

# Application of MongoDB Technology in NoSQL Database in Video Intelligent Big Data Analysis

Qing Tan<sup>1, a</sup>

<sup>1</sup>College of Information Technology, Luoyang Normal University, Henan Luoyang, 471934, China

<sup>a</sup>edutanqing@163.com

**Keywords:** MongoDB; NoSQL; Big Data; File Storage; Video

**Abstract.** MongoDB is a database based on distributed file storage, which lies between relational database and non-relational database. MongoDB is a very excellent database for document storage. It mainly solves the problem of storage and access efficiency of massive data. The MongoDB technology in NoSQL database in video intelligent big data analysis is flexible. Implementation of computing, storage resources cloud deployment, equipment ad hoc network, self-management, business never interrupted, flexible expansion on demand. Support distributed search engine, through distributed database MongoDB, support massive data storage and analysis.

## Introduction

NoSQL real-time historical data cloud can provide users with a series of data services, such as real-time data transmission, real-time compression, real-time storage, real-time computing, real-time synchronization, and real-time supply and so on. To help users build a "big data cloud" in a wide area, users can obtain data, calculate data, feedback display, and achieve high efficiency and high degree of freedom data statistics and mining calculation from anywhere in the network. Help users to achieve production decision, production optimization and production safety.

The real rapid development of NoSQL began in 2007 when Google and Amazon engineers published papers on BigTable and Dynamo databases, describing the design ideas of the new databases they have been using. From 2007 to now, more than ten popular NoSQL products BigTable have proposed a column storage model, which proves that data persistent storage can be extended to thousands of nodes.

The biggest defect of relational databases such as MySQL is extensibility. Although every database manufacturer has the solution of cluster, the scalability (IO bottleneck caused by storage capacity and the bottleneck caused by concurrent read and write access) is very limited. At present, there are two ways to solve the problem of database extensibility: the first is data partitioning (sharding) or functional partition, although it can solve the problem of database extensibility.

In the era of Web 2.0, the provider of website information has changed from a traditional webmaster to an ordinary user, and the information provided by the user is massive. SNS sites such as Facebook QQ space may generate tens of millions of levels of data every day [1]. It takes a lot of time to make SQL queries in a table containing billions of records in an RDBMS. Although the query problem can be partially solved by dividing the data into the database and table, it also brings some problems such as increasing the complexity of program development and data backup, and the complexity of database expansion and so on.

With the increase of data volume, even the analysis and retrieval of video content based on TB level video data, it may take several hours to calculate in the traditional technology mode, which is far from being able to meet the requirement of timeliness. Users hope to effectively manage and use massive data, quickly identify effective data, and improve the value of data utilization.

NoSQL system directly provides OPC client interface program to adapt to OPC protocol widely used in digital oil field. Users can easily import data into the database. For non-standard protocol, SQLNot provides all kinds of APIs, such as C#, Java, and provides interface programming template to reduce the user's development workload as far as possible.

This paper will analyze and study this new database, introduce its concept and theoretical basis, analyze the characteristics and applications of different types of NoSQL database, and focus on the application of MongoDB, the representative of NoSQL database in big data.

## Technology and Application of NoSQL Database

NoSQL is generally regarded as "not only SQL". Today, NoSQL refers to such a class of databases and data storage, which do not follow the principles of classical RDBMSs and are often related to big data sets on a web scale. NoSQL is not just a product or a technology; it represents a family of products, as well as a series of different, sometimes interrelated, concepts about data storage and processing [2]. Its significance lies in that the relational database is used when it is applicable to the relational database, and it is not necessary to use the relational database when it is not applicable, so we can consider using more suitable data storage.

MongoDB is a product between relational database and non-relational database. The data structure he supports is very loose, in bson format similar to json, so it can store more complex data types.

It's basically a distributed storage based on Key-value (i.e., at the expense of the "relational" database's greatest advantage-join). Non-relational databases are suitable for storage and real-time query services that require "structured data and semi-structured data", and are suitable for applications with higher data size and concurrent access requirements.

Wide columnar store, which includes BigTable Cassandra and Hbase, is a type used to process structured data [3]. It has several features: it has the ability to scale large scale, has a concept similar to column in a database, a very flexible schema, and uses the storage mechanism of memtable/sstable. And based on column storage, Cassandra adopts the idea of dynamic ultimate consistency, and draws lessons from BigTable's data model and implementation, so many people regard it as an open source version of BigTable Dynamo, as is shown by equation(1), and H is this type of KV store is our focus.

$$H_{1,L}(f) = 2 \sin^L(\pi f) \sum_{l=0}^{L/2-1} \binom{L/2-1+l}{l} \cos^{2l}(\pi f) \quad (1)$$

Mongoddb is a very excellent database for document storage. It mainly solves the problem of storage and access efficiency of massive data. The data is organized into a document in a bson format similar to json and stored in a collection. According to official tests, when the amount of data is over 50GB, Mongoddb accesses more than 10 times as fast as MySQL. Supported concurrency can reach 0.5 ~ 15000 times per second. Mongoddb comes with an excellent distributed file system gridfsto support massive data storage..

Analyze big data to produce small data. It is not enough for the camera to work 7 x 24 hours a day and record what is happening within the lens coverage, because most of the information may be invalid for the customer, and the valid information may be distributed within a relatively short period of time. According to the mathematical statistics theory, information presents power law distribution, or called information density, the higher the density is, the greater the customer value is. Real-time influx of massive data easily produces a large number of dormant data, waste a large number of storage resources [4]. The large amount of data is analyzed intelligently, and the value data fragment is extracted. Establish summary information; reduce the amount of data that users need to deal with, form metadata database.

MongoDB is a high performance open source document database and one of the most concerned NoSQL technologies. It is famous for its agility extensibility and enterprise-friendly (transaction support consistency and data integrity assurance large enterprise application cases). Some people even think that M in lamp should replace MySQL with MongoDB.

The collection of NoSQL also includes the data transmission API with built-in cache. it provides the service quality of "poor link after sending". At the same time, it has a built-in N2N database synchronization network, which has a perfect transmission design, and all transmissions are automatically adapted to the disconnected network [5]. In the previous military test, N2N has passed

more than 5000 km transmission test, and it supports DTN-delay tolerance network, so the data transmission can not be lost, so it can meet the needs of users.

There are usually two levels of consistency in NoSQL: the first is strong consistency, where all machine states in the cluster are synchronized, and the second is final consistency, which allows transient data inconsistencies, but ultimately data consistency. According to CAP theory and it is strong consistency and availability, partition tolerance at the same time. The ultimate consistency is a compromise to consider the user experience and is the biggest difference from the traditional RDBMS.

### **Application of MongoDB Technology in NoSQL Database in Video Intelligent Big Data Analysis**

The intention of NoSQL movement is not to replace the traditional relational database, but to combine it with the relational database according to the characteristics of the application. In the era of relational databases as mainstream database solutions, architects tend to focus only on the features and related maintenance of the database products used. But in the age of the Big Bang Theory, there are applications that generate huge amounts of data. Different applications have different characteristics and requirements for data access. Architects need to use different technologies to meet the needs of different applications according to the characteristics of the application.

Finally, because Mongo can support complex data structures and has powerful data query capabilities, it is so popular that many projects are considering using MongoDB instead of MySQL to implement Web applications that are not particularly complex. For example, why we migrated from MySQL to MongoDB is a real case of migrating from MySQL to MongoDB.

The first problem is performance. All databases are based on storage optimization rather than memory optimization. That is to say the best scenario for database application is that there is a small amount of memory and a large number of external IO. Even if you have a large enough cache to cache all your data into memory, it's still several orders of magnitude slower than a specially designed in-memory database or Key-Value cache. This is decided by the algorithm within the database, so don't expect to use the database as cache, as is shown by equation(2), where G is except for special memory databases, such as Oracle times ten [6].

$$G_N = \sum_{k=1}^8 |R(k+1) - R(k)| = 2, R(9) = R(1) \quad (2)$$

The NoSQL database is still in its infancy, and there will be great changes in the future. For the choice of NoSQL databases, the best strategy is not to bet on a solution, but to find solutions that have better portability.

As NoSQL is not a standardized mature product, there may be an unknown risk in the process of use. In order to avoid risk, data backup is the key. Data backup has two solutions. One is to backup the data in the NoSQL database and itself to the relational database. The two is to use two NoSQL databases at the same time to reduce the rate of data loss.

Big data is processing ability. Based on big data parallel or distributed processing techniques, such as Hadoop Hbase, a distributed storage and data management framework for big data is formed, and real-time and batch processing ETLs are provided to realize data cleaning / conversion / loading to form value information aggregation.

Visual presentation is ability. Efficient retrieval, comparison and presentation of video content, integration and display of views with GIS platform, information analysis platform, emergency processing platform and so on, to enhance the practicability of video data. Through big data technology, further mining the value information behind the massive video surveillance data, and quickly feedback the connotation knowledge to assist the decision judgment will be the development direction of the video surveillance application in the future.

BigTable, a non-relational database, is a sparse, distributed, persistent multi-dimensional sorting Map. BigTable designed to reliably handle data at the PB level and be deployed to thousands of machines.

In addition, there are a lot of redundant and useless data in the massive data, so the data must be cleaned, filtered and compressed to reduce the storage capacity and reduce the use cost of oil field users. NoSQLt real-time historical data cloud adapts to this demand. The arrival of real-time data will filter and clean the useless data according to user presupposition. In addition, the NoSQLt database has an efficient "waveform lossless compression algorithm", which can compress and store the data efficiently, and the compression ratio is as high as 1: tens of thousands.

In the field of traditional RDBMS, we must analyze data and build data model before storing data. To build a data model is to build a data table to determine the fields in the table, the data types of the fields and the relationships between the fields of the tables. However, requirements change over time in a real enterprise, and while a certain degree of reconfiguration is supported in traditional relational databases, if the application changes too much, the common idea of refactoring. NoSQL databases are to break the limits of this data model.

### **System Experiments and Analysis**

Open architecture, intelligent analysis plug-in, extensible ISV customization development simple and fast. Open architecture, third party intelligent analysis plug-in seamless integration; provide intelligent algorithms, intelligent analysis application services rest standard interface, support for ISV customization development [7]. Abundant intelligent analysis ability, support visual operation management and graphical analysis. It realizes automatic alarm of suspicious behavior, real-time distribution and control of whole network, accurate identification of person / vehicle, efficient retrieval of video content, intelligent case analysis, and self-diagnosis of equipment fault.

NoSQL carries out the principle of "two lines of revenue and expenditure", that is, the number of stored data is independent of the actual supply of data. Even after high efficiency compression, the interpolation algorithm built by NoSQLt can still restore the data with each millisecond precision. This interpolation algorithm is very efficient, as is shown by equation (3), where R is the user can not even feel the time difference between the real storage and the data.

$$R(W_j^\sigma, F_e) = \frac{E[(\xi_{W_j^\sigma} - E(\xi_{W_j^\sigma}))(\xi_{F_e} - E(\xi_{F_e}))^T]}{\sqrt{D(\xi_{W_j^\sigma})D(\xi_{F_e})}} \quad (3)$$

In a document-oriented database, we call a table in a relational database a collection and a record (tuple) in a relational database as a document. A number of keys and their corresponding values are stored together in an orderly manner to form a document, a plurality of documents (documents) constitutes a collection, and a plurality of collections constitutes a database. An instance of MongoDB can host multiple databases, each of which is completely independent.

The application scenario of NoSQL is very limited. A certain type of NoSQL is designed to perform the search function of inbox in facebook only for a specific type of application scenario, whereas relational databases are much more common, which means that NoSQL is very difficult to use. First of all, you have to figure out your application scenario.

DPU (data processing Unit): providing data access adaptation, data aggregation, unstructured data archiving, data conversion engine, data cleaning and conversion, In addition, according to the rule engine, the real-time alarm information is reported to DMU (data Mining Unit): rule model management, using visual interface to define business rules (Schema), in order to speed up the real-time data processing, Introduce efficient memory database technology. The data processed by the previous data processing layer is stored in the form of relational data.

MongoDB does not support connection queries and transactions; when creating and updating data with MongoDB, data is not written to the hard disk in real time, so it is possible to lose data.

## Summary

MongoDB's most important feature is that the query language it supports is very powerful, and its syntax is similar to that of object-oriented query language. It can almost realize most of the functions similar to single table query in relational database, and it also supports the indexing of data. After layer by layer optimization, a set of SQL real-time history database can store 200000 measured points for several decades, while the network deployed NoSQL real-time history data cloud has no upper limit, and users can deploy freely to realize massive data storage.

## References

- [1] Cai Jinhua. Analysis of NoSQL and its use. *Computer knowledge and Technology*, 2013, (7): 2757-2758.
- [2] Shashank Tiwari. *Giant. Deep into NoSQL*. Beijing: people's Post and Telecommunications Press, 2012.3-6.
- [3] Zhang Huaqiang. Relational database and NoSQL database. *Computer knowledge and Technology*, 2015 ,(7): 4802-4804.
- [4] Cheng Hua, Lin Ning. Using MongoDB to Implement Textbook Management System instead of MySQL. *International Conference on Communication Software and Networks, VOL02*, 2014:180-192.
- [5] Huang Xianli. Development and Application of NoSQL Non-Relational Database. *Fujian computer*, 2012, (7): 30-45..
- [6] Srivastava J, Cooley R, Deshpande M, et al. Web usage mining:discovery and application of usage patterns from web data, *SIGKDD Explorations*, 2011,1(2):12-23.
- [7] Chen Minmin. *Research on Forum Information extraction and Storage based on MongoDB Cloud Storage platform*. Shanghai: Shanghai Jiaotong University, 2013.