Research on Key Problems of Database in Cloud Computing

Hua CHEN*, Jing GE
Institute of Technology, East China Jiaotong University, China
*s021713@163.com

Keywords: Cloud computing; Database; Key problems.

Abstract. In recent years, the society has been developing rapidly, and it has also led many new industries. The database in cloud computing is one of these industries. At present, the relational database has become a bottleneck in cloud computing, and restricts the agility of WEB structure and the development of expansibility to a certain extent, so the NoSQL that can break the constraints of RDBMS paradigm are concerned by more and more people. According to the BASE theory, it is easy to be used and highly scalable, and supports the massive data as the main target and has been used widely in the field of cloud computing. This paper mainly analyzes the key problems and technologies of the database in cloud computing.

Introduction
With the continuous development of Internet, a number of related applications emerge. Cloud computing is a relatively new network application, showing a critical value of network computing. Cloud computing is the result mixed by virtualization, public computing, infrastructure, platforms, software and so on. Through this technology, network users can process a large amount of information in a very short time and provide more efficient services.

Research Situation and Significance of Database in Cloud Computing
With the advent of the information age, information resources have become a kind of wealth and a resource that all fields are required, and these information resources usually use the database to store and manage commonly, so there are the same way data storage and different data management methods. In the process of data management, the management methods generally differ with the change of different application fields. Cloud computing has produced a great application of database, large scale of data begins to appear, its high concurrency and high flexibility raise higher requirements for the relational database in the past, a new data structure also emerges and it is related to the solutions of NoSQL.

NoSQL is a non-relational database. Although its functions are not strong, its scalability is abundant compared with the traditional relational database. It can change with demands, better adapt to the actual situation, only provides some necessary functions for users. At the same time, it can get rid of the relational database with a complicated structure in particular situation. Although the past object-oriented database and XML database having a good performance is very good, because the database migration has risk and costs are relatively high, and the obtained data size is relatively small. These will hinder the stable development of the non-relational database, which causes the that the relational database still has a large market share at present and is almost a monopoly in the market. But the emergence of cloud computing has brought some changes. The operation of cloud computing needs to process the large scale of data, and the way of RDBMS processing in the past is unable to meet the requirement. The emerging MapReduce is a very good method, and a new way while Hadoop process the large scale of data under the framework of cloud computing. It is based on the non-sharing processing mode, and the processing capacity, the space size of disks and band width of the processor will not affect its actual efficiency, which enables some new data processing data to be used in old modes. The development of cloud computing has made the use of relational databases less and less. This is no longer a new thing. At present, the development of Amazon and Google has already verified this view.
Key Problems of Database in Cloud Computing

Although the relational database has a very important position in the domestic market, there are still some restrictions that affect the efficiency of its application. First of all, it is relatively difficult to expand. The core development of relational database system is to combine the reality, and it can build models based all data structures in principle, and ensure no redundant or missing information after the standardized procedures. But it also causes some problems. For example, in the process of actual query data, the multiple-table join query mechanism, which causes problems in the expansion of database. Secondly, the speed of reading and writing is slow. When the number of the whole database system reaches a certain extent, because the relational database system has a complex logic, so deadlock and concurrency often happen in the process. The speed in reading and writing will drop, which will affect the normal reading of data. Thirdly, the cost of use is high. Usually, the price cost of the database used by enterprises is relatively high, and with the increase of the system size, the cost will rise, which is also a difficult problem to overcome. Fourthly, the capacity of the support is limited. At present, relational databases have not been able to support massive amounts of data. For example, the company like Google requires a lot of data storage. The fifth is the unauthorized access. Unauthorized access mainly refers to that free access to network resources in computers without corresponding authorization. In general, the priority access is provided by the cloud providers, which will result in the unreasonable allocation of resources in cloud computing. The sixth is data synchronization. The database in cloud computing usually uses the relatively redundant storage methods, so users need to create more backups and then forward in the process of using. Although it has very high reliability and availability, but also ignores data consistency; the seventh is the confidentiality of data. The security is always the problem that cannot be neglected during the development of cloud computing. Because the transmission channel of cloud computing is open, so data is easy to be lost in the transmission and preservation of data, and many criminals will attack and steal data in the process, which is a big risk.

The Technologies to Realize Cloud Computing

Interface Display Layer

The main function of the display layer is to provide all the services offered by the middle level location for customers in a friendly way, including the following parts. First of all, these are HTML, CSS, JavaScript, which play a very important role in the WEB development process. It can be said a gold combination. HTML is a very standardized Web page technology, and the newly emerging HTML5 has made up for the problems in HTML4, such as the problems existing videos and local storage. At present, the Apple browser used by many people began to support some functions in HTML5. CCS can redefine the format of some elements in the page, which can control the appearance of Web page artificially, and separate the contents and the expression form of the page. JavaScript is a popular technology now. It can be used to enrich the dynamic language of Web page function, and control the page's logic and layout management. In addition, it is rich client technology. Nowadays, there are more Flex and Silverlight on the market. Flex technology is formed on the basis of Flash, users' experience is usually good, and they usually are used with JS, which can provide some rich applications that HTML technology can't provide. Although the utilization rate of Silverlight in the market is lower than that of Flash, it comes from the industry’s giant Microsoft. It can provide a good way of operation for developers.

Middleware Layer

The middleware layer plays a connecting role in the whole cloud computing. The underlying infrastructure layer provides the service that is used to support the display layer on the basis of providing the available resources. The main technologies of the Middleware layer includes the following types. The first one is REST technology, and it provides the services of middleware layer for the callers conveniently, and many non-relational databases support REST technology better. The second is parallel processing technology. The representatives in this aspect are mainly some
large-scale parallel frameworks, which can handle large quantities, and usually use cluster method to achieve this capability. The third is the application server. In cloud computing, the application server is generally optimized to some extent, and then more applications are carried out on this basis. The fourth is the multi-tenant form. It can allow many users to share the same application instance service, so that it can greatly reduce cost expenditure and reach the goal of maintenance cost finally. And users don't have to worry about the interaction between them, because cloud computing itself has good security measures and isolation measures. The fifth is distributed caching. In the application of cloud computing, this technology can reduce the pressure of the server, enhance the response speed of the server, and improve the efficiency of the server.

**Infrastructure Layer**

The infrastructure layer is mainly to provide the required computing and storage resources for the upper layer, basically including four technologies. The first one is virtualization technology. Affected by the virtualization technology, a super computer becomes multiple servers, and isolation is achieved between two servers. At the same time, the CPU and memory of the super computer can adjust and reallocate flexibly among multiple virtual machines. The goal of one to more is basically realized, which greatly reduce the use cost. The second is the distributed storage technology. The data in the cloud will grow rapidly, in order to satisfy this requirement, there are many specialized distributed storage systems, such as Google's GFS. These are the extensible and distributed file system and can be used in the large distributed software systems with abundant data. Moreover, if they are used on the cheap normal equipment, they can provide the highly quality services for users and achieve the basic functions of fault tolerance. The third is the relational database. Now, some characteristics and problems of relational database have become bottlenecks in cloud computing, so they need continuous optimization. The last one is NoSQL. For the cluster on the scalable and low-level PCs, cloud computing has raised some new requirements, require more new data models, which enables the key value database show up and many databases are almost non-relational no longer.

**Management Layer**

The first is safety management technology. Cloud computing has some characteristics in the data, all the data is maintained by the third party rather than the first party, so the safety of data has become a big obstacle, which needs to be solved urgently, otherwise, it will cause loss to enterprises. The second is SLA monitoring technology. The SLA technology in cloud computing focuses more on the content of virtual machine and quality of service. It is necessary to ensure that all indicators can be controlled within the prescribed limits of SLA. At last, it can provide users with better and higher quality services. The third is billing management. The resources should be recorded and counted timely and accurately in the process of using, which can facilitate to charge users the corresponding costs later. The fourth is the load balancing. The system runs between every virtual device, and it can accomplish the automatic load balancing among multiple resources, so that the traffic used by the system can be redistributed. The fifth is the operation. In the process of system operation and maintenance, it is necessary to ensure its specialization and automation as far as possible, thus reducing the operation and maintenance costs of the cloud computing center.

**Database Design**

**Disadvantages of Database Modeling**

The first is the disadvantages in performance. In the process of multi table access, because the connection between data is needed, it will increase an operation of database and an IO operation. So it will have an impact on its performance. The second is the Schema problem. The modes of relational database are all defined before, so if the actual business needs to change, it will make a great change to the whole system. Moreover, if a relational database is used in semi-structured data, many of the contents are empty, or there are a large number of operation of connection tables, which will lead to very poor performance.
**Description on Entity Relation**

Among the cloud computing databases, the most common one is one to one and one to more relationships. Because the composition of systems have an inclusion relationship, which is that each center involves several rooms, and a room including many building floors, each floors has many rooms, and each room has frames and servers, which is an obvious relationship of one to more. In relational databases, the relationships of one to one and one to more are described by foreign keys. Another is embed, which is only suitable for a relatively small number of entities. Otherwise, it will affect the performance of database because of too large documents, and if all the modifications of nodes are placed in this document, it will restrict normal operation.

**Document Update Validation**

There are some similarities in triggers in relational database. The function of CouchDB is to verify before document creation and update. Only passing the verification, documents can be saved to databases. The method of validation is also defined according to some fields in the design documents. In general, all documents will be used to this method when they are updated. If an exception is thrown with this method, the checkout failure is explained. At this time, CouchDB returns the error information in the exception to users.

**Control of Access**

At present, CouchDB only supports system administrator, and it can modify database based on any program, and can add account and password of system administrator in the configuration file of CouchDB. At the same time, CouchDB also supports the basic authentication of HTTP, and can also use this authentication method in the configuration file. At present, CouchDB is relatively weak in controlling accesses. The best way to use servers as the reverse of CouchDB, and then directly uses server directly to process the control access.

**Summary**

With the development of the society, the demand for information resources is becoming larger and larger, and a huge amount of information is transmitted every day. In particular, in some enterprises, the transmission and storage of information is needed at every moment, which raises more requirements to databases. At present, there are still some limitations in database of cloud computing, such as speed of reading, expansion, security and cost. By analyzing cloud computing and database technology, this paper can better support the rich client technology of display layer and the REST technology in the middleware layer, hoping to provide help for database development in cloud computing.

**References**

