www.ijemr.net ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962

Volume-6, Issue-1, January-February-2016

International Journal of Engineering and Management Research

Page Number: 21-24

Comparative Analysis of MongoDB Deployments in Diverse Application Areas

Sunu Mary Abraham
Assistant Professor, Department of Computer Science, Rajagiri College of Social Sciences, Kalamassery, Kochi, Kerala, INDIA

ABSTRACT

Several database technologies were developed to handle the present explosive growth of data. Many NoSQL databases evolved over time like Mongo DB, Cassandra, Hbase, Couch base etc for dealing huge unstructured data. This paper analyzes the deployment of MongoDB- a popular NoSQL database in different industrial application areas for the better understanding of its scope and to explore the reasons for employing MongoDB. This paper takes into account of the features like type of application, type of data handled, Scalability, Querying efficiency, data analytical tools used etc. for each application case study.

***Keywords*---** NoSQL, MongoDB, Hadoop, MongoDB deployment Comparison

I. INTRODUCTION

"Every two days now, we create as much information as we did from the dawn of civilization up until 2003", Google's Eric Schmidt words were not a myth, but a reality. Over the past few years, due to the explosive growth of data from terabytes to peta bytes to zeta bytes, new innovative database technologies evolved for scalability and availability. NoSQL databases were developed to deal with increased data volumes, the variety in data type and structures, performance and processing needs. Relational databases were not designed to cope with the scale and agility challenges that face modern applications, nor were they built to take the advantage of the cheap storage and processing power available today[1]. For enhancing the quality of services of the database, RDBMS is gradually getting replaced by NoSQL databases.

NoSQL means "Not Only SQL" and are poles apart from relational databases. Relational database systems define the schema first and insert data strictly according to the schema. Nowadays, most of the database applications would like to choose a data storage layer that would allow rapid iteration and evolution of schema over time. This enables prototyping over data modeling. NoSQL databases are built to allow the insertion of data without a predefined schema which makes it easy to make significant application changes in real-time and makes development faster. NoSQL databases are high performance, scalable systems and address issues that cannot be addressed by relational models like [1]:

- Large volumes of structured, semi-structured, and unstructured data
- Agile sprints, quick iteration, and frequent code pushes
- Object-oriented programming that is easy to use and flexible
- Efficient, scale-out architecture instead of expensive, monolithic architecture

This paper deals with MongoDB which is popular among the NoSQL databases and its deployments in various industries. MongoDB is an open source project held by the 10gen.company. It is a document-oriented, schema-free database, which stores data in BSON (Binary JSON) format [2]. MongoDB can deal with structured semi structured and unstructured data unlike RDBMS. MongoDB documents can vary in structure. Fields can vary from document to document. Similar documents are stored in collections. Here, *collection* corresponds to a *table* and *document* corresponds to a *record*. A collection is created on-the-fly when referenced for the first time. MongoDB can add, remove or change a field for a document without affecting other documents in the same collection. This saves the expensive ALTER table

operations that can lead to redesigning the entire set of schemas and the migration of existing database to the new schema.

MongoDB documents hold all data for a given record in a single document as against relational databases where data for a single record is spread across different tables. Therefore data in MongoDB is more localized, which reduces the need to JOIN separate tables [3]. Joins are avoided in MongoDB by embedding documents within the document. The result is increased performance and scalability as a single read to the database can retrieve the entire document.

Also, now-a-days, everyone works with objects (Python/Ruby/Java etc.) and we need Databases to persist these objects. Then, why not store these objects directly? MongoDB enables developers to design and evolve the schema through an iterative and agile approach [3].

MongoDB also provides horizontal scalability by a technique called Auto sharding and therefore chances of any node failure are almost nil. Most of the research studies reveal that MongoDB is much faster than MS SQL in writing (inserts/updates) and reading (retrieval)[2][4][5]. This paper is an attempt to recognize the kind of applications in which the NoSQL database -MongoDB can be deployed and also, explore the reasons for adopting or migrating to MongoDB technology.

II. DEPLOYMENTS- CASE STUDY

2.1 Expedia

Expedia is an online travel company. Expedia took an initiative to create new feature called scratchpad which made the travel search process fast, easy and personalized. Expedia employs MongoDB to eliminate the hand-written note-taking process for travel planning that helps them to finds deals, makes recommendations, and gives users better understand the changing marketplace of travel [7]. With MongoDB's flexible data model and simple horizontal scaling, the users were given a perfect vacation planning experience and its rich indexes powered analytics that made personalized suggestion to the users while planning and booking trips [6]. Here, Hadoop and Teradata served offline data analytics.

A team of only three developers built the first prototype for Scratchpad in less than two months, and launched it to production only two months later [6].

2.2 Metlife

MetLife is among the largest insurance companies in the world. They wanted to exploit their Big Data to create a 360- degree view of its customers so that they could know and interact with each client as an individual. Their effort to develop a centralized system using relational database was unsuccessful [8].

Exploiting MongoDB's flexible schema, it was easy to evolve schemas in real-time. It stored data in documents therefore managing customer data became easy. It handled the Big Data quite efficiently with auto sharding.

With MongoDB, they created a working prototype of a new system in two weeks and developed a new version of this new system, called "MetLife Wall" [6] in three months. It brings together data from more than 70 legacy systems. It runs across six application servers in two data centers and twelve database servers with a total storage capacity of 24 terabytes [9].

2.3 Shutterfly

Shutterfly is a popular, Internet-based, photo sharing and personal publishing company that manages a persistent store of more than 6 billion images with a transaction rate of up to 10,000 operations per second [11]. The company provides online photo service and products that enable consumers to share and cherish memories, in the form of photos and videos. Their products include photo books and personalized websites. Shutterfly faced an explosive data growth that exceeded the performance limits of its existing Oracle database. They slowly migrated to MongoDB which provided Shutterfly an agile, high performance, scalable solution at a low cost. Shutterfly's switch from Oracle to MongoDB resulted in the improvement of performance and substantial cost reduction. Also, Shutterfly developed the application in a very short span of time [10].

2.4 Google

MongoDB, together with Google created developer-friendly experience that enhances modern application development on Google Cloud Platform's Compute Engine. Together, they offer high performance, reliability and cost-effective platform for developing and deploying today's applications [12] [13].

2.5 Ebay

EBay has a number of projects running on MongoDB for search suggestions, metadata storage, cloud management and merchandizing categorization. For search suggestions, eBay made the suggestion list in a MongoDB document. This document was then indexed by word prefix, and metadata, such as product category. The multiple indexes provided them with flexibility in looking up suggestions and made the queries speedy. MongoDB helped the search suggestions to respond in less than 1.4 milliseconds [14] [15].

2.6 MTV

MTV Networks owns and operates several high-traffic web properties [6]. Adding new data types and tables to the schema significantly reduced read performance and other operations. The existing architecture RDBMS limited the evolution and success of their system. MongoDB's document storage model allowed to store hierarchical data (e.g.: TV episodes within a series) easily, without long expensive queries. MongoDB's flexible schema allowed MTV to concisely model the evolving data structures required by each brand. MTV found querying nested data in MongoDB easy to

understand and extremely fast [16]. The rich querying capabilities of MongoDB provided a combination of features unavailable in RDBMS.

2.7 Craiglist

Craigslist is a classified advertisements website with sections devoted to jobs, housing, and personals, for sale, items wanted, services, and community and discussion forums. Due to regulatory requirements, Craigslist has to retain digital records of its classifieds. With over a million new classifieds per day, significant amount of data has to be retained. It uses a MySQL data store to hold all the active information about classifieds, whereas MongoDB is used to store the archived data. As part of normal business changes, the data schema of stored data changes. By using MongoDB for archived data, Craiglist was able to effectively segment its data and alleviate problems due to schema migration [17].

2.8 Adhaar

India's Unique Identification project, Adhaar, is one of the world's biggest biometrics databases. Adhaar is in the process of capturing demographic and biometric data of over 1.2 billion residents. One of the database technologies at the back-end for the implementation of Adhaar Cards in India is MongoDB. Adhaar implementation enrolled thousands of Indians each day and adds terabytes of data to the Data Repository. Adhaar has used MongoDB as one of its databases to store this huge amount of structured and unstructured data. MongoDB was among several database products, apart from MySQL, Hbase, procured for running the database search. Here, MySQL is used for storing demographic data and MongoDB is used to store images [18]. Hadoop is the primary analytics and reporting platform that is used for the data processing and analytics.

III. COMPARATIVE ANALYSIS AND RESULTS

In this section, the information collected from the above case studies was analyzed. It was found that all the web and mobile applications prefer MongoDB for data storage and querying. This is because of MongoDB's rich querying capabilities, faster retrieval time and their ability to scale horizontally with the Big Data. Also, the schema for such applications is unstructured to some extent and has to get evolved over time.

In almost all the case studies, the transactional queries and sensitive data were dealt with the traditional SQL as they didn't want to compromise on the transactional data where ACID-Compliant data storage is a must.

Also, with MongoDB development and implementation time of applications with Big Data reduced to a time frame of weeks rather than months or years.

In all the applications with MongoDB as the database, mostly used data analytics platform used is Hadoop and Teradata. The reason is, both the analytical platforms use JSON format which is the stored data format for MongoDB.

Mostly, archival and content management applications tend to migrate from relational data models for faster and richer query experience. Organizations with Big data prefer MongoDB for their new apps, rather than existing apps migrating to MongoDB.

.

Table 1: Comparison of MongoDB Deployments

Organization	Application Description	Data Storage/ Search Queries	Transactional Queries	Data Analytics	New App/ Migrated from	Reasons for deploying MongoDB
Expedia	Planning Vacation	MongoDB	SQL	SAS, Hadoop, Teradata	New App	Storing unstructured data, Managing evolving schema, Data analytics
Metlife	360 degree customer view	MongoDB	SQL	Hadoop	New App	Scaling, Storing structured &unstructured data, Data analytics
Shutterfly	Web and Mobile photo services	MongoDB	SQL	Teradata	Oracle	Scaling, Storing structured &unstructured data, Data analytics
Google	Back end for Cloud launched applications	MongoDB	SQL	-	NA	High performance, reliability and cost-effective platform for developing and deploying applications

ebay	Search suggestion list, meta data storage	MongoDB	SQL	Hadoop	New App	Speedy Queries
MTV	Centralized Content Management	MongoDB	NA	-	Multiple RDBMS	Scaling, Storing structured & unstructured data, Managing evolving schema, Speedy Queries
Craiglist	Archival of Classifieds	MySQL, MongoDB	SQL	Hadoop	My SQL	Managing Evolving schema, Scaling, Storing structured &unstructured data, Speedy Queries
Adhaar	Storing and searching, demographic biometric data and images	SQL,MongoDB, Hbase	NA	Hadoop	New App	Scaling, Managing evolving schema, Data analytics, Storing unstructured data set which includes biometric data and images

IV. CONCLUSION

In conclusion, for unstructured big data related web or mobile application that requires horizontal scaling and which needs fast and rich querying capabilities, MongoDB is the mostly preferred NoSQL database. For applications with critical transactions and sensitive data handling, relational databases are still preferred. Also, with MongoDB, prototyping, development and implementation time drastically reduced when compared to the traditional SQL databases. Since organizations are starving for knowledge from the big data repositories, several big data analytical tools have emerged in the market and most of the user preferred analytical tools have support for integrating with MongoDB database.

REFERENCES

[1] http://www.mongodb.com/nosqlexplained
[2] Chieh Ming Wu1, Yin Fu Huang, John Lee: 'Comparisons Between MongoDB and MS-SQL Databases on the TWC Website', American Journal of Software Engineering and Applications 2015; 4(2): 35-41.
[3] https://www.mongodb.com/mongodb-architecture
[4] Rajat Aghi, Sumeet Mehta, Rahul Chauhan, Siddhant Chaudhary ,Navdeep Bohra:' A Comprehensive comparison of SQL and MongoDB databases', International Journal of Scientific and Research Publications, Volume 5, Issue 2, February 2015
[5]Cristina Bazar, Cosmin Sebastian:'The transition from RDBMS to NoSQL. A Comparative Analysis of Three Popular Non-Relational solution: Cassandra, MongoDB and Couchbase', Database Journal Vol V,no,2/2014
[6]https://www.mongodb.com/expedia
[7]http://blog.expedia.com/how-expedia-is-simplifying-travel-planning/
[8]]https://www.mongodb.com/metlife
[9]http://www.metlifegto.com/news/Built-in-record-time--the-MetL
[10] https://www.mongodb.com/shutterfly
[11]https://gigaom.com/2011/01/28/real-world-nosql-mongodb-at-shutterfly/
[12] https://www.mongodb.com/google
[13]https://cloud.google.com/solutions/mongodb-on-compute-engine?hl=en
[14] https://www.mongodb.com/ebay
[15] https://dzone.com/articles/ebay-secret-database-scaling
[16] https://www.mongodb.com/mtv
[17]]https://www.mongodb.com/craiglist
[18] http://usf.vc/wp-content/uploads/2014/04/USF-Aadhaar-Report-Jan-2015.pdf