

Optimization and Use of Cloud Computing in Big Data Science

Ying Lin

Northern Arizona University, The City of Flagstaff, America

Keywords: Cloud Computing; Big Data Science; Applications and Challenges

Abstract: Cloud computing is widely used in big data science and has become the key to achieve data collection, data processing and application in various industries. Therefore, understanding cloud computing and big data science helps to further explore the core mechanism of cloud computing development. In addition, in the face of the exponential growth of data and cloud computing, the development of big data science also faces many challenges and opportunities. Based on the existing experience, the author will further understand the characteristics of cloud computing and the current status of the application of cloud computing in big data science, so as to clarify the challenges faced by cloud computing in big data science and put forward suggestions for improvement, thus promoting the active application of cloud computing in big data science.

1. Background of the study

The development and construction of the Internet has promoted the rapid development of information technology, and on this basis, the scale and performance of networks and computer systems have also been greatly improved. In this regard, the development of data science has also put forward higher requirements in terms of professionalism. On the one hand, the amount of manual data analysis and processing has increased dramatically. On the other hand, data storage and analysis are also facing stricter requirements. Accompanied by the explosive growth in the rate of data generation will be the application of cloud computing to big data science has become an inevitable trend in the development of data processing. Not only that, through the generalization of relevant data types and sources can be found. The data covers a wide range of industries, including social media, e-commerce, healthcare, transportation, and so on. The level, therefore, need more powerful computing power and efficient analysis ability. The data is analyzed in a comprehensive and professional manner. Cloud computing and big data science are more efficient and flexible than traditional data processing methods. Cloud computing technology itself has many advantages such as flexibility, efficiency and scalability. In addition, it can provide large-scale computing and storage resources, which can be dynamically allocated according to the actual demand. Therefore, the integration of cloud computing and big data science can build a more ideal scale for the development of data and the construction of the platform In this regard, the integration of cloud computing and big data science can provide more powerful computing and storage resources for the development of data, so as to efficiently process and analyze the data and improve the overall efficiency. Not only that, the use of cloud computing can reduce the original data maintenance costs

by sharing resources. Cloud computing can effectively reduce the maintenance and management costs of hardware equipment and facilities. Through cloud computing for data processing and analysis, and its application to the discipline of big data, can accelerate the scientific research process, faster to achieve the stage of research results. Thus, the flexibility and scalability of cloud computing can encourage more data innovation and promote the overall development and progress of science and technology.

2. Current status of cloud computing in big data science

Cloud computing is widely used in big data science. It has gradually entered the mature stage from the primary stage^[1]. And it is getting wider and wider in terms of application scope. First of all, cloud computing is used more in the field of scientific research. Cloud computing can be used for earthquake monitoring, marine information monitoring, astronomical information. Computational processing, etc., can timely understand the geographic environment and climate change. Therefore, it is often used in the field of more precise calculation^[2]. Secondly, cloud computing and big data science integration is also commonly used in network security protection^[3]. For example, it can be used for virus database storage spam shielding. At the same time, in the multimedia image and audio processing, cloud computing can be used to store and analyze animation materials, high simulation animation production, massive image retrieval and a series of. In addition, in the field of Internet, cloud computing can provide the corresponding mail service, online real-time translation, network retrieval services^[4], etc. In the field of medicine, cloud computing can also be used as DNA information analysis, massive case storage analysis. In the medical field, cloud computing can also be used as an important technical support for DNA information analysis, massive case storage and analysis, and medical image processing^[5]. With the continuous deepening of cloud computing technology, the application of cloud computing to big data science has gradually become a daily routine. Big data and cloud computing have become an important part of daily life, and they can bring more convenience to human life^[6]. Therefore, the combination of cloud computing technology, its own strong reliability, serviceability, cost-effective and so on. Characteristics of its application to big data science has become an inevitable trend in the development of global information technology, so as to achieve data visualization and efficient service. It can bring great impetus to the development of various industries.

3. Challenges of Cloud Computing in Big Data Science

3.1 Uneven quality of data

With the development of The Times, the gap between the explosive growth of the amount of data and the existing technical means is getting wider and wider^[7]. The amount of data collected by cloud computing is increasing exponentially, which also brings more difficulties for data screening, storage and processing. First of all, the sources of big data are diverse. In the process of data collection, the existing data needs to be accurately divided and marked, so that the structured data, semi-structured data and unstructured data can be sorted out and calculated in the statistical analysis of the data. Not only that, the type of data format is also facing more stringent requirements, so in the process of data analysis, cloud computing must also combine different types of data for analysis. By processing different formats and types of data to do a good job of system and algorithm identification work. There is still a large gap between this and the current development of cloud computing, big data science construction and application. Therefore, there is a need for higher. Scalability and performance of the cloud computing system for data processing. Secondly, due to the large amount of data collected, the quality of the data will be uneven, so in the process of data

collection and statistics, there will be data inaccuracy, incompleteness and inconsistency and so on. These will not only directly affect the accuracy of the overall analysis of data and the reliability of the final results, but also cause confusion in the process of data analysis, impact on the overall quality of data. Finally, the need for effective data resource management is mined from large amounts of data. Combined with professional knowledge and keen insight, therefore, powerful algorithms and tools are also needed in the process of cloud computing. Thus, it will be deep into the relevant field knowledge and business aspects of the corresponding management to enhance. However, in the current cloud computing and big data disciplines, it is not possible to fully ensure that in the massive amount of data can be accurately mined with the data related to this research or demand there will still be algorithmic analysis of the loopholes^[8].

3.2 Low data security

The collection and processing of data requires attention to both the accuracy and source of the data, while managing the security of the data and customer privacy is also the current top priority for good data processing. The information collected by big data will contain a variety of information types. For example, personal identity information and related business secrets, so how to do a good job through the platform between cloud computing and data processing. It is a key issue that currently affects data privacy and security, which is also one of the key issues in the current statistical analysis of data processing^[9]. Secondly, then the data security storage and transmission process also needs to take into account the actual algorithm of cloud computing and the overall scale, to ensure that the overall collection of information, statistics and grooming can be compliant, but also in the management of rights and applications to play a good articulation of the role of the current cloud computing resources can not fully satisfy the data in the demand for complex scientific calculations, therefore, the software and hardware facilities are improved accordingly through multiple calculations. The more intensive tasks need to do a good job of computing and large-scale data processing. Therefore, remote network access to data is required. Thus, in the process of statistical analysis and processing will also be subject to geographical and spatial limitations, commonly bandwidth limitations. If the bandwidth can not meet the actual needs of cloud computing, it will cause the overall calculation can not be carried out smoothly^[10].

3.3 Lagging data information

Cloud computing in the application of statistical analysis of big data, not only need to have a strong performance of its processing and analysis, but also to pay attention to the security of the data. In addition to this, how to fast real-time processing of the collected information? Identify and analyze, and promote the dissemination of overall information is also one of the important challenges currently facing the application of cloud computing in big data science. First of all, the large amount of data as well as the faster growth of the dissemination range will require data. Using cloud computing for data processing, it is important to screen the data with sufficient professionalism as well as to do a good job of transforming the data in a timely manner, so as to realize the distribution of multiple ports in terms of images, audio, text, and so on. However, in the current process of cloud computing processing, data information processing requires in addition to sufficiently powerful hardware equipment. Many cloud computing processors are not able to completely process the information at the fastest speed, so it will present information processing lag. Secondly, in the process of information processing and dissemination, differences in geography and equipment can lead to differences in methods of use and maintenance. The management of information screening will also be due to a variety of issues such as layer by layer auditing, resulting in information lag. Thus, it is impossible to show the processed information in real time.

In summary, in the process of cloud computing and data processing, it is not possible to do a good job of real-time processing and construction of information, so in the current process of cloud computing and big data science construction, there is still a lag in information management and other situations.

4. Optimization Measures of Cloud Computing in Big Data Science Applications

4.1 Optimization of screening pathways

The optimization path of cloud computing in the application of big data science needs to be combined with the current situation presented by cloud computing to improve and enhance the development of cloud computing and big data science data quality issues, can be optimized by optimizing the screening path to improve the overall data quality. First of all, conducting cloud computing to process data requires strengthening data collection and pre-processing to effectively improve data quality while ensuring data quality. Cloud computing can provide more powerful computing power and storage capacity to support efficient cleaning and preprocessing of large-scale big data, so cloud computing can carry out data cleaning and preprocessing by eliminating duplicated data, correcting erroneous data, and filling in missing values. Therefore, it can remove duplicate data, correct erroneous data, fill in missing values and so on. Improving the overall accuracy and completeness of the data requires professionals to accurately screen the data in the pre-processing stage. Through keyword setting, the labels and images of the data are accurately divided to ensure high quality and completeness of the data. Secondly, to do a good job of data storage and management, cloud computing can provide distributed storage and centralized management. Therefore, it can fundamentally ensure the reliability and security of big data. Distributed storage data will be dispersed to multiple nodes. It can improve the overall reliability and scalability of the data, and at the same time, centralized processing can effectively achieve unified management and maintenance of the data. It can also effectively reduce data inconsistency and redundant data. Third, the analysis and mining of data can further promote the data to visualize the interactive state in the cloud computing platform to show the overall high-quality development of data. Cloud computing can provide efficient data analysis and mining tools, through distributed computing and parallel processing technology can ensure that the cloud computing platform can do a good job of large-scale data analysis and mining quickly and accurately, so the choice of appropriate cloud computing resources and platforms can effectively deal with massive amounts of data at the same time, to find the development of data in the laws and patterns can be a good way to do a good job in the future of the data filtering, organizing and summarizing. Provide scientific research and decision-making support. In this regard, the choice of a suitable cloud computing platform. It is crucial to improve the quality of big data. As we all know, different cloud computing platforms have different advantages and features, so you need to make a specific choice according to the actual data application scenarios and needs. If you need to classify cloud computing, you can do so by choosing a platform that focuses on big data processing and analysis. To carry out the use of cloud computing, it is necessary to ensure that the platform has a strong computing and storage capacity, so as to introduce a wealth of data services and solutions to the overall data management through the cloud computing platform, which facilitates the user's application of data screening. In this regard, cloud computing through the provision of data development visualization and interactive tools to promote the use. It can help users understand and apply big data visualization technology in a better way, and can present complex data to users in a more intuitive form. Convenient for users to analyze and mine data, while the interactive design can promote the formation of good interaction between users and data. So as to do a good job of personalized analysis and exploration. In this regard, the above choices can ensure that cloud computing in big

data science in a more accurate form, through distributed storage, centralized management. Efficient filtering and pre-processing, to do a good job of high-quality analysis and mining of data. To visualize the state and interaction and other forms of high-quality data development, but also for the subsequent integration of cloud computing and big data disciplines to provide high-quality screening pathway, laying a good foundation.

4.2 Enhancement of equipment safety

Cloud computing in big data science applications for good management of data and equipment security. It plays an important role in the reserve of data. In the era of big data, the rapid growth in the amount of data and the diverse types of data bring great challenges to the collection, processing, and analysis of data. When dealing with data, it is important to consider both the source, type, and application scenarios of the data, as well as the ability and scope of cloud computing to deal with complex data. Cloud computing as an emerging technology model can provide a more efficient, flexible, and reliable way of dealing with many applications, and therefore can do a good job of improving the performance of data security through cloud computing. Therefore, it can effectively avoid the risk of data loss and leakage caused by the loss or damage of physical equipment. At the same time, according to the cloud computing data encryption and access control mechanism, the development is centered on protecting data security to ensure data management security and user privacy security. Secondly, cloud computing in enhancing data processing capabilities, according to the cloud computing environment and the development of the data itself to do a good job in the dynamic expansion or increase or decrease of computing resources and storage resources, so as to meet the actual needs of the data in different application scenarios. In this regard, in the process of processing large-scale data, it is necessary to. Strengthen the impact of the scene on the data, so as to choose the appropriate. Calculation methods and processing modes to achieve efficient and rapid processing and analysis of data. In this regard, the multi-tenant model provided by cloud computing can further screen compliant and secure platforms, allowing multiple users to share computing and storage resources. The use of cloud computing can effectively do a good job of securing equipment control while reducing overall costs and improving resource utilization. Finally cloud computing can be used as a data sharing and cooperation. The key bridge to connect data and users together, so in terms of equipment management needs to strengthen regular maintenance and network security settings, thus ensuring that data in the storage process can be more easily shared in a way to show to the organization and the user, to achieve a wider range of data cooperation and collaborative work, while also reducing the risk of exposure of information to achieve the interoperability of data in different areas and sharing, and to promote overall data development and cloud computing research and construction. In this regard, cloud computing in the application of big data science can be based on the actual needs of equipment security and data application scenarios. Do a good job of screening data, enhance data processing capabilities, improve data security, and promote data sharing and cooperation, so as to provide more support for future cloud computing and data development.

4.3 Enhance information flow to improve processing speed

The application of cloud computing in big data science can effectively improve the processing speed and quality of information. In the face of ever-changing technological changes, good data collection and organization of statistical analysis is also the current good. Data and information management is the key to using cloud computing, so it is necessary to strengthen the management of information flow, improve the speed of data processing, and effectively enhance the application of cloud computing in big data science. For example, by enhancing the processing speed of

information flow, the overall efficiency and accuracy of analyzing data can be improved. Through large-scale data processing and analysis tasks can be distributed to multiple computing nodes. At the same time, cloud computing is also able to take advantage of existing information processing to achieve faster data processing speed and efficient analysis. First of all, between cloud computing and big data science can improve the overall processing speed through powerful computing capabilities, the traditional data processing model will be limited by hardware resources. Since traditional models cannot handle large-scale datasets, the scaling of data storage and the distribution of computational tasks need to be realized through cloud computing. It can be placed on multiple servers to achieve more powerful computing capabilities, and can effectively handle larger-scale data. Thus, it can also provide more possibilities for the future to do a good job with a large number of data processing. Do in-depth analysis and research of data through cloud preprocessing. Second, cloud computing can achieve higher and better data processing speed. The optimization of information flow is used as a prerequisite for data processing, which can distribute data to multiple computing nodes. And realize the speed of data processing based on the existing computational advantages. On the one hand, it is possible to obtain faster analysis results and understand the phenomenon presented by the data, and at the same time, it can also be used as an important part of decision-making, and make overall decisions based on the data analysis decisively. Third, through the optimization of data and information flow can improve the overall flexibility and scalability, big data science can be adjusted by adjusting the scale of computing resources to meet the corresponding data processing and analysis tasks when the overall cloud computing and big data science integration. Do a good job of disaster data field aspects of management. Therefore can effectively reduce the cost of data processing and analysis through big data science and cloud computing. Overall data development services can be enhanced through the use of cloud computing.

5. Summary

The application of cloud computing in big data science can effectively improve the data processing performance and expand its scope and depth, and it can also further integrate cloud computing and big data science to realize the diversified processing between data, and realize the diversified fusion of data and precision management. The input and use of cloud computing can provide an effective management model for expanding data construction.

References

- [1] Shang Duxin; Wang Tao; Zhang Tianyu; Tang Jinjian; Zhang Zhihou; Zheng Xijian. *Application of big data technology in campus management [J]. Science and Technology Innovation Herald*, 2019(32)
- [2] Jiali Jin. *Research on the development of information management system in the context of big data[J]. Information System Engineering*, 2019(10)
- [3] Yang Siyu. *Discussion on the professional construction of information management and information system under the background of big data [J]. Science and Technology Wind*, 2020(12)
- [4] Ma Dongbo. *Application and Prospect of Big Data Technology in Smart Campus[J]. Industry and Technology Forum*, 2020(23)
- [5] Gu Jiale; Xuan Junru. *Design of student portrait system integrating big data technology and deep learning[J]. Information and Computer (Theoretical Edition)*, 2022(18)
- [6] Jiang Minxu; Wang Yuxin. *Research on the construction and application of smart agriculture based on big data technology [J]. Modernized Agriculture*, 2022(04)
- [7] Liang Lindong; He Zhengdong; Wang Fenshou. *Brief analysis of the utilization strategy of big data technology in oilfield development [J]. China Petroleum and Chemical Standards and Quality*, 2022(15)
- [8] Xu Yanping. *Trying to talk about the application of big data technology in medical informatization [J]. Information Record Material*, 2020(12)
- [9] Xu Xiangming. *Application of big data technology in school management[J]. Information Record Material*, 2020(12)
- [10] Li Jian. *Application analysis of big data technology in civil engineering[J]. Information Record Material*, 2021(03)