MongoDB - NoSQL Database

IDS 521 - Term Paper

By
Vyshali Madduri - 665525782
Somya Maheshwari - 652842577
Pravaagini Mani - 661805879

Abstract:

The requirements for data management for web-based applications have greatly altered in recent years. In-depth data consistency and a broad range of capabilities are features of relational databases. Due to the exorbitant cost of storing and managing data in conventional relational database systems, noSQL databases have been developed. NoSQL databases provide more scalability and heterogeneity when compared to RDBMS. NoSQL database MongoDB provides high levels of scalability, performance, and availability. The popular NoSQL database MongoDB is examined in this essay. The overview, characteristics, contrasts, benefits and drawbacks, and uses of MongoDB.

1. Introduction

The phrase "NoSQL database" is frequently used to describe any non-relational database. NoSQL is sometimes referred to as "non SQL," but it is also sometimes referred to as "not merely SQL." In any case, the majority of people concur that NoSQL databases are those that store data in a different format than relational tables. The advent of NoSQL databases coincided with a sharp decline in storage costs in the late 2000s. The days of having to build a complicated, challenging data model in order to prevent data duplication are long gone. NoSQL databases were designed with developer productivity in mind because developers (and not storage) were starting to represent the majority of the expense of developing software.

A strong, adaptable, and scalable general-purpose database is MongoDB. Secondary indexes, range queries, sorting, aggregations, and geospatial indexes along with the capacity to scale out features are incorporated by MongoDB. This essay discusses the key design choices that gave MongoDB its current form[1]. For storing large amounts of data, MongoDB is a document-oriented NoSQL database. MongoDB uses collections and documents rather than the tables and rows seen in conventional relational databases. The fundamental unit of data in MongoDB is a pair of key-value pairs, which make up documents. Collections are the equivalent of relational database tables since they include sets of documents and functions. A database called MongoDB first became popular in the middle of the 2000s.

In the following sections we will see the overview features, comparisons, advantages and disadvantages, applications of MongoDB.

2. MongoDB

2.1. Overview

- The majority of NoSQL databases share the extremely adaptable MongoDB database architecture. It enables you to work with and save several data kinds in a single document. MongoDB is used to store a lot more data than conventional databases. These fundamental traits of MongoDB are brought about by the JSON document storage format that the database system uses.
- The open standard file and data exchange format known as JSON (JavaScript Object Notation) uses readable language to store and transfer data objects. It is a common data format that is used in many electronic data exchanges. Despite being derived from JavaScript at first, it is now language neutral because many popular programming languages contain codes that permit data to be saved in the JSON format. extension json

2.2 Data Storage in MongoDB

Collections are used by MongoDB to store large amounts of data. Documents that resemble rows in a table in a relational database system can be found among the collections. Keys, the most fundamental data type in MongoDB, are the building blocks of documents (which can contain many, distinct fields and data types). By adding and deleting new and existing fields, one can change the organization of data that has already been recorded in a document using MongoDB, a NoSQL database. The same document can also use different data kinds. Therefore MongoDB has an adaptable structure. MongoDb operates in the application layer and the data layer[2].

2.3 Features

- **Document Oriented:**
  MongoDB saves all data as documents unlike RDBMS, where each document has a unique ID. The data in these documents is kept in key-value pairs rather than rows and columns, making the data far more flexible.

- **Schema-less Database:**
  Schema-less database is an useful feature offered by MongoDB, which allows for the storage of several document types in a single collection. A single collection holds numerous documents in MongoDB, each of which may have a varied amount of fields, kind of content, and size. A document need not be comparable to another document, as is the case with relational databases. MongoDB offers databases a lot of flexibility.

- **Scalability:**
  Through a process known as sharding, MongoDB offers horizontal scalability. A vast quantity of data is divided up into manageable chunks using the shard key, and these chunks are then evenly dispersed across shards that are spread over numerous physical servers. This process is known as sharding. It can also include new devices into an active database.

- **Indexing**:
  Every field in the documents is indexed by the MongoDB database with primary and secondary indices, which speeds up and simplifies retrieving or searching for data from the pool of data. If the data is not indexed, the database must manually search each document using the given query, which is time-consuming and ineffective.

- **Aggregation**:
  Users of MongoDB can additionally conduct operations on the aggregated data to obtain a single result or a computed result. A map-reduce function, a single-purpose aggregation method, and an aggregation pipeline are the three types of aggregations that are offered.

2.4 Use Cases of MongoDB

MongoDB can process large amounts of data with a high performance. It is a schemaless document type database. It lacks a rigid structure and can easily accept any data type. Due to these reasons, MongoDB is perfect for big data and real time analytics. Hence, it has found its applications in various prominent organizations [3].

- **MetLife:**
  A leading insurance and financial services company released its customer service application, The Wall. This application holds and displays information about their customers, their policy information and payment statements. At the backend, it uses MongoDB to store and process large amounts of database information.

- **Indian Aadhar:**
  The Aadhar project is India's Unique Identification system that holds massive amounts of demographic and biometric data of the Indian population. MongoDB was the original database system used in this project. It allowed Aadhar to store, process and analyze large unstructured data.

- **eBay:**
  eBay is an ecommerce platform that facilitates B2C and C2C sales. It ranks among the most popular shopping apps in the United States and hence has petabytes of data that needs to be processed efficiently and without latency. MongoDB provided eBay the database it needed in order to scale its platform and execute tasks like big data analytics, search recommendations etc.

- **EA:**
  EA Sports FIFA uses MongoDB to manage the massive amounts of data generated and to improve their gaming experience. Due to MongoDB's flexible schema, new items can easily be added to the data layer.

2.5 MongoDB Compatibility and Integrations

[4] MongoDB offers compatibility with all the popular operating systems, programming languages, frameworks, Graphical User Interfaces, Integrated Development Environments. MongoDB Using the necessary APIs provided by MongoDB drivers, one can easily connect MongoDB databases directly to the application code. For example, MongoDB supports Python using pymongo driver, it supports Java using the MongoDB Java Driver. It is also compatible with popular development frameworks such as Angular, Flask and Django. [5] Angular is a development framework used to build desktop web and mobile applications. It is one of the most popular JavaScript frameworks in the market. Angular can easily be integrated with MongoDB using either MEAN stack (where the client is built on Angular and the database is MongoDB) or Serverless architecture. Several Runtime Environments are also supported by MongoDB namely Java Runtime Environment (JRE), Node.js . All major operating Systems like Windows, Linux and macOS are well supported by MongoDB. Integrated Development Environments (IDEs) like JetBrains and VS Code can be connected with MongoDB databases with the help of extensions. This way, you can easily create, manage and query MongoDB databases in these IDEs. In addition to this, MongoDB also provides its own GUI, MongoDB Compass where you can query, optimize and analyze your data.

3. Cassandra vs MongoDB vs HBase

MongoDB is undoubtedly one of the most well-known open-source NoSQL databases, but Cassandra and other wide column databases might offer better query performance and always-on features. The availability of managed DBaaS services, which allow you to delegate database management and maintenance to the provider and concentrate solely on developing your application, is another crucial consideration when selecting a NoSQL database. In this regard, HBase falls short of MongoDB's DBaaS offerings like MongoDB Atlas, which are much more developed. HBase, on the other hand, can be a very effective solution for write-intensive applications and massive records[6].

Name	Cassandra	MongoDB	HBase
Architecture	Wide Column Store	Document Store	Wide Column Store
Server OS	FreeBSD, Linux, OS X, Windows	Linux, OS X, Solaris, Windows	Linux, Unix, Windows
Distributed System Consistency	Eventual and Immediate Consistency	Eventual and Immediate Consistency	Immediate Consistency
Owner and Developer	Apache Software Foundation	MongoDB, Inc.	Apache Software Foundation
Replication	Masterless Ring	Master-Slave Replication	Master-Slave Replication
Programming Language (Base Code)	Java	C++	Java

Fig 1. Difference between different NoSQL database systems [6]

4. Advantages and disadvantages of MongoDB

4.1 Advantages

- **Flexible Database**
  The schema of MongoDB is not fixed. It has a dynamic and change-friendly architecture that supports storage of both structured and unstructured data without the need of a predefined schema. Since both businesses and the data they maintain are always changing, a flexible database model that could adjust to these changes is crucial.

- **High Performance**
  Most of the data in MongoDB is stored in the RAM. And so when queries are made, the data is directly gathered from RAM rather than the hard drive. Due to this reason, it enables a faster performance.

- **Higher Availability**
  MongoDB contains features like gridFS and replication. During system crashes or software failures, the replication process can help restore and backup important information.

- **Sharding**
  Sharding is used by MongoDB when working with huge datasets. Data from a huge collection is distributed across multiple machines. This process is called sharding. In the event that the bulk of the data presents a problem for the server, it automatically breaks the data into more manageable chunks without interrupting the action.

- **Easy to use**
  MongoDB is user-friendly and easy to use. It is easy to set-up and install.
  It offers a variety of useful features including GUIs. It also offers integration with popular frameworks, languages etc. Due to this reason, it is also easily accessible[7].

4.2 Disadvantages of MongoDB

- **Joins**
  MongoDB also does not inherently support joins. Since It is not a relational database, it cannot handle joins. It can be implemented by manually inserting the code but that process is messy and extremely time consuming [8].

- **Indexing**
  Another major limitation of MongoDB is that it can only perform quickly and efficiently if the right indexes are used. MongoDB will operate at a very slow speed in case the indexing is carried out poorly or contains any errors. Moreover, fixing such an issue is again a tedious process.

- **Duplicates**
  Since the relations are not clearly defined in MongoDB, there is always a chance of data redundancy. Thus, data is not managed efficiently. Such an issue is not ACID compliant and may eventually result in corruption.

- **High Memory Usage**
  Because of the lack of join functionality, there is data redundancy and hence MongoDB requires high storage.

5. Bug reports and criticisms

- ### Security

  Data from tens of thousands of MongoDB installations have been stolen as a result of the default security setting of MongoDB, which grants anyone complete access to the database. In addition, a lot of MongoDB servers have been held hostage.

  B. Jose and S. Abraham mentioned “In September 2017; updated January 2018, in an official response Davi Ottenheimer, lead Product Security at MongoDB, proclaimed that measures have been taken by MongoDB to defend against these risks." in 2017 [9]

  The official MongoDB RPM and DEB packages' binaries by default bind to localhost as of the MongoDB 2.6 release. This default behavior was made available to all MongoDB packages on all platforms starting with MongoDB 3.6. As a result, unless specifically set by an administrator, all networked connections to the database will be declined.

6. Conclusion and Future Scope

With the release of MongoDB, people recognized the importance of NoSQL databases and their simplicity and flexibility. Because of their benefits and added advantages over SQL databases, MongoDB and NoSQL databases have shown rapid growth. MongoDB is now being used by big companies like Forbes, eBay, Marcello and Adobe.

[10] However, despite being a popular choice, MongoDB has certain security issues that one cannot ignore. MongoDB Rails is vulnerable to hash-injection attacks which can lead to authentication-bypass, denial of service, or timing attacks. Although MongoDB is a NoSQL database, it is also vulnerable to SQL injection attacks. This occurs when an attacker sends a malicious request to the database via SQL queries and the database returns the information requested by the attacker. This creates a vulnerability that has the potential to destroy your system from within.

As NoSQL databases gained popularity, other companies also launched competitors in document databases. For example, DocumentDB was released by Amazon Web Services (AWS) with MongoDB compatibility, but is its competitor in reality. Other popular alternatives to MongoDB in the market are Couchbase, MariaDB, Cloudera etc.

7. References and Citations

[1]Kristina Chodorow, Michael Dirolf, 2010, MongoDB:The Definitive Guide

[2]“What Is MongoDB? Working, Architecture, Features, and Use Cases |.” What Is MongoDB? Features and Use Cases, 5 Aug. 2022, www.spiceworks.com/tech/cloud/articles/what-is-mongodb.

[3] *MongoDB Real World Use Cases: Advantages & Top Companies [2023]*. upGrad blog. (2022, November 23). Retrieved December 2, 2022, from https://www.upgrad.com/blog/mongodb-real-world-use-cases/

[4] *MongoDB Compatibility & Integration*. MongoDB. (n.d.). Retrieved December 2, 2022, from https://www.mongodb.com/compatibility

[5] *Angular & MongoDB: Support & Compatibility*. MongoDB. (n.d.). Retrieved December 2, 2022, from https://www.mongodb.com/compatibility/angular

[6]“Cassandra Vs. MongoDB Vs. Hbase: A Comparison of NoSQL Databases | Logz.io.” Logz.io, 17 June 2019, logz.io/blog/nosql-database-comparison.

[7]“Top Advantages of MongoDB.” Top Advantages of MongoDB, www.knowledgehut.com/blog/web-development/advantages-of-mongodb. Accessed 2 Dec. 2022.

[8]Debra Bruce, View all posts by. “Ultimate Guide MongoDB: Definition, Advantages and Disadvantages.”Knowledgenile,21Apr.2021, www.knowledgenile.com/blogs/pros-and-cons-of-mongodb.

[9] B. Jose and S. Abraham, "Exploring the merits of nosql: A study based on mongodb," 2017 International Conference on Networks & Advances in Computational Technologies (NetACT), 2017, pp. 266-271, doi: 10.1109/NETACT.2017.8076778.

[10] Singh, S. (2019). Security analysis of mongodb. *International Journal for Digital Society*, *10*(4), 1556–1561. https://doi.org/10.20533/ijds.2040.2570.2019.0193