Comparative analysis of data management system

YIN Cheng du[1,2,a] JIA Hongli[1] Gaolili[1] Linxin[2]

(1.Department of Equipment Command and Administration, Ordnance Engineering College, Shijiazhuang 050003.China

2.Troop No.77606 of PLA Lhasa 850000.China

a.ycd1124@163.com)

Keywords: traditional data management system; big data management system; thinking method；data processing；data storage

Abstract: Based on the traditional data management system and big data management system was introduced, the comparative analysis of the traditional data management system and big data management system from the aspects of thinking method, data storage technology and data processing technology. Make people realize that with the development of the information age, the advantage of big data management system is better than the traditional data management system.

With the development of information-based society, networking and intelligent and the popularity of Internet, networking and intelligent terminal, people's lives gradually entered by the knowledge explosion times to the explosion of data, data has become today's new production factors and strategic resources, much of the industry and relevant government departments to focus on [1]. Data management system is the most direct data processing tools, but with the increase in the amount of data, the type of complex, the data management put forward higher requirements. Many of the industry in the face of massive, multi-source heterogeneous data, the traditional data processing system is facing storage and analysis processing bottleneck, and data management system appeared to solve the many technical problems in the traditional data management system. The traditional data management system and data system is to deal with the analysis of data in valuable information and knowledge for the purpose, but technology development background, environmental data, analysis processing depth, breadth and the thinking mode of[2], the traditional data management system and data system there is a big difference.

1 . Traditional data management systems and large data management system

The traditional data management system and data management system there is a big difference, first of all on the definition are compared, the data management system can be seen as the upgrading of the traditional data management system, a higher data processing.

1.1 traditional data management system

The traditional data management system is refers to from a large number of real data through a specific algorithm analysis of each data, within the acceptable computational resources to find the law, the automatic extraction of implicit, past the unknown, valuable potential information and analysis is the real data may is incomplete, noisy, fuzzy and random [3].

1.2 data management system

Data management system can be understood as many complex data together analysis and processing, inside the potential value, knowledge is extracted, make people realize that things in the global, and can predict the future trends of the system. Big data system for today's data stream processing has its unique advantages, unapt make people lost in the information society of the massive complex data in the tide.

The traditional information management system with large data platform from three aspects:

（1）the traditional information management system is generally used for a realistic generated data, big data management is based on the analysis of existing data.

（2）the traditional data management system is a linear analysis, big data analysis processing system using parallel logic.

（3）the traditional information management system is focus on simple data applications, large data management system for global analysis and application of [4].

2 traditional data management systems and big data management system of thinking difference

The big data era, opened a new era of transformation, and has brought about great changes in people's work and life thinking. The era of big data thinking embodied in three aspects.

(1) the transition from causal relationship to. Change the idea to solve the problem, big data concerned about the relationship between data, rather than tangled in difficult to figure out the cause and effect relationship.

(2) from the exact transition to hybrid. In the face of the rapid growth of multi-source heterogeneous data, the analysis of the data is constantly expanding, ignoring the micro level of precision, chasing the macro level of the law of the development of things and direction.

(3) analysis of transition from random samples to all data. When analyzing the data, not close to a small part of the data, to make use of all the data, and to make full use of [5].

The traditional data management thinking that data to be stationary, once after processing and analysis, the data would become old and useless data, and data systems that data is in motion, not static nor divided into the old and the new and pay more attention to the cross correlation comparison of old and new data.

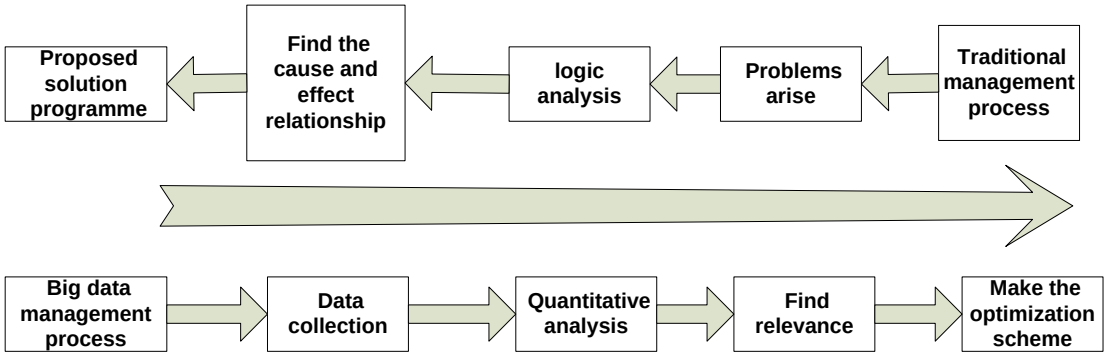

Figure 1 traditional data management and big data management thinking difference

The traditional method of small data is found, by logical analysis, the causal relationship is proposed to solve the problem. The method of big data is through large-scale data collection, through quantitative analysis, put forward the associated relationship, make the optimization scheme. Big data management and the traditional management mode of thinking difference as shown in figure 1. Different thinking patterns are mainly embodied in the causation of liberation from the people to a great extent, and more focus on the use of the associated relationship and discovery. Big data thinking is mainly to tell us what is rather than why, do not have to know the reasons behind the phenomenon.

Sampling has been hailed as one of the greatest achievements of the twentieth Century[6], its advantages are through the sample, the survey, access to data, and then on the overall inference. Sampling deliberately avoided the detail investigation. However, in many areas, part of the data collection has as much as possible to the data collection transformation, sampling has been unable to achieve the depth of data and analysis of wide crossing with the use of. Random sampling is a temporary data, as you collect more and more data, the prediction results will be more accurate. Thus our data processing has occurred earth shaking change, and our thinking has not changed. Therefore, we should abandon sample analysis this shortcut, complete and comprehensive data collection.

3 Comparison of storage and processing technologies

Data management technology has experienced the stage of manual management, file system and database system, with the wide application of data, the data environment of data management is faced with more and more complex, analysis and processing efficiency of timeliness requirements more and more high, and the traditional database and analysis processing technology exposes more and more problems. The emergence of large data management systems to data management technology has brought fresh blood.

3.1 traditional information management system and data storage method of large data management system

Traditional data storage is usually the storage structure of the database. Records are usually in the form of existing database relational tables, the traditional relational database needs to be neat database tables, and high requirement of hierarchy and consistency, these are the traditional database query language (SQL) necessary, traditional database established in the past are focus on data stored in a large disk, its shortcoming is to take advantage of the physical host rate is low, hardware resources to expansion and maintenance costs higher. The large data non relational database (NoSQL) no longer has the hierarchy and consistency requirements of [7]. Large data processing requirements high speed of data loading, high speed of query processing, and the efficient use of storage space, for the aperiodic of the big data is the structured data, structured data, semi-structured data and data management system using distributed file system, distributed database system, data stream processing system of divide and rule of storage.

Traditional data management system and data systems due to the data environment and therefore deal with the object, there is a big difference, traditional data management system data source is mainly a specific range of information management system, to generate passive based and data structure of a single, in a structured data based. Data management system data comes from a variety of sources, in addition to the sensing equipment of the information management system and web information system, perception of information system to generate various types of data, compared to the traditional data sources are broader, number more, more complex structure, data in addition to passively collect more is active acquisition, the acquisition is a more comprehensive range, high throughput data, faster processing speed, the requirement of real-time rendering, due to the requirements of the data accuracy is not high, so high data redundancy, high uncertainty, storage, processing object the main difference see Table 2 shows.

Table 2 Comparison of traditional data management system and large data system

contrast	Traditional data management system	Big data management system
data source	less	Numerous and extensive
Collection scope	Local sampling	Global acquisition
Collection way	Passive dominated	Initiative and automation
data type	Relatively single, mainly to structured data	Complex and diverse, with a semi - structured and unstructured data based
Data redundancy	low	high
Total data volume	Calculated by TB	Massive data, calculated by ZB or EB
Processing unit	With MB as a unit	With PB as a unit
Data precision	High demand	demand moderately
Treatment efficiency	Longer time	Real time, fast

Due to the traditional relational database, in the management of large data encountered a large amount of data and data type of complex a series of difficulties and obstacles, based on distributed data storage is a hot issue today, is the effective way to solve the large data storage, so with the trend of dramatic increase in the amount of data types and data, large data storage technology will be the inevitable choice of the future of data storage.

3.2 traditional data management system and data processing methods of large data management system

Data mining, from the technical level is concerned with a large number of fuzzy, random, uncertain data, the extraction of knowledge and information related to the existence of the unknown potential use of information and information process, the process as shown in figure 3.

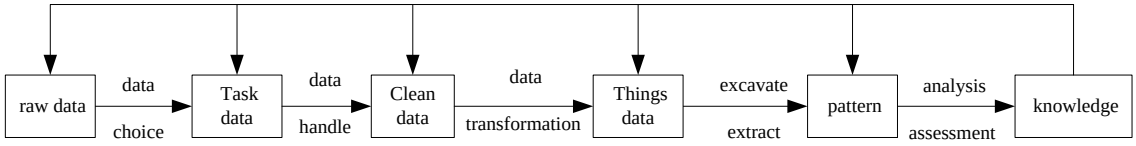

Figure 3 data mining process model

The traditional data analysis of statistics collected first-hand and second-hand data for processing and analysis. By analyzing and processing the knowledge hidden in the seemingly chaotic data, information extraction, extracted and find out the internal rules of the object of study, in order to achieve the maximization of the use of data. Its flow chart is shown in figure 4.

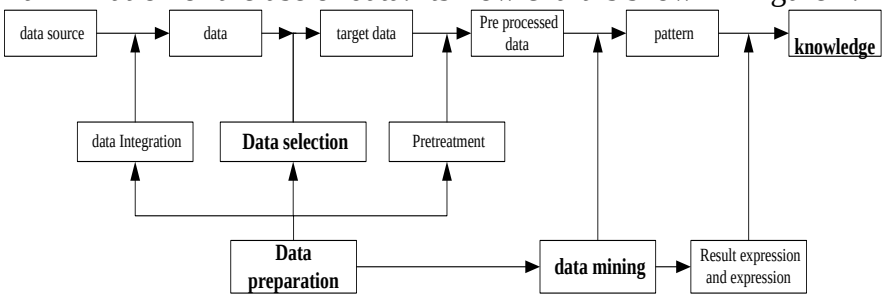

Fig. 4 the process of traditional data processing

Can be seen on a special data analysis of large data analysis, so a lot of traditional data processing methods can also be applied to large data analysis. Data mining refers to data from large, complex type, dynamic fast growth, low value density in mining has great potential value of information and knowledge, and services to users. Hadoop system as the most widely used data processing, mature and effective tool, with multiple advantages:

(1) can solve the bottleneck problem cannot be solved by traditional method of data sources and data form diversification and the amount of data in the rapid growth, especially HDFS high fault tolerance, high scalability, MapReduce programming model is of high data processing seemed at ease.

(2) secondly in data processing, low cost consumption, in hardware, because Hadoop is particularly suitable in the architecture running in cheap servers; in software, Hadoop is based on Apache's open source free program, strong controllability, modify, plasticity is strong.

(3) once again, the Hadoop system covers all aspects of the data processing process, and improve the ecological system, is more conducive to the compatibility between different products and high reliability.

(4) Big data processing technology in general is to make an accurate prediction of the future, and the current individual differences or the same reason, to make a correlation analysis. The task map analysis, carries on the induction and then reduce, using a divide and conquer, in order to win the processing of thinking. Finally, the knowledge representation, the data into the homes of ordinary people.

4 Concluding remarks

Now the development of data systems have been widely in-depth including finance, Internet, communications, energy, manufacturing, health, education, transportation, agriculture, health care, retail and other of our lives in various fields, which will for our lives produced great influence, and to the traditional number according to will bring a great challenge to the management and analysis. In this paper, the traditional data management system and the large data management system, the definition of thinking, storage technology and processing technology were compared and analyzed. To make people have a further understanding to the data management system is the practical demand of the era of big data explosion; modern new equipment; direct impetus to the modernization of the new; promote an inevitable choice for the transformation of the four modernizations.

Reference

[1] Liu Weiwei, Deng Zhonghua, Lu Yingjun, based on the research of the connotation and solution of large data mining based on cloud computing [J]. information system, 2015,38 (7): 103-108
[2] Liang feng lan Research on data management methods of beam bridge in the era of data management [J]. computer development and application, 2013,26 (11): 34-36
[3] Wang Yu Research on Key Technologies of large data processing flow based on cloud platform [J]. information technology, 2014 (9): 143-146
[4] Yang ju long big data technology complete solution foundation, design, development and practice [M]. Electronics Industry Publishing House, 2014
[5] Chen hai song Wang Xiaofei Zhang Daokun, Zhang Lin based on UML equipment integrated information management system analysis and modeling [J]. Journal of Sichuan ordnance, 2011,32 (2): 1-4
[6] Wang Xiangbo Research of equipment maintenance information management based on data mining in [D].2013,12
[7] Feng han chao, Zhou Kaidong, research on the optimization of large data storage structure under distributed system [J]. Journal of Hebei University of Engineering, 2014,31 (4): 69-73