GIS Application based on Cloud Storage for Atmospheric Environmental Monitoring

Mei Han[1, a *],Ke Feng[2,b]

[1] Jining University Jining Shandong China

[2] Jining University Jining Shandong China

[a]hanmeisd@126.com, [b]fk831@163.com

Keywords: environmental monitoring, GIS, cloud storage, MongoDB

Abstract. GIS has been introduced to Jining's air quality automatic monitoring system. The system provides the client with easily available services of high quality. In order to meet the need of big data storage of the monitoring system we designed the distributed cloud storage framework to deploy MongoDB storage service. We designed the interface programs of GIS applications. The interface program make the client's cluster access to the MongoDB service more convenient and efficient and lays a foundation for the GIS to access the storage system rapidly.

Introduction

China is one of the serious air polluted countries, and the air quality in Jining is not optimistic either. In March 2015, the average concentration of fine particulate matter（PM2.5）in Jining is 71 ug/m^3, the average concentration of inhalable particulate（PM10）is 153 ug/m^3, the average concentration of SO_2 is 58 ug/m^3, and the average concentration of NO_2 is 41 ug/m^3. The number of the days with visible blue sky and white clouds is only 12. Facing the increasingly serious environmental pollution, the environmental protection workers of Jining have a lot of hard work to do. The most arduous one is the real-time, accurate monitoring on air quality and its changing trends and the supervision of pollution sources and their control. As the requirement of environmental monitoring is increasing day by day，environmental monitoring instruments can not only complete the online real-time monitoring and analysis but also adapt to multi task and multi parameter measurement, change monitoring objects at any time, accomplish different monitoring tasks, analyze the data transmission and exchange among monitoring instruments in order to realize the real-time measurement and control. There are 7 automatic monitoring systems for environmental air quality in urban areas of Jining. Based on this, Jining promotes actively the construction of the automatic monitoring stations for the atmospheric environmental quality in its administrative countries or cities. It has been planning or constructing new automatic monitoring systems for environmental air quality in 10 administrative countries or cities. The automatic monitoring for environmental air quality in the whole city will be realized in the near future.

In recent years, the computer technology and terminal equipment have been developing rapidly. GIS started a new field of application in environmental protection, effective utilization of resources, natural disaster forecast, and military command etc. The initial GIS managed geographic information data simply in a documentary way. In the 70s of the 20th century, with the development of computer cartography, the using of databases to store and manage geographic data gradually replaced the traditional GIS data management [1]. At the end of 20th century, the Internet developed vigorously. GIS focused on system and function also kept its focus on service applications. At present GIS is heading towards distributed data storage, networked applications and popularization. With the help of cloud computing, GIS applications pay more attention to the improvement of efficiency. It makes full use of network data resources to create the greater developing opportunities.

The Introduction of Geographic Information System to Environmental Monitoring

Geographic Information System (GIS) combines the functions of computer acquisition, storage, management, analysis, display and application of geographic information together to form a dedicated system [3], which fuses computer graphics with database system. Besides, it has been applied extensively in the field of environmental monitoring. The technology of GIS can be used to analyze and process the data of environmental monitoring. With its unique spatial analysis and visual expression function, GIS can meet the requirements of information processing that environmental monitoring demands, and can provide its users with readily available services of high quality.

GIS consists of the following five parts: hardware, software, spatial data, user and application model. It has 5 basic functions: data input, data edit, data storage and management, spatial query and spatial analysis, visual expression and output [4]. Among these functions, the key one is the spatial query and spatial analysis. The spatial analysis of GIS is unique, which makes it different from other computer graphic software. GIS is not only an information system, but also a management system for the analysis and description of geographical space [5].

Comparing with other computer technology, GIS has a relatively late start. But it is successfully applied in many fields [2] with its amazing development speed. According to the different ways of research and application, GIS can be divided into two types. One is to process the user's data by GIS. Another is to provide the user with services by dedicated GIS software which is further developed based on GIS.

Cloud storage framework to build MongoDB

GIS is an interdiscipline of geographic information technology and computer technology. Its rapid rise has led to the development of geographic spatial data storage. The status of big data in GIS applications is becoming increasingly important. The development of cloud computing drives the development of cloud storage services. With the development of cloud storage, NoSQL achieves rapid popularity. Many large enterprises and network giants select different database products according to their actual needs. MongoDB, as the leader of NoSQL, naturally becomes their preferred database. The author elaborates the method of spatial data storage from the aspects of GIS' development, applications and characteristics. According to the development of NoSQL, MongoDB is chosen as the database to store the environment monitoring data. A distributed cloud storage framework is used to build MongoDB. The performance advantages of storing geographic spatial data based on Cloud Storage are completely demonstrated.

MongoDB is a document database among the four types of databases of NoSQL. It has all the characteristics of NoSQL, and it plays a leading role in NoSQL with its rich functions. Some of its characteristics are different from other products of NoSQL. It has a high place in the practical applications of NoSQL. The service support of MongoDB can run under both Linux and Windows, no matter 32 bit or 64 bit. When running under the system of 32 bit, MongoDB has a limit of maximum file size, that is 2GB. There is no size limit to its documents while running under the system of 64 bit.

MongoDB replaces the concept of "line" in the relational database to "document" model. Such a storage model would be more flexible. Its unique data storage structure makes it possible to nest a document or an array. The complex hierarchy can be represented by one record. What is stored in is the very loose data structure similar to JSON format of BSON. It is easy to adjust data structure and data storage model flexibly. Some complex types of data structure can be easily stored. MongoDB has a powerful language query manner similar to object-oriented query, which can satisfy most of the queries in relational database[6]. In order to simplify the management of the database, MongoDB's autonomous management of the server is designed. In operation, there is only one action to start the server at the very beginning. If the main server is out of work, MongoDB will start a backup server to complete the work of the main server. These are completed automatically under the MongoDB's internal mechanism. Meanwhile the backup server will be upgraded to the current service supporting

server. Under the distributed environment, the cluster server will automatically configure a new node in the case of knowing an add node.

Deploy Storage Server

MongoDB cluster consists of config server, mongod, mongos, clients and arbiter etc. Specific configuration is analyzed and deployed based on the characteristics of different data and the amount of data. Fig. 1 show the structure of MongoDB's distributed cluste.
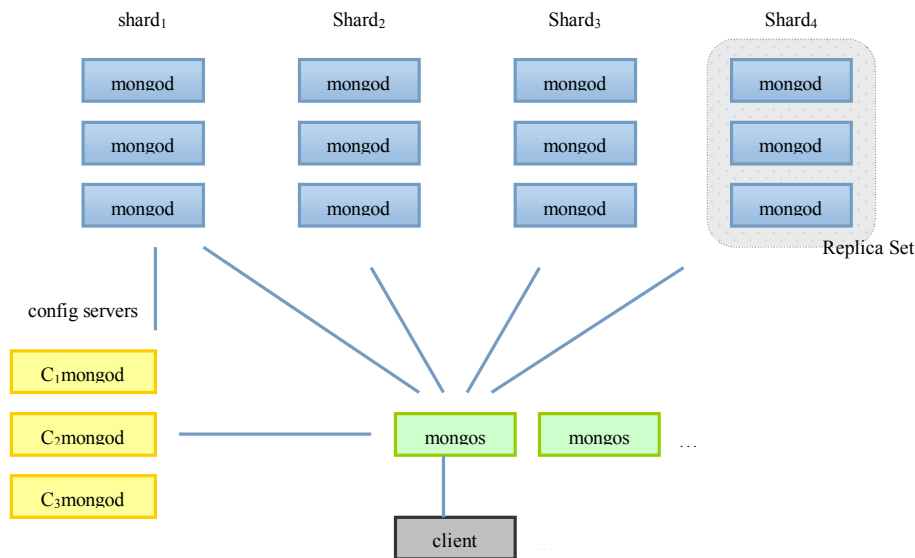

Fig. 1 the structure of MongoDB's distributed cluste

Config Server. It mainly stores the cluster metadata information, including each shard, shard's server and the basic information of chunk. There is a segment on config server to backup information, which ensures the consistency of the data on config server[7].

Mongod: It is the shard server of MongoDB. It mainly stores fragment data, namely data block. Shards: It is the fragment of MongoDB. Two mongods form a fragment, that is a group of data. The data in the two mongods that belong to one fragment is the same. A group of mongod can be designed into a master-slave structure or mutual master-slave, which ensures the consistency of data. In addition, the data on each fragment belongs to certain range. For environmental monitoring data, the data of a city can be regarded as one fragment. According to the size of the data, the fragmentation deals with the load balancing by automatic segmentation or data's autonomous transfer.

Mongos: It is the route server of MongoDB. It is mainly responsible for routing and coordination work as a console. So the whole cluster looks more like a database services system. In practical applications, there can be multiple mongos, and it can be placed on any server. After starting the mongos, it gets the metadata information of fragment or block from config servers, and then it responds to the clients' requests. The content of the request is routed to the corresponding data storage server via shard server. Finally the searched results are fed back to the clients.

In the deployment scheme of distributed cloud storage for the environment monitoring data, we adopts the method that the insertion and update are separated from the query and access and it deploys separately multiple reading and writing server. This separation method has solved the problem of the efficiency of monitoring data storage and the reliability and extension of the database service. Using the deployment architecture model of Replica Set + Sharding ensures the high usability of the storage of monitoring data in MongoDB[9]. The data of environmental monitoring is numerous. To update spatial data regularly in the change of geographic spatial information, this architecture opens up the possibility for the horizontal extension of the database. At the same time, if the fragment needs adding just to change the corresponding replSet name. This is very convenient for the expansion of our data

storage. At present a lot of large data storages are basically using this architecture to deploy MongoDB storage system.

The Designing of Clients Module

Environmental monitoring data is stored on a distributed server of Cloud. If GIS application on the clients have a need to visit or access to environment monitoring data, it should be achieved through the interface program. Thus, We need to design a file interface system for this kind of storage architecture on the clients to adapt to the clients' access requirements. And a mature and efficient interface program is rather important to promote the performance of the storage system. Our interface program is written in C++. By this program, we can write new environmental monitoring data and a large quantity of data initially, update data, query and delete outdated data according to different requirements.

MongoDB Service Configuration Access. The interface program needs to be connected with the MongoDB server first. Distributed deployment requires users to open the MongoDB service, and then access to the Mongo DB service configuration through the program. In order to achieve access to the data, the program needs to handle the information of port, Ip and user ect. first, and then access to Mongo DB server. The program is connected to the Mongo DB database through the Get Connection() function. Specific configuration services are stored in the configuration file gis_conf. By obtaining the configuration information in gis_conf, the default database connection is obtained. Specific core pseudo codes are shown as Get Connection():

```
Get Connection()
  if(CONF.empty()) {
   Read Config("gis_conf", CONF);
  }
  string host = CONF["host"];
  string port = CONF["port"];
  string database = CONF["db"];
  string user = CONF["user"];
  string password = CONF["password"];
  DBClient Connection * dbconn = new DBClient Connection(true);
  conn(dbconn);
  dbconn ->connect(host + ":" + port);
  dbconn ->auth(db, user, password, err);
  return conn;
}
```

Write Data Interface Program. The single geographic spatial data or a batch of spatial data is written to the MongoDB storage files. Geographic spatial data and attribute data of geographic information are written to Mongo DB. Write () function and related parameters are as follows:

int write (char *fname, unsigned int type,Point center, Point * boundary, unsigned int boundary_size, Relation *relations, char *attributes, unsigned int attributes_size, File Object * fo)

Among those, fname is file name, center is the center coordinate of spatial object, boundary is the anchor point coordinate of spatial geographical object, relations are the ones of an object and the other one; attributes are data sets of object's attributes, fo is a file object set.

Read Data Interface Program. The specific implementation of the reading interface is based on the needs of GIS applications. The reading of environmental monitoring data can be divided into the following categories: query by file name, query by point including this point, query objects related to certain point by relation, and query all the objects included in the range by range. For different requirements of applications, different query interfaces are designed.

Delete Data Interface Program. The deleting of interface program is similar to the reading of interface program, which requires to delete according to different fields. It can be divided into 3

categories: delete according to file name, delete object data including the point according to point, delete object data included in the range according to range.

The client is the operating platform of environmental monitoring system. The sample data obtained from the server are displayed in various forms in the client, which provides data for monitoring analysis. The client is an analysis Center as well. All kinds of analysis functions can help to draw a conclusion. GIS application program is responsible for receiving the client's logic control to the program, proposing transfer requests, controlling of the data analysis methods and results' showing forms. The designing and using of the interface program for GIS applications realizes the clients' convenient access to the MongoDB's cluster service. At the same time, the access to the data is optimized at the code level. All those lay a solid foundation for the GIS to access the storage system rapidly.

Summary

The integrated application of distributed cloud storage framework and GIS system for environmental monitoring data has obvious advantages comparing with the traditional storage structure in dealing with large amounts of data. In a certain period of time, the system can meet the requirements of Jining's environmental monitoring system in the storage performance and efficiency. It will greatly promote the healthy development of Jining's environmental protection work.

References

[1] Quantum GIS. Development Team, 2012. Quantum GIS Geographic Information System. OpenSource Geospatial Foundation Project [J]. URL: [http://qgis.osgeo.org], 2011.

[2] Kosters G, Pagel B U, Six H W. GIS-application development with Geo00A [J]. InternationalJournal of Geographical Information Science, 1997, 11(4): 307-335.

[3] Zhong-wei Sun, Li-wei Shao, Chun-liang Hou. Information and communication technology on the development of GeographyJournal of Hebei Normal University: Natural Science Edition, 2004, 28(5): 533-537.

[4] Jian-shong Li. Geographic Information System Principle [M], Wuhan University press, 2006.

[5] De Smith M J, Goodchild M F, Longley P. Geospatial analysis: a comprehensive guide to principles, techniques and software tools [M]. Troubador Publishing Ltd, 2007.

[6] Membrey P, Plugge E, Hawkins T. The definitive guide to MongoDB: the noSQL database for cloud and desktop computing [M]. Apress, 2010.

[7] Shuai Tian. Design and implementation of a large scale remote sensing data storage system based on MongoDB and HDFS [D]. Zhejiang University, 2013.

[8] Yang C, Good child M, Huang Q, et a1.Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing [J]. International Journal of Digital Earth, 2011, 4(4): 305－329.

[9] Lin J X,Chen C C,Wu X Z, etal. GeoKSGrid: A geographical knowledge grid with functions of spatial data mining and spatial decision[C].2011 IEEE International Conference on Spatial Dat a Mining and Geographical Knowledge Services(ICSDM),2011,121－126.