Research on Data Sharing of Water Conservancy Informatization Based on Data Mining and Cloud Computing

Daqian Li*

Xi'an University of Technology, Xi'an, Shaanxi 710048, China

*Corresponding author: 2756116759@qq.com

Abstract. The wide application of networking technology has brought great convenience to people's daily life and promoted the development of international economy. The innovation of information technology has profoundly changed people's traditional way of life. In water conservancy, water conservancy informatization is the foundation and important symbol of water conservancy modernization. As an important tool of information construction, data warehouse and data mining technology have unique advantages in dealing with massive data. Therefore, it is widely used in water conservancy industry with massive data. Combined with the application characteristics of data mining and cloud computing, this paper analyzes several key supporting technologies of water conservancy informatization data sharing. Through the analysis of the existing system integration and application mode, the preliminary design of the operation framework is put forward in theory, which provides reference and suggestions for the construction of water conservancy information data sharing.

Keywords: Data mining; Cloud computing; Water conservancy informationization; data sharing

1. Introduction

Nowadays, the rapid development of information technology and Internet has brought great changes to people's lives. In the new environment, network applications related to social networks, e-commerce, mobile games and other concepts have been welcomed, and have been well integrated into life, changing people's lifestyles [1]. For many people, the Internet is almost as important as public infrastructure services such as water, electricity and gas.

Water conservancy informatization refers to the use of information technology to collect, process, store and share water conservancy information, so as to improve the efficiency of water conservancy work in an all-round way. Water conservancy management adopts the management mode of combining watershed and regional management, and water conservancy informatization construction is the main way to realize the integration of watershed information resources, realize the sharing of water resources basic information, geospatial information and water management information, and serve the whole social and economic development [2]. At present, driven by the development of Internet of Things, cloud computing, mobile Internet, etc., the digitalization level and degree of water conservancy industry have been continuously improved, and the accumulated data resources have...
increased in series, forming a large number of data sources, thus announcing the coming of the era of water conservancy big data.

Traditional data mining can't meet people's demand for information. After the exploration and research of experts and scholars, cloud computing has super computing power and storage capacity, and its operating cost is relatively low, which is favored by the society. Therefore, building a data mining platform based on cloud computing is an inevitable trend in the new era, which makes outstanding contributions to promoting the development of national science and technology [3]. Starting with the connotation of big data mining, this paper analyzes the data sharing countermeasures of water conservancy informatization based on data mining and cloud computing, so as to promote the better and faster development of big data and meet the needs and services of water conservancy informatization.

2. Big data mining and cloud computing technology

2.1. Big data mining

In fact, big data mining is to mine some valuable and potential information for customers from big data with rich types, dynamic updates, high density and low value, so as to meet customers' needs and serve customers. The real purpose of big data mining is to mine valuable information.

With the help of the Internet, cloud computing and mobile intelligent terminals, data mining and development can be carried out better. Nowadays, although the application maturity of big data mining is obviously superior to that of traditional data mining, the research on big data mining is still in the process of continuous improvement and perfection. Big data mining still needs the help of cloud computing and other related technologies to realize mature application.

The growth background of big data mining is quite different from that of traditional big data mining, so its processing objects are also different. Big data mining has a wide range. Besides the data processing of management information system and Web information system, it also needs to process and explore some sensing equipment information similar to perceptual information system. Big data mining fully embodies the characteristics of its wide range of data sources, large storage capacity and rich and diverse types [4]. Because of this, big data mining can collect information more freely, without limitation, with a wide range of data collection and fast processing time, but at the same time, the data accuracy of big data mining is not very high.

2.2. Cloud computing technology

As a new technology, cloud computing technology can not only provide support for massive data processing, but also effectively improve the data processing level. Usually, cloud computing technology itself has the following characteristics:

First, through the application of cloud computing technology, users can be provided with the required self-service.

Second, with the help of cloud computing technology, access to cloud computing network system can be realized anytime and anywhere with various network facilities, and the cloud computing technology can be used to build a resource pool shared by many people [5].

Third, under the action of cloud computing, it can be deployed quickly and its application functions can be brought into full play [6]. Through the research, it is known that cloud computing technology can provide virtual resources support for both parties of online transactions. In the process of building data mining platform, by applying cloud computing technology, it can give full play to its due application value.

3. Problems faced by water conservancy information sharing

Information storage is the core of water conservancy informatization construction. The application scope of water conservancy informatization construction in different places and departments is narrow, and there is no unified storage technology standard [7], which leads to the disconnection of
the whole country. DBMS database is widely used in water conservancy system, but its development is limited by its spatial information expansion ability, and there are many differences in hydrological data storage between different departments, which often makes it difficult to achieve seamless connection. Resource integration is difficult. Due to the lack of integration and coordination concept in the early stage of construction, the coordination ability between various departments and various types of data is poor [8]. Although the investment in hardware has been widely valued in the process of water conservancy informatization, it is an urgent problem for water conservancy informatization management departments to fully integrate various hardware resources and improve resource utilization rate.

The management of early water conservancy informatization construction is scattered, and the integration of debris of built and not built information data centers is a primary task of current water conservancy informatization construction [9]. Backward infrastructure, mainly reflected in the backward data collection methods and low efficiency, affects the quantity and accuracy of water conservancy information. With the advancement of water conservancy informatization, more and more water conservancy information systems have been established or will be established soon, and there are more and more potential network security problems [10]. Improving the security and defense capability of water conservancy information system has become an unavoidable problem for water conservancy departments.

4. Application of data mining and cloud computing technology in water conservancy informatization data sharing

4.1. Data warehouse construction

The establishment of data warehouse is for us to process the hydrological data, and more importantly, to complete the quick and accurate query, drawing, statistics, analysis and printing of hydrological data, and to provide effective reference for managers to make decisions [11]. In response to this requirement, the construction of data warehouse for hydrological data information mainly includes the following steps:

The first step is to collect and analyze business requirements. Understand the latest hydrological information, so that water conservancy institutions at all levels and flood control and drought prevention departments can master hydrological information and forecast.

The second step, physical design of data model and data warehouse. According to the characteristics of water conservancy information resources, the corresponding tables of rainfall (including time period, daily and monthly rainfall), rivers, dams, reservoirs, entropy and pumping stations, as well as the corresponding tables of rainstorm report, hail situation, wind power, gate opening and closing, reservoir opening, etc.

The third step is to extract, transform, purify and load the data into the data warehouse [12]. In lakes and rivers, we get the corresponding data through telemetry devices, and specially write the relevant hydrological information translation system, which translates the data obtained by telemetry devices, processes the error messages, and finally loads them into the data warehouse.

The fourth step, select some connection software to provide necessary data for other applications, analyze the existing data online, and update the data warehouse continuously during the use.

4.2. Construction of integrated management information system

For the purpose of collaborative management and efficient application, the integrated management information system is a system platform suitable for the production, operation and management of this unit, which is organized by experienced personnel to form a development team and completed with the joint participation of all units.

Any data describes the characteristics of an objective world entity or a dimension of a concept, and water conservancy data is no exception [13]. The objective world entity or concept can also be called an object, and the data of a certain field is the sum of the characteristics of domain objects. As a field
of water conservancy, water conservancy data is the sum of the characteristics of water conservancy objects. Through water conservancy objects and their relationships, water conservancy data are turned into organic water conservancy data resources. The water conservancy data model mainly includes water conservancy object classification system, entity object model, metadata model and multidimensional theme application (Figure 1).

Figure 1 Schematic diagram of water conservancy data integration and sharing framework

The construction of integrated management information system has been completed, which has realized the integrated management of office management, operation management, technical quality management, planning and design project management, surveying and mapping project management, digital archives and other daily work; At the same time, it realized the sharing of water conservancy information data supporting the operation of the above modules, and initially entered the threshold of water conservancy big data management.

Because the data in the integrated management information system can be consulted, compared and (selectively) output according to the divided roles, the utilization rate of data and the value of data are improved, which provides reference for decision makers to grasp the engineering design progress and the workload of personnel in real time, and provides an effective basis for their decision-making and management, thus further improving the management level, product quality and work efficiency.

4.3. Design of cloud computing water conservancy information data sharing system based on P2P technology

In view of the slow download of FTP in the center at present, this paper proposes a cloud computing water conservancy information data sharing service system based on P2P technology [14]. The system adopts the combination of C/S structure and PP technology to realize water conservancy information data sharing, as shown in Figure 2. In the system, the central server provides the source data, and the clients join the network by establishing a connection with the server to form a cloud and provide services to the outside world in a unified way. Clients include super nodes and ordinary nodes.

Figure 2 Cloud computing water conservancy information data sharing system based on P2P technology

(1)Super node
Supernodes are clients conditionally selected from a large number of ordinary nodes, which have the functions of information storage, search, network relationship maintenance, network resource management and allocation. When selecting super nodes, according to the physical topology of the network, the adjacent nodes are clustered into one class, that is, a cloud is established for users, and the node with the best comprehensive conditions such as performance and bandwidth is selected as the super node in the cloud. In order to improve the search efficiency, a DHT network is built between super nodes, and the nodes use DHT to search for directional information [15].

(2)Ordinary node

Common nodes constitute the whole bottom P2P water conservancy information data sharing system, and connected clients can exchange resources with each other, thus reducing the download time of data files and the load pressure of servers [16]. Ordinary nodes upload their own file information, user information, etc. to super nodes for distributed storage, or send search commands to super nodes to find resources.

(3)Node administration

The node management module is mainly responsible for selecting super nodes. When the super nodes are overloaded, the super nodes exchange node information and dynamically adjust the network to achieve load balance. In order to solve the problem that some selfish nodes in the system only use network resources, but do not provide resources themselves, the system designs the following rules: download data below 50GB, and there is no limit on the amount of sharing; If the cumulative downloaded data volume exceeds 50GB, the shared data volume is at least 1/2 of the downloaded data volume, otherwise the system will refuse to download.

(4)Data download

Resources in P2P network are shared. Resource owners upload their own file information and user information to super nodes first [17]. When a node in the system needs related resources, they search in P2P network and establish a connection after searching. When no available resource node is found, they download resources from the server. This resource downloading strategy makes the network have good scalability, reduces the server load and minimizes the server bottleneck problem.

4.4. Share content

(1) Catalogue of water conservancy data resources

According to the characteristics of each type of water conservancy object, the data resource catalogue is constructed with reference to the usual water conservancy government affairs and business classification (such as administrative management, watershed, scale of water conservancy object, etc.), and the water conservancy data center uniformly constructs the data resource catalogue database as shared content to provide water conservancy data resource catalogue service to the outside world.

Data sampling depends on the size of sample data. Generally speaking, we use sampling accuracy to calculate the sample capacity, which is mainly produced by reliability and error, with reliability expressed as \( 1 - \mu \) and sampling accuracy expressed as \( (\mu, \nu) \). When the sample size \( m \) approaches infinity, the average number of samples can be treated as a normal distribution, namely:

\[
1 - \mu = p \left\{ \frac{|\overline{v} - \bar{v}|}{T(\overline{v})} < \alpha \right\} \quad (1)
\]

In which \( T(\overline{v}) \) represents the standard deviation of \( \overline{v} \), and the confidence area of \( \bar{v} \) when the confidence level is \( 1 - \mu \):

\[
\overline{v} - \alpha * T(\overline{v}) < \bar{v} < \overline{v} + \alpha * T(\overline{v}) \quad (2)
\]

\( T(\overline{v}) \) is defined as:

\[
1 - \varphi(T(\overline{v})) = \frac{\mu}{2}; \varphi \quad \text{obeys normal standard distribution, and} \quad \alpha \quad \text{can be found}
\]
from $\mu$ value of normal distribution table [11]. According to the sampling accuracy:

$$\alpha_t \ast T(\bar{v}) < v$$

(3)

It is known that the standard deviation of samples in simple random sampling is:

$$T(\bar{v}) = \sqrt{\frac{1}{m} - \frac{1}{N}} S^2$$

(4)

Simplify formula (3) and formula (4) to obtain sample capacity $m$:

$$m \geq \frac{1}{\left(\frac{\alpha_t}{S^2}\right)^2 + \frac{1}{N}}$$

(5)

In formula (5), $N$ represents the total sample capacity, and $S$ represents the population standard deviation [12]. When $N$ is very large, this formula is close to:

$$n = \frac{\left(\frac{\alpha_t}{\beta}\right)^2}{\alpha_t S^2}$$

(6)

(2) Basic database of water conservancy

The water conservancy data center builds a unified water conservancy basic database, which mainly covers the data sets used to identify individual attributes of water conservancy objects (such as water conservancy object code, name, location, river basin, etc.) and water conservancy object relationships (such as spatial data describing spatial relationships and various business relationships, etc.) as shared contents, providing authoritative, comprehensive, complete and consistent basic water conservancy information data sharing services for water conservancy business or government affairs applications.

(3) Shared database of water conservancy application

The water conservancy data center uniformly builds an application sharing database, which mainly covers the attribute data of water conservancy objects generated by specific water conservancy government affairs and business activities that need to be shared among different water conservancy government affairs and businesses (such as precipitation, hydrology and flow generated in hydrological business activities, water resources bulletin data generated in water resources government affairs management work, etc.). As a shared content, it provides authoritative, comprehensive, complete and consistent application water conservancy information data sharing services for water conservancy government affairs and business applications.

5. Conclusion

As a new data application mode, big data technology is being popularized and applied in the fields of scientific calculation and collaborative design in the water conservancy industry. The informatization practitioners in the water conservancy industry should form their own realizable development ideas according to the needs of the industry, the promotion of application means and the informatization development goals. Using the most advanced database technology is the key to process and utilize the massive hydrological data information to get all kinds of useful information we need. In the new era, building a data mining platform based on cloud computing technology can greatly improve the efficiency of data processing and provide better experience and service for users. The construction of water conservancy data sharing cloud platform should focus on the construction of data storage standards, data security and resource management. At present, the establishment of water conservancy data sharing cloud platform lays a foundation for subsequent research and application, and will accelerate the pace of water conservancy informatization construction.
References

[1] Hu Shan. Research on data mining platform architecture and its key technologies based on cloud computing. Wireless internet technology, vol. 17, no. 05, pp. 125-126, 2020.

[2] Zheng Wen, Wang Lin. weather radar data sharing service based on water conservancy cloud storage technology. Water conservancy informatization, no. 2, pp. 34-37, 2020.

[3] Baofeng, Hui, Guoqing, et al. Research on the Random Corresponding of Privacy Data Mining in the Association Rules of Cloud Computing[J]. Journal of Physics Conference Series, vol. 1187, no. 5, pp. 052006, 2019.

[4] Li Houmei. Research on Massive Data Mining Based on Cloud Computing. Information Technology and Informatization, vol. 226, no. 01, pp. 127-129, 2019.

[5] Zhu J. Research on data mining of electric power system based on Hadoop cloud computing platform [J]. International Journal of Computers & Applications, vol. 41, no. 3-4, pp. 289-295, 2019.

[6] H Zhang, Z Zhang. Research on the Big Data Cloud Computing Based on the Network Data Mining [J]. International English Education Research: English Version, no. 2, pp. 72-74, 2019.

[7] Xue Shixiang. Discussion on the application of big data technology in water conservancy informatization construction. Science and Informatization, vol. 000, no. 006, pp. 42-42, 2019.

[8] Li Ya, Song Bo, Lv Pengju. Thoughts on the integration and sharing of water conservancy information resources in Henan Province. Henan Water Conservancy and South-to-North Water Transfer, vol.48, no. 10, pp. 88-89, 2019.

[9] Zang Min. Give full play to the role of hydrological measurement and control system in water conservancy informatization. China Flood Control and Drought Relief, vol. 29, no. 05, pp. 17, 2019.

[10] Xin Huawei, Wang Jinguo. on water conservancy automation and water conservancy informatization. encyclopedia forum electronic magazine, vol. 000, no. 024, pp. 792, 2019.

[11] Wang Yuli. Application of data source physical concentration technology in water resources information integration. Digital Technology and Application, vol. 38, no. 10, pp. 53-55, 2020.

[12] Xu yuehua. application of big data technology in water conservancy informatization construction. science and informatization, vol. 000, no. 005, pp. 22-22,24, 2020.

[13] Kumar R. Assessment of Water Quality Parameters and Identification of Microbial Diversity in the High Altitude Sacred Lake Dodi Tal, Garhwal Himalaya. Taiwan Water Conservancy, vol. 68, no. 4, pp. 22-33, 2020.

[14] Nobora J, Abracia A, Daag AS, et al. Flood Inundation Modeling of the Pasig-Potrero River Draining from Mount Pinatubo to the Pampanga River Basin using HEC-RAS. Taiwan Water Conservancy, vol. 68, no. 2, pp. 28-44, 2020.

[15] Donald E, Mbekee F. Water Quality and Plankton Assessment of Ossah River, Umuahia, Southeast Nigeria. Taiwan Water Conservancy, vol. 68, no. 4, pp. 34-46, 2020.

[16] Bhat M A, Abbasi T, Abbasi S A. A Household-scale on-site Embodiment of the Novel Bioreactor SHEFROL for Treating Greywater. Taiwan Water Conservancy, vol. 68, no. 2, pp. 26-35, 2020.

[17] Kumar R, Sharma R C. Exploration of Psychrophilic Microbial Diversity and Physicochemical Environmental Variables of Glacier-Fed Lakes in the Garhwal Himalaya, India. Taiwan Water Conservancy, vol. 68, no. 3, pp. 38-51, 2020.