The Application of Computer Information Processing Technology in the "Big Data" Era

Zhihui Liu[1, a]

[1] Xi'an International University, Xi'an Shaanxi, 710077

[a] email

Keywords: Computer Information Processing Technology; The Era of Big Data, Data Processing Model

Abstract. The next era of big data, computer information processing technology for processing large data stream processing model and the model is divided into two types of batch mode, computer information processing large data including extraction, analysis and interpretation of data in the three big work. Facebook and data processing, for example, analyzing the application of computer information processing technology, challenges summarize large data computer information processing technology face.

Introduction

Large data era, hundreds of millions of computers and mobile devices to create amazing amount of information. Data include not only human information, but also other things related information. Moreover, the amount of information continues to increase, never stagnant. The amount of data large data era surge in computer information processing technology capability also need to continuously upgrade. US computer giant IBM believes that the era of big data processing computer information technology requires the ability to handle huge amounts of data. The conventional computer information processing technology to measure the amount of data (megabytes, MB) can not meet the demand [1]. Therefore, this article on computer information processing technology under the background of big data analysis.

Big Data Processing Method

Computer processing of data into the process both serial and parallel mode, which is also divided into stream data processing model and batch processing both models.

It means a direct stream processing model information data processing technique, the process model information storage process is omitted. Refers to the stream processing model data as water, the continuous data to be processed composition obtained data flow architecture [2]. After obtaining the data, the model immediately processed data and real-time feedback results. Stream processing model is mainly used to handle high speed requirements of response data. Should the information stored after processing, data processing result of the loss of effectiveness, the data results are not applicable to the environment, and therefore do not have the actual data processing value. Industries such as securities, stocks and other such industries as real-time data processing system data exchange system, the data processing speed have higher requirements, and therefore more suitable for stream processing model [2]. Since the stream processing model for the effectiveness of the data with high requirements, data need not be stored before processing, which are usually larger data capacity, and adversely affect the processing speed of the stream processing model, resulting in reduced efficiency. Thus before design flow processing model, we need to combine the characteristics of data to be processed, properly designed data structure to ensure the efficiency of data processing.

Batch model required the data into blocks, and then through different treatment zone data from different modules for processing. As it can be seen from the batch processing model, the basic principles of the process model for the data division processing, and then were treated by different regions, so that the batch model can reduce space and time consumption data.

The effect of the two treatments on the large data model is better, and have their own advantages. However the practical application of both models to handle large data, the data type is extremely complex, and therefore need to use two models to process data, rather than individual using one of the models. Big data such as e-commerce make, requires root data timeliness requirements choose different processing model. For the processing speed in seconds or milliseconds of data, we need to use stream processing model processing; and for the period in days or weeks of the issue, can be used in batch mode for data processing. In addition, in minutes or hours cycle data can stream processing model or batch mode processing.

Big data processing flow roughly divided into three steps. First, the data source information extraction; secondly, according to the characteristics of the data to choose the right way to store data; Finally, according to the data processing speed and space requirements, select the appropriate processing mode data processing, the final results will feedback to the user. User refers not only to people, but also indicates call Big Data processing program or task. According to the three steps of data processing, computer information processing large data including extraction, analysis and interpretation of data in the three big work.

First, data extraction. Due to the different large data sources, namely a large organization with a complex variety of data with the characteristics of data to be processed there may be valid data and invalid data, thus model requires data processing to extract valid data from the mass of data that describes the entity as well as an analysis object relationships between entities, and then the relationship between the entity and its data is processed to form a unified organizational structure for storage, to give the intermediate data can be further processed. Current source for data extraction and unified technology is more mature, you can refer to and draw on existing prior art extraction and integration of intermediate data.

Great work on the core data processing for the data analysis, which is the intermediate data processing and analysis. Traditional middle data processing technology is more mature, machine learning, data mining techniques are able to deal effectively with the intermediate data, the above techniques can be used for the intermediate data of a large data processing and analysis. But big data technology in the middle of the data is large, and therefore can not be used directly in the middle of traditional data processing, would need to be optimized and improved according to the characteristics of Big Data, Big Data eliminate the useless and invalid data before use. Therefore, the intermediate data analysis and processing, the first link is to clean up and filter the data.

Second, analysis of the data. Since the object of computer information technology for big data processing, data processing speed and thus a higher status, but also cause for concern. In addition, the characteristics of big data also determines the algorithm design of computer information technology data processing needs to coordinate the handling of the relationship between accuracy and speed. In addition, the processing of different objects is substantially the same, the algorithm also need to consider its portability algorithm [3]. And when the data reaches a certain size, does not have universal between different amounts of data processing methods, only small data processing method may not work for big data processing.

Third, data interpretation. Data interpretation belong results show, in general, the data processing results were highly professional features usually only professionals understand the meaning of the data processing results. For the convenience of non-professionals to understand the results of data processing, the results need to make a reasonable explanation. Traditional data interpretation method for the terminal display or store the text, but the huge amount of data big data and complex contact between the entities and more complex, and thus the terminal display and article storage visualization technology can not achieve an effective interpretation of results, and thus the actual tag cloud can be used in place of the historical flow, etc.

The Computer Information Technology and Data Processing Tools

Specific computers cloud computing concept cloud computing integration of distributed computing, parallel computing and grid computing, and many other concepts, and on the basis of the above calculation on the concept of change accordingly. Cloud computing "cloud" is a virtual container to

cloud cover hardware, system development platform, the input and output services and other available resources. Cloud computing as a concept broader calculation, its implementation also needs more support conditions, such as the various concepts of computer operating systems, storage, processing, indexing techniques [4].

The best way to handle legacy data to a relational database schema, but ineffective relational database schema is used to manage and store large data, and thus does not apply to large data processing. Current large data processing in the field most of the more common processing tools for the Hadoop, Hadoop data processing technology integrates several large data processing technology required, Hadoop Big Data processing tools can be used as a benchmark developed. For the current development of large data processing tools, innovation, mainly based on the characteristics of data objects for Hadoop effectively improved. In addition to Hadoop, there are many large data processing tools. But other big data processing functions and Hadoop tool similar to a limited part of the large range of data processing tools, mainly domain-specific data processing [5].

Application of Computer Information Processing Technology - A Case Study of Facebook

Facebook is currently the largest global social network platform, currently Facebook's more than one billion users worldwide, a large number of users can also cause a lot of network platform to produce the amount of data per day, the amount of data available over 25PB, and the amount of data to 60TB per day rate increase. Facebook problem to be solved very massive data processing problems. Figure 1 is a Facebook data processing platform framework.

In facebookda data processing platform, the role of relational database technology is to obtain data, provide data source for large data processing platform. At the same time the relationship between library technology in the backend role in the OECD data, users can query data and query results are returned to the user terminal display platform. Core platform for data processing and analysis, the part Hivc-Hadoop cluster implementation. After the preceding relation database technology to obtain data stored in the corresponding back-end relational database, and the system can query Hivc Hivc-Hadoop cluster, its role is to reduce the ad hoc query Hivc system pressure. Copy the data to the backup Hivc system, backup system responsible for ad hoc queries.

Challenges and Opportunities of Computer Information Processing Technology

Big data to the users a richer experience, but also the advantages of computer technology and the Internet have become more prominent. But also to the large data information processing technology has brought new problems, but also to computer information processing technology to produce a new challenge.

Under the era of big data, the object information processing technology faces for mass distribution of a very wide range of heterogeneous data, pre-processing of heterogeneous data need to integrate heterogeneous data. Compared with conventional data processing objects, data types and large data era of unstructured data, but a variety of complex structured, semi-structured and unstructured data. And with the popularity of mobile devices increasingly wide range of spatial and temporal data generated are constantly changing, resulting in data storage also to convert storage pattern, in order to meet the massive data storage requirements.

The larger the amount of data, the higher the complexity of the data, the higher the probability of attacks. From the current computer virus and network attack situation, big data has become a major new target viruses and network attacks. Therefore, data security issues become big key influence for the trust, and the impact on the development prospects of big data. The larger the amount of data, the higher the possibility of data loss. Conventional data processing in the stream processing model to store large data processing needs of centralized data storage, data leakage risk is greatly increased.

In addition, the era of big data, user network behavior will be recorded in the form of specific data, the user's personal privacy but also face greater security risks. Enhanced interaction between users, data loss at some point, although not a threat, but a plurality of time information disclosure

may be linked to the user's behavior, causing the user privacy is compromised [6].

Mass data processing performance computers have high requirements for processing huge amounts of data, the computer performance handle high-speed operation state, the power consumption of the computer is also higher. Thus, by reducing the energy consumption of hardware, software upgrade functions of the computer information processing technology brings restrictions.

Conclusion

In summary, under the background of big data, computer information processing technology and improvements need to be made to optimize the characteristics of the data, computer information processing in the era of big data to calculate better Yingying and development. While computer information processing technology there are many challenges, but the computer information processing technology still has great potential to bring the era of big data challenge also contains a lot of opportunities, I believe that the computer information processing technology under the "big data" in the background of the future there will be greater development.

References

[1] Lu Jingquan. "big data" background Computer Information Processing Technology [J]. Information and Computer (Theory Edition), 2013,06: 126-127.

[2] Tian Maolin. computer information technology research "big data" era of treatment [J]. wireless Internet technology, 2016,02: 144-146.

[3] Yang Zong. computer information processing technology "big data" Time Analysis [J]. Information and Computer (Theory Edition), 2015,12: 133-134.

[4] Yang Shangyong. Computer information processing technology in the "big data" era of application [J]. Suppliers, 2015,40: 215.

[5] Li Chunhui. Computer information processing technology under the "big data" background analysis [J]. Electronics and Software Engineering, 2014, 09: 33-34.

[6] Wang Haiyan. Application and Practice of Computer Information Processing Technology under the background of Big Data [J] Information and Computer (Theory Edition), 2016, 03: 22 + 24.