gen began developing MongoDB database in 2007 as a 'Platform as a service' which is known as MongoDB.inc. In 2009, it was introduced as an open source database server.

a. Features of MongoDB

1. Ad-hoc queries: Field, range queries and regular expression supports in MongoDB for searching. This query gives specific patterns and also supports Javascript functions.

**2. Indexing** : The primary & secondary indexes are available in MongoDB document.

3. Replication: MongoDB support for replication of document. It can be consists of two or more replicas. All read & write functions are done in primary replica by default. Using built-in replication, Secondary replica makes copy of primary replica. When any primary replica fails, through election process, the primary replica has been decided from secondary replicas. Secondary replicas server to read function.

4. Load Balancing: MongoDB uses horizontal sharding. MongoDB can be used over more than one server, balancing the load or replicating data which gives throughput when system fails.

5. File Storage: MongoDB used for storing files with load balancing and replications features. This is called grid file system.

6. Aggregation: Map-Reduce are used for aggregating data in mongoDB. In SQL, usually used group by clause for getting these kinds of results aggregation function is used. The aggregation functions are keeping together to form a pipeline.

b. The MongoDB Web Shell
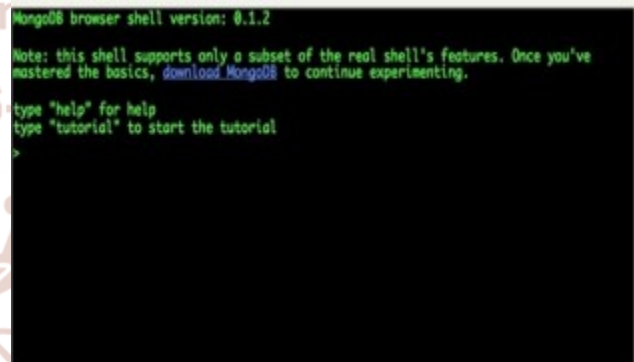
The MongoDB web shell is application for described the features of MongoDB terminal. It mainly consists of 3 major parts:

User can access new MongoDB web shell through any of the web browser. Most of the code is written in Javascript and executed on the web browser. The browser-based shell uses the back-end Mongodb instance. The Back-end gives us:

➢ For Session management, the sandboxing is used and accessing the multiple resources for the database.

➢ Provides a framework to preload the data to the assessable resource.

➢ Verifying the state of data.

Below three components helps us to familiarize the concept and uses of mongodb web shell

1. Try.mongodb.org

The primary goal of this project is replacing existing browser to try.mongodb.org the new version provides the functionalities of JavaScript code and new features for mongodb. This provides new features of JavaScript and new tutorial of system.

2. 10gen Education

MongoDB web shell can be used on 10Gen Education platform where users can develop programmes for Mongodb very quickly. The assignments are given to the students who want to learn MongoDB and they can solve the assignment with individual installation of MongoDB with the help of web shell.

3. MongoDB API Documentation

The MongoDB Web Shell with API reference pages, which will part of documentation. This will be used for fast testing for detailing on the page and evaluating its effects comparing data sets.

c. Storage Engines of MongoDB

Now days, user dealing with massive database which is also complex in nature so it's needed to exclude the redundancy.

Using plug-gable storage engines, we can achieve the new applications storage problems increasing capability of application and using a specific hardware with the help of storage engine, it can reduce the complexity of operations which are done of MongoDB, so that users can access the multiple applications across different application, so that they achieve data modeling, scaling of data, providing security to data, providing security to data and provides different tools. Here we can run multiple storage engines within in replica set makes easy to manage the data. In the MongoDB, the data automatically migrates between primary and secondary replica set members, they are independent of data format.

d. MongoDB architecture
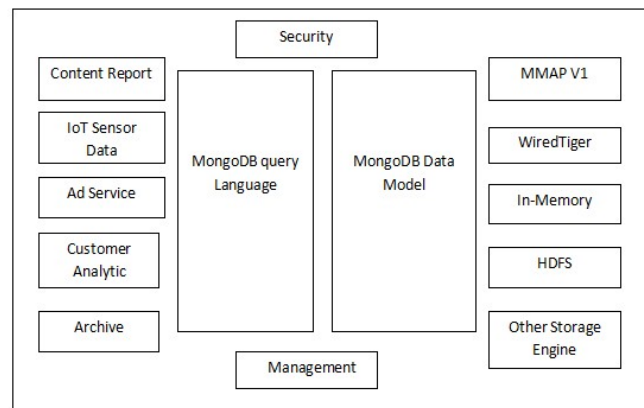
MongoDB 3.0 supports two data storage engines

➢ **MMAPv1** : This is improved version of the storage engine used for mainly concurrency control.

➢ **WiredTiger**: WiredTiger storage engine, the benefits where the hardware utilization, higher output and it.

Above are the data storage engines can be exist in single replica set, which are managed by MMS(MongoDB Management service) RockDB engines provide key-valued pair of data storage. HDFS storage engine and fusionIQ engine used for file system.

Wired Tiger storage engine, it is new storage engine for MongoDB which is developed by Berkley, the most deployed software in the World. Its scalability works on multi-processor architecture. It uses various techniques ex. Message passing algorithm, lock free algorithm, and hazard pointers and so on. To reduce the overhead on CPU and Disk I/O management, it uses particular file format and compression technique. After updating the versions of wiredTiger, the storage and data of replica will not disrupt.

e. Features of New Driver For MongoDB

➢ **Wire Protocol Compression**:

MongoDB used the wired protocol compression oplog data can be compressed 20x allowing replica's four times faster. In this process, server uses the snappy algorithm which will give benefit with speed and compression of data.

Here, we requires the driver which compress the conversions which done with servers some of application uses snappy algorithm, but zb1b algorithm, the most widely used and easily available, when driver connects with server they should save compression format. Between the client and server can be avoided with the help of compression techniques.

f. Drivers for MongoDB

1. OP.msg
This is one of the wire protocol is OP_MSG. This will give high-performance with replacement of old wire protocol.
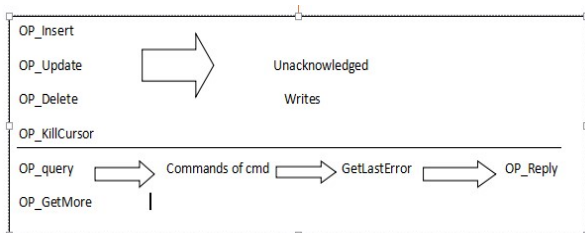
2. Ye Olde Wire Protocol
This protocol holds three components write messages, unacknowledged and a disposing cursor message.

Two messages are generate at the server side – one for create cursor with query and another get more results from cursors.
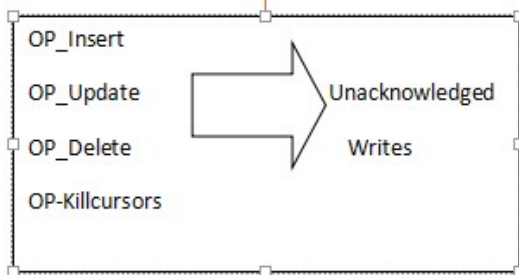
After innovating protocol, new messages are added into ye-olde protocol i.e. commands. We are reusing OP_Query, which explains query about cmd commands. Users want to get acknowledged about their writes, which implement and called after the getLastError command.
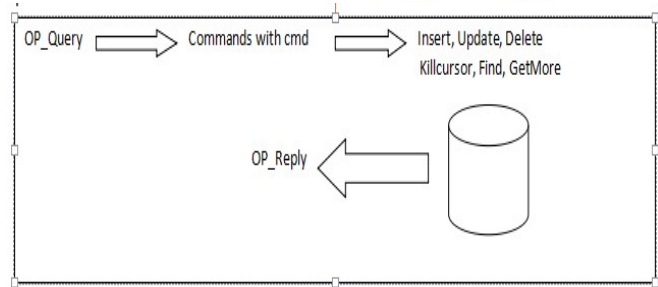

Ye-Olde wire protocol's OP_Insert msg

Inserting the document in this protocol is very easy and efficient. Message header followed by documents without delimiter.

3. Middle Wire Protocol
In the MongoDB 3.2 version, Middlewire protocol is been used in which all the process has commands.
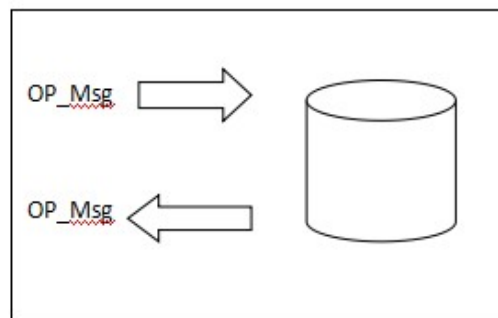

It has features of uniformity and flexibility which helps us to enhance to add new features into the MongoDB. It has some disadvantages:

1. It uses OP_Query on cmd.
2. After every writing it use acknowledgment
3. Efficiency is less.

4. Modern Wire Protocol
Modern wire protocol, release in 3.6 version of mongoDB with new protocol.

In this, client and server uses same type of message. This new message type added the combination of old and new types of wire protocols. It is designed by Mathias Stearn, which helps to create a robust protocol as compare to new protocols

OP_Msg contains:

Header	Flags	Body	Document	....	Checksum

g. Case Study on MongoDB: Craigslist

Craigslist is popular job posting website which are used about 570 cities, 1.5 Million new jobs are posted every day. Craigslist holds millions of records in

different formats. Craigslist uses MySQL to stored information, but later on the management and flexibility of data cannot be reached, then they migrate over 2 billion of data into mongoDB to achieve scalability and flexibility. MySQL was used for data storage, but does not gives scalability. MongoDB has built in scalability, so Craigslist chooses MongoDB to achieve scalability. In MongoDB, each post which are included in record of Craigslist database that are saved as single document so that are saved as single document so that mongoDB can save these changes without any high cost migration

VII. Conclusion

In this Research Paper, We have gone through NoSQL databases and its type, MongoDB with its architecture and drivers. From above analysis we can conclude that, NoSQL database can perform more better and fastly than SQL database. In NoSQL, mostly JSON language is used for CRUD functions, so that it become much easy for updating documents. The compression of data has been done in NoSQL database automatically, so that it will take less space to store and less time to access the data from database. The SQL database can be transferred to NoSQL database, that's why many companies migrate there data into NoSQL database.

VIII. REFERENCE

1) Divya Chauhan and K. L. Bansal, "Using the Advantages of NOSQL: A Case Study on MongoDB", International Journal on Recent and Innovation Trends in Computing and Communication Volume: 5 Issue: 2

2) Mr. Sushil Soni ,Mr. Mayuresh Ambavane, Mr. Shamal Ambre , Mr. Shirshendu Maitra, "A Comparative Study: MongoDB vs MySQL", International Journal of Scientific & Engineering Research Volume 8, Issue 5, May-2017

3) Mongodb architecture, https://www.mongodb.com/mongodb-architecture

4) The mongodb web shell, https://www.mongodb.com/blog/post/the-mongodb-web-shell

5) New Driver Features for MongoDB, https://dzone.com/articles/new-driver-features-for-mongodb-36

6) MongoDB Case study, https://www.mongodb.com/post/15781260117/mongodb-case-study-craigslist