

Application and Development of Database Technology in the Background of Big Data

Bing Li¹, Juan Wang², Ning Li³, Minghua Zhao¹, Feng Wang¹

¹School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China.

²Xinghuo Road Primary School, Lianhu District, Xi'an 710014, China.

³Renhouzhuang Primary School, Beilin District, Xi'an 710000, China.

Abstract. The development of big data makes us face many problems when we operate traditional databases and use hundreds of TB or PB data. Many industries such as finance and Telecommunications now involve a lot of data, and a lot of data for the original ones. The database is a great impact. At present, the database technology in the world has made great progress. Many kinds of database technology emerge in endlessly. This paper mainly studies the development status of database technology under the background of large data.

Keywords: big data, database, current situation, trend.

1. Preface

Large data, also known as mass data, is widely used all over the world, which has caused a wave of new data technology reform. In five years of 2011, the concept of big data was introduced in the United States. In December 2011, in the 12th Five-Year Plan of the Internet of Things launched by our country, some technological innovation projects were launched, many of which are related to big data, which fully demonstrates our country's support for the development of big data. With the change of the times and the progress of science and technology, database technology support under the background of big data is required in many fields, such as railway ticket purchasing system, Taobao Shuang11 activities, etc. Many aspects of our life must be applied to stable and reliable database. However, in order to achieve these, we must rely on the current big data. Data technology, the old database structure we used in the past is no longer suitable for the current social development. We must rely on the technology of big data and strive to upgrade the technology of database, so that the development of database can be more complete and fully adapt to our life. This is also our database development and scientific research. The direction of personnel efforts.

2. Overview of Big Data

2.1 The Concept of Big Data

Big data, which usually means that the common software can not deal with the data set in the prescribed time, needs a new processing method to deal with a large number of data. In summary, the characteristics of big data are 5V, which are large, high-speed, diversified, low-value density and authenticity.

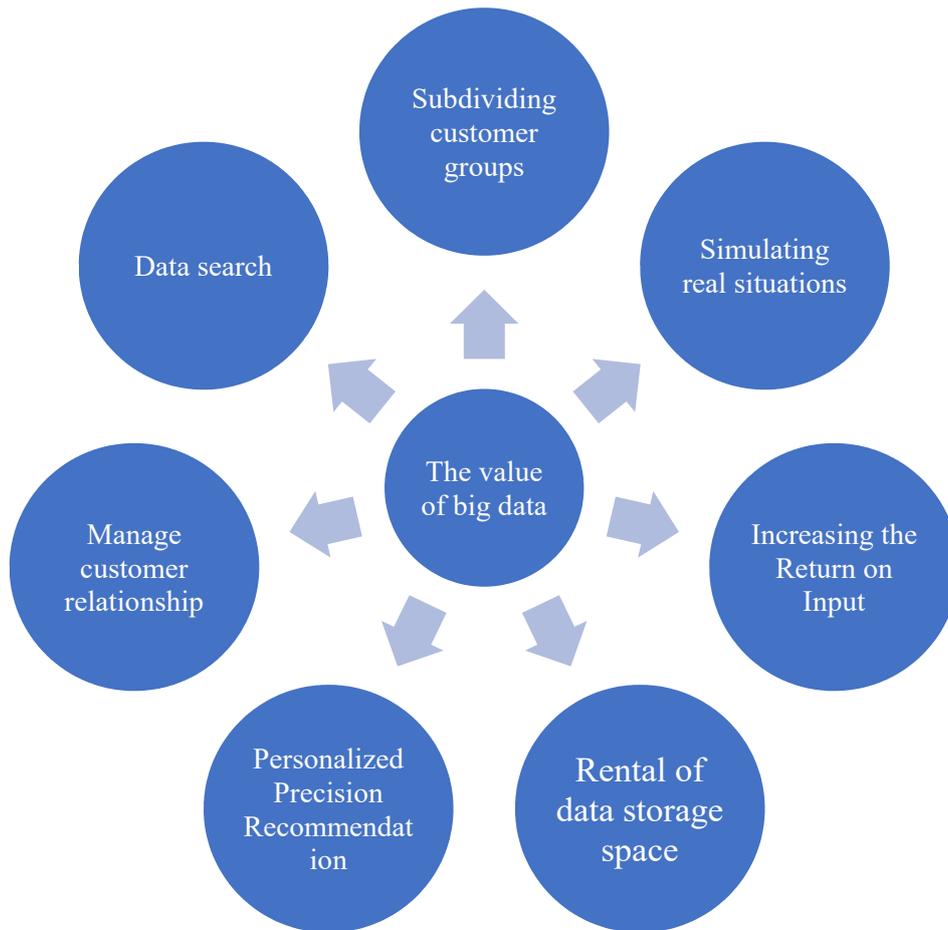


Figure 1. Business value of big data

The main characteristics of big data include the following aspects:

Huge quantities: Big data involves a lot of data content and is growing every day and every minute, but our ability to process and analyze data is not growing so fast.

Types are tedious: There are many applications of large data, which come from all walks of life, and the types of data are different.

Fast: We are now changing the types and quantities of data collection, so the corresponding processing speed has also been growing rapidly, because only the accelerating processing speed can match the existing data development model.

Commerciality: In our age, a lot of data information is with strong business information, and all have a fixed life cycle. With the passage of time, their value is getting lower and lower.

2.2 Classification of Big Data

It is necessary for us to classify big data. We usually divide big data into three parts: network data, decision data and scientific data. Network data, which we use most, is usually related to the network data, including web pages, logs and so on, as long as the data related to the network are in this range. Decision-making data, which is usually referred to as some content of old database management. Scientific data is the scientific data that we usually produce in various aspects such as scientific research [1].

Based on the above analysis, we can represent the big data as follows.

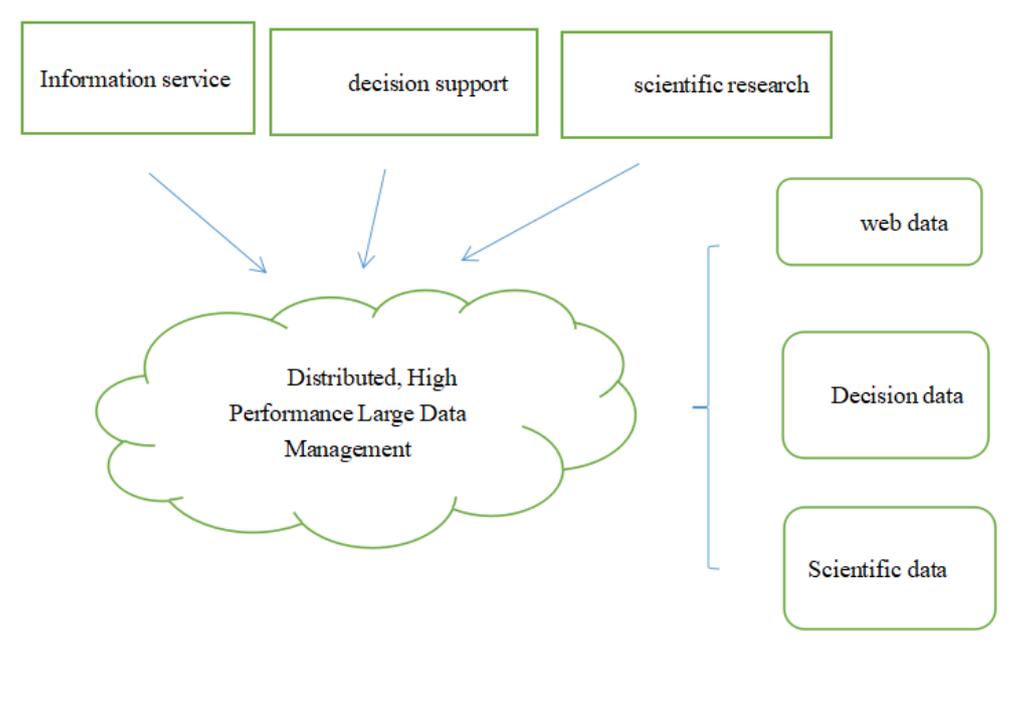


Figure 2. Classification of large data

3. Innovation and Development of Database in the Age of Big Data

At present, the indexing and database undertakings in the era of big data have created many important basic new technologies, as follows:

3.1 Cloud Computing

Cloud computing is a very widely used technology. It uses the idea of sharing infrastructure, can maintain and manage computer resources by itself, and provide users with convenient, fast and efficient resources. Cloud computing is the future direction of development in the field of database. At present, there is no clear definition of cloud computing. According to the authoritative technical departments of the United States, cloud computing is a model in which users can easily access related resources. These resources can be quickly used and made public. We generally call cloud computing "cloud", which is generally referred to as the related resources of servers on the Internet, these resource packages. Including two aspects, hardware resources and software resources, users can use the corresponding resources when they use them locally and send out the corresponding request information. Cloud database has the following advantages: it has infinite scalability, good use effect, according to the need to use payment, can take parallel processing. Cloud computing and big data are closely linked. The platform of big data gives the platform of cloud computing. Cloud computing brings data processing solutions for big data.

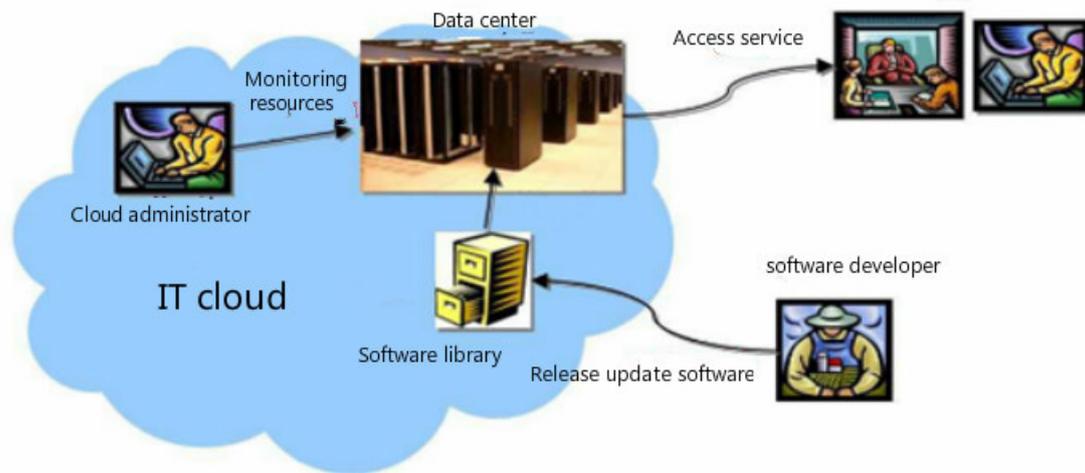


Figure 3. Cloud computing architecture technology

3.2 Virtualization

Virtualization technology is a key part of cloud computing, including network, computing and storage virtualization. By virtualization technology, several virtual computers can be run on one host at the same time, and different users can use high performance devices, which greatly saves the related hardware costs. Server virtualization technology can fully improve the efficiency of hardware resource utilization, make system performance fully mobilized, and more applicable. These are all very important means to effectively improve network stability and performance. Of course, in the development of virtualization, there are also some problems, such as loss, overload, and so on. These are late needs. The key to further research. From the user's point of view, the system architecture of data virtualization mainly includes three aspects: application layer, data virtualization layer and source data layer. The application layer is a data query and access application for front-end users. It is responsible for feeding back users' data query requests to the data virtual layer. The application layer provides users with a variety of data access interfaces. Users can choose corresponding data access interfaces according to their own needs. As the core of data virtualization system architecture, data virtualization layer mainly covers two key planes: data virtualization plane and management plane. Data virtualization plane is mainly responsible for data abstraction, modeling, data source mapping and other functions. The management plane configures, manages and maintains the system. Source data layer is like a comprehensive interface, which is responsible for the unified management of different forms of data provided by all data holders, and data acquisition and transmission through API, JSON and other interfaces [2].

3.3 Distributed Cluster Database

Distributed clustering database is a great achievement in the development of database technology. In distributed cluster databases, each computer is often stored in a separate location, each computer can save a copy of the database, and each computer unit has its own independent local database. Each computer is connected through a network, thus forming a large database which is logically complete and physically dispersed.

The database in distributed cluster system is a logically complete database for users. In distributed cluster system, database integrity, security and data consistency are guaranteed. The distributed database can manage the shared data uniformly, and the client can also operate the non-database part. External link technology is the main way to control distributed cluster database remotely. In distributed cluster database, each computer transfers information by means of network. Users can access the data of each database unit remotely and transparently in the system, and can also access the data of multiple databases in a combined way. Because each database unit is physically distributed independently, external linking technology can logically connect these previously independently distributed database units into a complete entity. From the overall perspective, distributed cluster

database is a complete system. Users' single access to the database will be “automatically decomposed”, “automatically addressing”, “automatically transformed” into corresponding network requests [3].

Distributed cluster databases usually adopt hybrid network topology structure, and use hybrid network protocols to automatically complete network protocol conversion. In order to ensure that the database can store large amounts of data, the distributed cluster database uses high reliability and high availability cluster to improve the reliability and practicability of the database, so as to meet the needs of Internet users.

3.4 Data Analysis

Data analysis and processing under the background of large data should be developed continuously according to our current requirements for data processing. On the one hand, it should be able to process related data quickly, on the other hand, it should be able to process real-time data. At present, there are two technical directions for the processing of large data in the industry, one is parallel relational database, the other is distributed database. Of course, the technical aspects of database still need to be carried out according to the requirements of specific applications. Data analysis in the context of large data is the most promising application technology, which mainly includes six aspects, as shown in Figure 4, which are visual analysis, data mining algorithm, predictive analysis ability, semantic engine, data quality and data management, data storage[4].

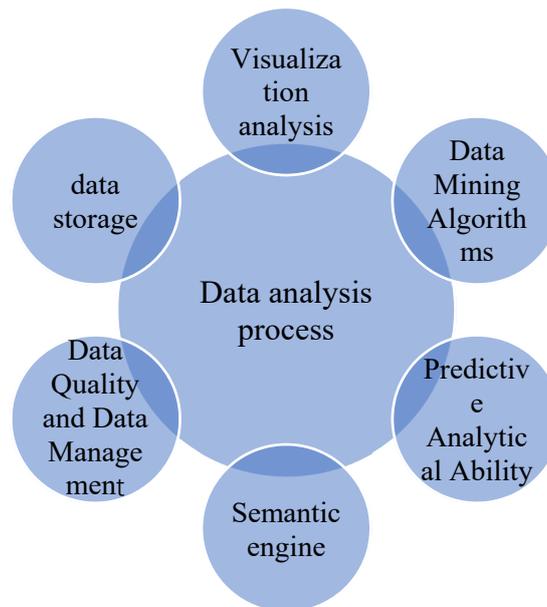


Figure 4. Data analysis process

Data visualization analysis, that is, using computer technology and image technology, presents complex data to users in the form of graphics, so as to facilitate their analysis of data. Data mining algorithm corresponds to data visualization. Data visualization is to enable users to analyze data more clearly, while data mining algorithm is to enable machines to analyze and process data more quickly and accurately. Generally speaking, clustering and segmentation algorithms are commonly used in data mining and analysis. Predictive analysis ability, which is a data analysis method based on data mining algorithm, aims to help analysts make some accurate prediction of data mining results. Semantic engine, unstructured data has diversity, it needs corresponding tools to analyze and extract data, and semantic engine is to achieve this function. Data quality and data management is to process data by means of standardized processes and tools, so as to ensure data analysis to achieve a desired result. Data storage is to build a data warehouse, and then store the data in the data warehouse according to a specific mode. The stored data can be queried and accessed by users according to the subject.

3.5 Information Indexing Technology

At present, all kinds of conditions of our computer are very superior, which have laid a solid foundation for the development of indexing technology, enabling information indexing technology to develop to a high level. At present, considerable progress has been made in engine intelligence, multilingualism and so on.

4. Summary

The emergence of big data has led to a great improvement in our database technology, which is also a great technological breakthrough. It also gives our users a new experience in using the database technology, pushing the application of database technology to a higher level. Big data is a broad concept, which needs to be closely linked with the actual application. Relevant research should be carried out in various fields. The first thing we need to do is to have a clear choice for the industry of big data application, so that big data technology and systems can play a full role, so as to promote the progress of database technology.

Acknowledgements

Science and Technology Project of Xi'an City [2017080CG/ RC043 (XALG021), 2017080CG/ RC043 (XALG011)].

References

- [1]. Li Zhanhuai, Wang Guoren, Zhou Aoying. Research progress and trend of interpreting big data from the perspective of database [J]. *Computer Engineering and Science*, 2013, 35 (10): 1-11.
- [2]. Liu Jun, Zhang Xin. Construction of the database of sci-tech periodicals from the perspective of big data [J]. *Editorial Journal*, 2014, 26 (01): 59-62.
- [3]. Wang Shufeng, Fei Xianju, Hu Zhixi. Research on inquiry teaching method of database course based on big data [J]. *Computer Education*, 2014 (11): 99-102.
- [4]. Liu Wei, Xia Cuijuan, Zhang Chunjing. Big data and related data: the coming data technology revolution [J]. *Modern library and information technology*, 2013 (04): 2-9.