

Research on Big Data Service and Its Key Technology in Cloud Environment

Ying Liu

Xijing University, Xi'an, Shaanxi, 710123, China

Keywords: Big data service, Key technology, Cloud environment

Abstract: With the continuous development of society, a large amount of information data has been produced in human life and production. Especially in the current era of the Internet of Everything, the widespread promotion of computers and mobile terminals has transformed information data into an indispensable part of people's life and production. It affects people's daily life, work, and learning. Under the environment where these information data are constantly growing and information data is becoming more and more common, there is an urgent need to strengthen research on data storage and data processing to improve people's data extraction efficiency, quality, serving people's lives. By elaborating the relationship between big data and cloud computing, the article analyzes cloud computing and its key technologies, and discusses the application of big data processing technology under cloud computing. It aims to provide some ideas for studying how to promote the orderly and healthy development of big data.

1. Introduction

In recent years, the rapid development of big data in the world has aroused great attention of academia, industry and governments of various countries. Big data will bring significant development opportunities and technical challenges. On the one hand, by effectively managing big data and analyzing and extracting its value, it can provide high-value-added applications and services to the industry and realize huge economic and social value. On the other hand, big data brings huge technical challenges while bringing development opportunities. Traditional computing technology will face many technical difficulties when processing big data. Therefore, it is necessary to research and find new effective technical methods to complete the analysis and value discovery of big data. Big data processing is a complete technology that includes and integrates multiple levels of big data storage, calculation, and analysis. Academician Xu Zongben pointed out in the keynote speech of the 2016 Big Data Academic Conference of the Chinese Computer Society: "Big data technology requires multidisciplinary comprehensive research, involving data acquisition and management, data storage and processing, data analysis and understanding. Combining big data applications in the field, etc. "

In today's era, with the development and popularization of computers and information technology, the flow of people, funds, commodities, and information in the production and living activities of human society is presented in a digital manner, and society is being fully digitized, which has triggered industry data. The explosive growth has promoted the rapid development of big data. Big data has been applied to many industries, and huge data resources have become the strategic resources of countries and enterprises. In 2012, a report released by the World Economic Forum pointed out that big data is new wealth and its value is comparable to that of oil. The McKinsey consulting report believes that data is a means of production, and big data is the next frontier of innovation, competition and productivity improvement. The author of "Big Data Era" believes that big data will start a major era transformation. He pointed out that big data will bring tremendous changes, change our life, work and thinking, change our business model, and affect our political, economic, technological and social aspects.

Large-scale data resources contain huge social and commercial value. Effective management of these data and mining of the in-depth value of the data will have a huge effect and influence on national governance, social management, corporate decision-making and personal life. Therefore, the ingenuity and application of big data has become an important driving force for global

technological innovation and economic development. However, large-scale data resources bring new development opportunities to people, but also bring many new technical challenges. Industry big data with diverse formats, complex shapes, and large scale brings many technical difficulties to traditional computing technology. Traditional database and other information processing technologies have been difficult to effectively deal with large-scale data processing. To this end, people urgently need to find effective big data processing technology methods and means to effectively process and analyze the application of big data in the industry.

2. The Basic Definition of Big Data

Regarding the concept of big data, it is difficult to have a very quantitative definition. Existing definitions are qualitative descriptions from the perspective of data scale and supporting software processing capabilities. For example, the qualitative description of Wikipedia is: big data (refers to a data set that cannot be acquired, managed, and processed within a certain period of time using traditional and commonly used software technologies and tools; the definition given in McKinsey Consulting's big data report Yes: Big data refers to data sets whose size exceeds the capabilities of conventional database tools to acquire, store, manage, and analyze. In fact, the focus of the term "big data" today has gone far beyond the definition of data scale. It represents the development of information technology to a new era, represents the new technologies and methods required for massive data processing, and represents the new services and new values brought by the application of big data.

3. The Main Technical Characteristics and Difficulties of Big Data Processing

The technology is comprehensive and cross-cutting; big data processing is a comprehensive computing technology involving many levels of computer technology. As pointed out by Xu Zongbenyuan, big data technology requires multidisciplinary comprehensive research, which involves data acquisition and management, data storage and processing, data analysis and understanding, and big data applications in the combined field. A complete big data processing and application system is usually a complete technology search that includes and integrates large-scale hardware resources and infrastructure management, distributed storage management, parallel computing, analytical mining, and application services. Therefore, big data processing is highly technically comprehensive and cross-cutting.

The large scale of data, the failure of traditional computing methods and systems, and outstanding computing performance issues: Big data has brought many new challenges to traditional computing technologies. The huge amount of data will cause huge calculation time overhead, which makes it difficult for traditional calculation methods to complete processing within an acceptable time when facing large-scale data. The huge amount of data or the amount of calculation presents a huge challenge to big data processing technology in terms of computing performance.

Application demand-driven features: Many problems of big data applications come from specific industries. Big data processing has strong industry-driven demand-driven features. Therefore, big data processing must be closely integrated with the actual scenarios and needs of industry applications. Starting from the actual application needs of the industry and combining the actual application needs to solve the technical problems in big data processing, so as to effectively use big data technology to improve industry information processing and Service level, explore the deep value of the industry. Because big data technology has typical industry application-driven characteristics, this also requires cross-fusion between the application industry and the computer field. As summarized by Xu Zong's hospital: data resources are the foundation, the processing platform is the support, the analysis algorithm is the core, and the application benefits are fundamental.

4. Big Data Processing Technology in the Context of Cloud Computing

Traditional data management is mainly based on collection and storage. Under the background of cloud computing, big data management methods have been changed and innovated. The focus is on data analysis and mining to provide a basis for relevant personnel to make decisions.

Big data collection technology. According to different collection forms, big data collection can be divided into two types, namely centralized collection and distributed collection. These two collection methods have their own advantages and disadvantages, and we take the advantages as an example to illustrate. Centralized collection can control global data, and in terms of flexibility, the advantages of using a distributed collection mode are more obvious. In the process of big data collection, it includes both the collection of internal data of the enterprise and the collection of information data between enterprises. Through distributed parallel computing mode, several collection modes are mixed to improve the overall efficiency of data collection. That is, in the process of big data collection, a centralized collection mode is adopted for enterprises, and a distributed collection mode is adopted between enterprises. In each enterprise, multiple central servers are set up to store information and data shared by the enterprises. For the organization between the central servers, the distributed data collection mode is adopted.

According to different structure types, big data can be divided into structured data, semi-structured big data and unstructured data. In the process of data collection, the data types should be analyzed first, according to different types, through the advantages of cloud computing expansion, fault tolerance, etc., to achieve the homogenization of the data, to achieve the integration of data of various structures.

Big data storage technology. One big characteristic of big data is “big”. If traditional data storage technology is adopted, it will be difficult to meet the storage requirements of big data. The reasons are as follows: (1) In the context of the era of big data, the amount of data is growing rapidly, and the traditional single-node data warehouse appears to be unable to deal with massive amounts of data. (2) Row-based storage is the main form of traditional data warehouses, but the maintenance of views and indexes requires excessive costs. In the context of cloud computing, big data storage is dominated by columnar storage. This method is mainly stored according to the attributes of the data. In contrast to storing by rows, each attribute is stored in a column. In the process of data projection, only the attribute columns involved can be accessed, which greatly improves the input / output efficiency of the system. Due to the high similarity of the data types of adjacent columns, the use of this storage mode can improve the compression rate of the data, and can further reduce the cost of data transmission after compression.

Big data online analysis technology. Online analytical processing technology is a key content in the big data warehouse system. The complex data analysis process focuses on decision-making analysis to provide users with actual results. Adopt online analysis means, starting from comprehensive data analysis, establish a multi-dimensional model, get comprehensive data analysis results, and provide reference for decision makers. One of the characteristics of online analytical processing is data analysis. Combining data warehouse with online analytical technology can not only calculate massive data, but also analyze data.

Big data mining technology adopts online analysis technology, and can often only obtain surface knowledge information, but little is known about the potential connection of data information. However, in the context of cloud computing, using data mining technology can learn about the data itself, and understand the potential connections between the data, using concepts, laws or patterns to express it. At present, the big data mining technology is mainly parallel mode, which has great advantages in large-scale data processing. In the past, serial data mining processed data with a small scale and took a long time, but distributed parallel data mining technology, distributed systems, clustering, splitting and other methods were used to improve the efficiency of data calculation. In addition, data mining technology in the context of cloud computing can take advantage of parallel mining. Compared with other serial methods, parallel mining technology can split parallel tasks in distributed systems through machine clusters, split the tasks and then process them. More machines

perform sub-task processing to improve data processing efficiency, which can save data processing costs to a certain extent.

Big data visualization technology. Through data mining technology, it can realize multi-dimensional and deep analysis of big data, which is convenient for obtaining more effective information. On the basis of the cloud computing platform, visualization technology can be realized, and the above information is embodied to make it more vividly displayed, and the relationship between data information is more intuitively displayed to the user, which is convenient for the user to understand. Visualization technology refers to displaying the database and related data in the database in the form of graphic images in the data storage space, and at the same time, using some means in the display process to dig out the relevant hidden information contained in the images. The traditional data processing process is only based on the data itself, to observe and analyze the information contained in the data. However, with the help of cloud computing visualization technology, not only can multi-dimensional images of non-spatial data be displayed, but also a direct retrieval process can be realized during the graphic display process, helping users to better mine data information, understand data information, and improve information retrieval effectiveness.

5. Conclusion

Today, with the continuous development of information technology, people have entered the era of big data. For the processing and analysis of massive big data, the traditional data processing technology has obviously been in decline. The emergence of cloud computing has just solved this problem and provided the possibility for big data processing. Through big data collection, storage, analysis and mining, it can realize the reasonable use of big data and provide customers with higher quality data services.

Acknowledgment

Research on effective classroom teaching process design based on hybrid teaching mode (JGYB1919) for the school level teaching reform project of Xijing university.

References

- [1] Ren Lei. Research on big data architecture based on cloud computing in power industry [J]. Technology and Enterprise, vol.9, no.3, pp. 94, 2015.
- [2] Gong Xu. Discussion on big data processing technology based on cloud computing [J]. Electronic Technology and Software Engineering, vol.11, no.6, pp. 198, 2015.
- [3] Zhang Lan. Research on the cache processing of e-commerce data based on cloud computing [J]. Computer Knowledge and Technology, vol.11, no.2, pp. 87-88, 2016.
- [4] Ren Liang. On Big Data Processing System under Cloud Computing Technology [J]. Electronic Technology and Software Engineering, vol.7, no.3, pp. 27, 2014.
- [5] Guo Qun. Research on the application of big data in the medical field based on cloud computing [J]. Fujian Computer, vol.8, no.4, pp. 131-132, 2016.