A Study on Mongodb Database

Kavya. S

M Tech. Student

Department of Computer Science

Mount Zion College of Engineering, A. P. J. Abdul Kalam Technological University, Kadammanitta P.O, Pathanamthitta, Kerala, India

Abstract— This paper trying to focus on main features, advantages and applications of non-relational database namely Mongo DB and thus justifying why MongoDB is more suitable than relational databases in big data applications. The database used here for comparison with MongoDB is MySQL. The main features of MongoDB are flexibility, scalability, auto sharding and replication. MongoDB is used in big data and real time web applications since it is a leading database technology.

Key words: NoSQL, MongoDB, auto sharding, aggregation

I. INTRODUCTION

Relational database management systems came into existence since 1980's.They are a common choice of storage of information in new databases used for financial records, manufacturing and logistical information personnel data and other applications. They work efficiently when they handle a limited amount of data. Due to the emergence of applications that support millions of users simultaneously an appropriate database is required. To handle huge volume of data traditional relational database is inefficient. To overcome the difficulty in handling huge volume of data ,the term NoSQL was introduced by Crlo Strozzi in 1998.It refers to non-relational databases. More recently,the term has received another meaning namely Not Only SQL.The main advantage of NoSQL database is that it can handle both unstructured(e-mail,multimedia,social media) and semi structured data very efficiently. Mainly there are four categories of databases namely Key –value store, document store, column oriented and graph database. MongoDB is a cross platform document oriented database, first developed by the software company MongoDB Inc., in October 2007 as a component of planned platform as a service product. The company shifted to an open source development model in 2009.Since, then MongoDB has been adopted as a backend software by a number of major websites and services. These include Craigslist, eBay, Foursquare and Newyork Times. MongoDB is written in c++ and provides high availability, easy scalability and better performance. MongoDB works on the concept of collection and document. A database is a physical container for collections. Collection is a group of MongoDB documents. It is equivalent to RDBMS table. MongoDB database contain multiple collections. A document is a set of key-value pairs. Documents have dynamic schema. That means, documents in the same collection do not need to have the same structure and common fields. MongoDB supports dynamic queries on documents using a document based query language that is nearly as powerful as SQL.It stores the data in the form of JSON documents. Auto sharding, replication and high availability are the main features of MongoDB.It is commonly used in big data, content management and delivery, mobile and social infrastructure, user data management and data hub. MongoDB supports different data types such as String, Integer, Boolean, Double,Min/Max keys, Arrays, Timestamp, Oject, Null, Symbol, date, Object ID, Binary data, code and regular expressions.

II. MONGODB DATA MODEL

MongoDB stores data as documents which are in the BSON format. BSON is binary representation of JSON document. Documents having similar structure are organized as collections. Collection is analogous to a table in relational database. Documents and fields in MongoDB are represented using the terms row and columns respectively in MySQL. The difference between the relational database, MySQL and non-relational database ,MongoDB is that in relational database information for a given record is usually spread across many tables, whereas in MongoDB the documents tend to have all data for a given record in a single document.

MySQL	MongoDB
Table	Collection
Row	BSON document
Column	BSON field
JOIN	Embedded documents and Linking
GROUP BY	Aggregation
Primary key	Primary key

Table 1:

III. FEATURES OF MONGODB

MongoDB has a flexible data model. That means data can be stored in any structure. This feature also allows modification of data in an easy way. Another main feature is elastic scalability. All NoSQL databases contain some form of sharding or partitioning.This allows the database to scale out on hardware. Thereby allowing almost unlimited growth. MongoDB provides high performance than traditional relational databases. The performance of MongoDB is measured in terms of both throughput and latency at any scale. MongoDB does not use join operation, instead they use embedding of documents and linking. Because the data in MongoDB is more localized.This localization dramatically reduces the need to join separate tables. Each document structure in MongoDB database can vary from one another. If there is a need to create a new field in any one of the document, then the field can be created without affecting a central system catalog and without taking the system offline. In MongoDB, field updates can be done easily. It provides rich data model. Data locality and dynamic schema are other main features of MongoDB. The main feature of MongoDB includes querying, aggregation, indexing and auto sharding.

Indexes play a major role in providing efficient access to data,for both read and write operations,which are

supported natively by the database rather than maintained in application code. MongoDB supports many queries, mainly for highly scalable operational applications. The result of query execution can be a document or subset of specific fields within the document.

Different types of query provided by MongoDB include key value queries, range queries, geo spatial queries, search queries, text search queries, aggregation framework queries and map reduce queries. Replica sets are another feature of MongoDB which is a fail over mechanism. Only the primary database allows write operation. Multiple secondary servers are used for read operation. For a replica set, minimum three servers is required. Of the three servers, one is primary server, other is secondary server and the remaining one is arbiter server. Arbiter server is not used for storing data. They are used only during failover time to determine which server will be the next primary server. Another feature is auto sharding. This feature is used to overcome the hardware limitations. Hardware limitations means bottleneck in RAM/disk I/O. This feature of MongoDB helps to distribute data across physical partitions. These physical partitions are called as shards. Thus data is automatically balanced in the clusters as the data grows. In relational database sharding is not built into the database. An aggregate is a group of related entities and value object. Maximum document size in MongoDB is 16 MB and large documents are handled with Grid FS.MongoDB runs on OSs such as Windows, Linux, Mac and Solaris.

IV. COMPARISON OF MONGODB VS MYSQL

	MongoDB Commands
SELECT * FROM table	db.collection.find()
SELECT * FROM table WHERE user='Akshay'	db.collection.find({user="Akshay"})
SELECT * FROM table ORDER BY Age	Db.collection.find.
DISTINCT	.distinct()
GROUP	.group()

Table 2: Fig (a) Retrieval of data in MySQL and MongoDB

Modeling of data in MongoDB database differs from relational database. Different modeling styles can be applied depending on the requirement of the application. Most common modeling styles are embedding of documents and normalization on collections. The embedding feature has a disadvantage. That is, it may cause the situation that documents grow in size after creation which may degrade the performance of database.

Col 1

Col 2

Col 3

Table 3: Fig(b)Data modeling by embedding of documents

Example of Embedded documents having one to one relationship is shown below.

{_id:1,Name:"Akshay Anand",Address :{ City":"Kochi",Country:"India"}}

Example of Embedded documents having one to many relationship is shown below

{_id:1,Name: "Akshay Anand",Children :[{Name:"Aravind",Age:2},{Name:"Anupama",Age:4]}

This shows that array of values can be stored easily in MongoDB.

MongoDB supports denormalization.It is a process of reducing number of physical tables which are accessed more frequently to reduce the query processing time.This process reduces number of joins required to design the query to get desired output.

Col-1	Col-2

Table 4: Norm.

Col 1	Col 3

Table 5: Fig(c) Normalization.

Col 1	Col 2	Col 3

Table 6: Fig (d)D normalization.

In MySQL,the concept of normalization is used. This concept was first introduced by E.F.Codd. The objectives of normalization process include well organization of data, minimizing update anomalies and maximizing data accessibility. In ©,a common key is used to refer the tables Table1_Norm and Table2_Norm.In the next figure, the tables are merged together. Embedding is similar to denormalizationbut still little variation is there. Embedding of documents give better performance than normalization on collections.

V. ADVANTAGES OF MONGODB

It is schema less. MongoDB database belongs to document store category in which one collection holds different different documents. Number of fields, content and size of the document can be different in each document. The main advantage of MongoDB database is that structure of a single object is clear. It does not contain complex joins.It has deep query ability. Easy of scale out is another major advantage. In this type of database, conversion/mapping of application objects to database objects not needed. MongoDB uses the internal memory for storing the work set there by enabling faster access of data. It provides index on any attribute. The secondary indexes supported by the MongoDB database make them transparent to developers.

VI. APPLICATIONS OF MONGODB

They are widely used in big data and real time web applications such as Facebook, Yahoo, Google and Amazon. It is also used in content management and delivery. It can be used in mobile and social infrastructure. For user data

management the best choice among NoSQL database is MongoDB. It finds application in data hub also.It is the best choice for a small or medium sized non –critical sensor applications, especially when write performance is important.

VII. CONCLUSION

As NoSQL trend is relatively new, many researchers are attracted to this category of databases. NoSQL databases such as MongoDB and its key-value stores provide an efficient framework to aggregate large volumes of data. MongoDB can store complex data like array,object or reference into one field. Mapping of objects is very easy in this type of database. The features of MongoDB like auto sharding and replication of data make the development faster than MySQL. MongoDB provides flexibility, horizontal scalability, auto sharding and replication. MongoDB is a better choice for big data applications than MySQL database. It gives better performance than relational database. Depending on the requirements of application, we can choose the suitable NoSQL database.

REFERENCES

[1] Mrs.Anuradha Kanade, Dr.Arpita Gopal,Mr.Shanthanu Kanade"A Study of Normalization and Embedding in MongoDB"Advance Computing Conference (IACC), 2014 IEEE International
[2] Cornelia GYORODI,Robert GYORODI,George PECHERLE,Andrada OLAH "A comparative study:MongoDB vs MySQL"Engineering of Modern Electric Systems (EMES), 2015 13th International Conference on11-12 June 2015.
[3] K.Sanobar,M.Vanita,"SQL Support over MongoDB using Metadata"