The application of NoSQL database in Air Quality Monitoring

Mei Han[1, a]

[1] Jining University Jining Shandong China

[a] hanmeisd@126.com

Keywords: Air Quality Monitoring, Big Data, NoSQL, MongoDB

Abstract. In air quality monitoring system, various monitoring devices generated tremendous amount of date which make traditional relational database very inefficient in reading and storing it. In order to solve this issue, this paper compares the characteristics of various NoSQL databases and selected MongoDB to process the data. In addition to that, methods and feasibility of using NoSQL database in the air monitoring system are analyzed.

1 Introduction

Currently, Jining city use scientific methods to monitor the pollution of industry. So far, Jining already built 327 online monitoring systems in 80 high pollution companies which can monitor the operation of drainage facility and the result of desulfuration. In the first half of year 2014, Jining developed the current air quality monitoring system which can monitor the air quality of the whole city automatically. This development put more pressure on the database, making the original relational database couldn't satisfy the requirement.

2 The requirement for air condition monitoring system

The initial requirement for air quality monitoring station is to be able to monitor the content of SO_2 NO_2 PM10 PM2.5 O_3 and CO in the air. In addition to that, the air quality monitoring system should be able to monitor wind speed, wind direction, temperature, humidity and air pressure. Based on the information, automatic monitoring station generates data in every minute which can be used as a suggestion for people's traveling plan. Big data and high concurrent are the basic characteristics of air quality monitoring station. However, the consistent specification for data is not that tough.

2.1 Highly concurrent requirements. Because of the development of environment monitoring system, lots of monitoring devices are included in the system, monitoring the data becomes more complicated and the external visit becomes more often. The database workloads become very high, which could possibly reach 10000 read and write requests per second. Relational database could handle at most 10000 SQL search request, but to handle 10000 SQL write data request, the IO of hard disk even couldn't handle.

2.2 Huge storage requirements. After the implementation of automatic monitoring system, there will be tremendous amount of information which could possibly reach millions of record every day. For relational database, doing SQL search in a form which contains more than 100 million records is super inefficient. It even couldn't handle use log in system for large website.

2.3 High scalability and high availability requirements. The relational database is the hardest to extend in the web-based platform. In order to do join handling, relational database has to store the data in one server which is not good to spread the data across more servers. In the past, our database couldn't work as the web server or app server to add more hardware and server to enhance the performance and the workload capacity. For continuously automatic monitoring system, updating the relational database system is very painful, usually need to migrate data. In addition, hardware updating cost is very high.

Because of the difficulty in handling big data and highly concurrent requirements, traditional relational database couldn't be used in current systems. So far, the NoSQL database is becoming mature and it is capable to work as the database for Jining's air quality monitoring system.

3 The analysis of NOSQL database

Because of the development of web 2.0, various web activities created the need for highly concurrent and big data. NoSQL database was developed as a solution to this issue. The NoSQL started in 2009 and developed to Not Only SQL later. Start from 2010, NoSQL database becomes quite popular in Internet and Database. The characteristic of NoSQL database is not using relational modeling, which is good for our project. Non-relational database can be classified into 3 categories.

3.1 Key-Value database. In this database, data are organized according to the key-value. It is similar to Hash map in traditional programming model. By recoding the key-value, data can be added searched or deleted quickly. The processing efficiency is very high for a reasonable set of Key and value. NoSQL attracted the attention of many developers since it was born. There are many projects based on NoSQL, for example, the Google's Bigtable, Amazon's Dynamo, Riak Redis etc. The first two find its commercial use in its enterprise and the design methodologies have set a model for the future developers.

Key-Value database is relatively simple and fast. It is also good for storing user's information, such as conversation, configuration file, parameter and shopping cart. The kind of database like Riak and Redis can provide an efficient way handling real time data. The Large content often stored as a value in key value database. This storing method made it can't query by key value effectively, making it not suitable for dynamic conditional search. The good thing is NoSQL database is strong in that part.

3.2 Document-Oriented database. Document-oriented database is a kind of NoSQL database. It saves data in the format of document. Each document contains the unit data. Every data has a name and value corresponds with it. The value can have the format of string, character, number, date or complex type data, such as the ordered list and related data. The smallest unit for data storage is document. In one table the type of document are not necessarily the same, data can be stored using XML, JSON, JSONB format. In document database, by searching according to aggregating data, it can be found instead of the whole content. Also indexes can be created according to aggregating data, which can efficiently support dynamic searching requirement.

Document NoSQL includes MongoDB CouchDB RavenDB etc. Among them, MongDB is the most popular one. As a high performance database, the document type NoSQL database is easy to deploy, so easy to master. It regards records as a document object, the storing operation is simple. It also support some friendly search function of SQL. High performance and convenient search making it as the first choice among NoSQL database for customers. The customers include SAP, Desiney SourceForge .etc. MongDB scarified some consistence in data consistency in order to achieve high performance. By default, it can support high speed writing and updating without error noticing. Most of the drivers are written asynchronously which need the built in function to achieve the error message manually. In addition, when cluster scale and workload reached a limit, the stability of system will be affected.

Because of the document oriented storage mechanism of NoSQL database, it is very good for the search and store of log data. The search efficiency reached a higher level, Building and using becomes comparatively easy. It got supported from various NoSQL user since it was born. But as the cluster scale and workloads increases, the robustness and maintainability got challenged a lot for document based database. Its nonsupport transaction also harms the usability.

3.3 Column-Oriented database. A column-oriented database is a non-relational database that stores data tables as sections of columns of data rather than as rows of data. In comparison, most relational DBMS stores data in rows. In practice, row-oriented storage layouts are well-suited for OLTP-like workloads which are more heavily loaded with interactive transactions. Column-oriented storage layouts are well-suited for OLAP-like workloads (e.g., data warehouses) which typically involve a smaller number of highly complex queries over all data.

In the practical project, HBase (Hadoop database) has been widely used as a database under Hadoop platform. HBase is a high reliable, high performance, column-oriented distribute storage system. Through HBase, the large storage cluster can be built in low coast PC-Server. HBase could

satisfy the need for big data storage. However, because of the complicated structure, the learning cost is relatively high in order to satisfy certain need.

4 The storage scheme of environmental monitoring data in NOSQL

Based on the above analysis, document oriented database has the most advantages for the environment monitoring system. In this paper, MongoDB is used as an example to introduce the design of NoSQL database. MongoDB is developed through C++ and developed based on Google BigTable distributed file storing system. It mainly provides the solution for expansive high performance database. According to the testing result of MongoDB website, when the data is larger than 50 GB, the visiting speed of MongoDB is 10 times faster than the speed of MySQL, which could support 5000-15000 parallel operation per second. The greatest advantage of MongoDB is the searching language it supported is very powerful. The grammar of MongoDB is similar to object oriented searching language, which could achieve most search functions in relational database. In addition to that, MongoDB could support building indexes of data. MongoDB is an open source high performance database, it is also the most popular NoSQL techniques which is famous for the fast, extensibility and friendly to enterprise. It combines the advantages of SQL and NoSQL database.

4.1 The data storage in air quality monitoring. The data of SO_2 NO_2 PM10 PM2.5 O_3 and CO concentration mainly belongs to numerical or character. In MongoDB data correlation is described using document. Document objects can be regarded as a data record in RDBMS (Relational Database Management System). But they are not exactly the same. The structure of table is fixed which restricted the update and extend of data and correlation. But MongoDB do not have such restriction, and one document can be embedded as subdocument. In the end all data can be described using the format similar to BJSIN (Binary Java Script Object Notation). Such data are arranged according to the time sequence, so the searching performance can be improved by defining time as index.

MongoDB stored the data as document. It could store the monitoring data to the document every hour. This is quite similar to table used in relational database. Compared with storing single data in a document, this method improved the efficiency dramatically. The reason is by using bigger document, it could read fewer amounts of documents while reading data.
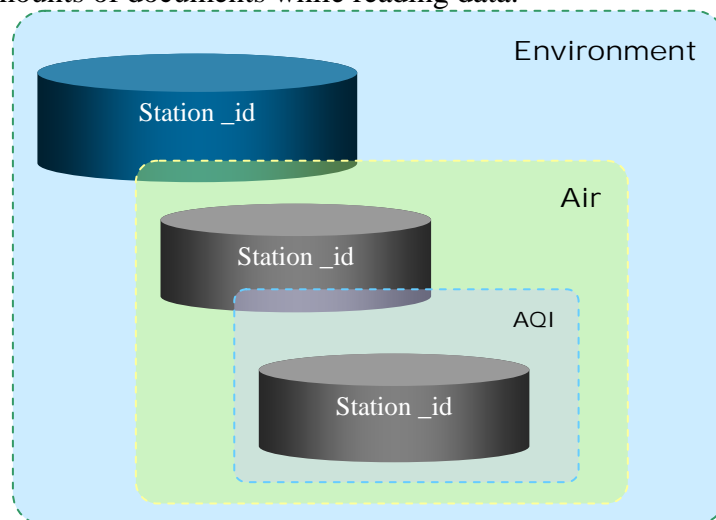

Fig.1 Support embedded subdocument

4.2 Big data storage in environment monitoring. Compared with ordinary air quality data, the data of Acid precipitation, surface water, underground water, acoustic environment, aquatic life etc. are mainly pictures, audios and videos. The traditional database is very inefficient in handling these semi-structured data and unstructured data. In MongoDB files are saved as GridFS which can be regard as a large hash map, file name are saved as key and file contents are saved as value. GridFS provide a mechanism which can divide big data to several small files, which can allow saving large

files. When various large files are stored in MongoDB database, it will be automatically broken up to many part, each parts only store one place in memory systems.

5 Summary

NoSQL database was developed to remedy the deficiency of relational database. We will still using relational database when relational database can satisfy the need. I chose Mongo to solve the visiting efficiency of air quality monitoring system. In order to satisfy the need for traditional database user, Various NoSQL database tried to compatible with SQL to reduce the difficulty. Making NoSQL developed from functional oriented to the scale of user groups. Along with the evolution of big data, we can predict the users of NoSQL database could expand from IT business to traditional business. We hereby only do analysis in environment protection.

References

[1] ROBBIN A Discussion of NoSQL Database —— Why Should Non-Relational Database [EB/OL]. (2009—1 1-25). [2013-08-16]. http://robbin.javaeye.com/blog/524977.

[2] Official Website. NoSQL [EB/OL].[2014-11-16].http://NoSQL-database.org/

[3] FAN Kai. An Overview of NoSQL Database[J]. Programmer, 2010(6): 76-78.

[4] CHARLES A BELL In-Depth Understanding of MySQL 2010

[5] Official Website. MongoDB[EB/OL].[2014-11-20]. http://www.mongodb.ors/.

[6] LI Guo-jie, CHENG Xue-qi. Study on Big Data: the Major Strategic Fields of Science and Technology and the Development of the Social Economy in the Future-Research Status and Scientific Thinking of Big Data[J]. Bulletin of the Chinese Academy of Sciences, 2012, 27(6): 647-657.