

Construction of Data Sharing System of Water Information Based on Big Data Technology

Zhiming Li

Jiangxi Water Resources Institute, Nanchang, Jiangxi, 330013, China

Keywords: Big technology; Water information; Sharing system; Data storage

Abstract. As a new technology data management and application model, the development of big data technology provides a new feasibility of technology for the new situation and the new task of water conservancy industry information construction. The practice of water conservancy informatization construction in China was analyzed, together with the difficulties and challenges in the construction of future data sharing platform. The integration of basic resources is realized with the help of big data technology, and infrastructure, water business applications and other resources are planned as a whole part to achieve the sharing system completion. In addition, the development of data storage standards and data security also provide the basis for the protection of data sharing system and successful operation.

1. Introduction

Water information is using the information technology to carry out water information collection, processing, storage and sharing, and thus achieving the purpose of improving the efficiency of water conservancy. Water conservancy belongs to a social welfare business, which gets support from the country. Water conservancy management is a combination of watershed management and regional management. The construction of water information is the main way to realize the integration of river basin information resources, and also it is an major approach to achieve the goal of sharing basic information of water resources, geospatial information and water management information [1, 2]. The water information sharing system will play an important role in serving the whole social and economic development.

Water information construction can be traced back to the late 1990s, it developed from the earliest CAD two-dimensional map by the universal computer to the use of three-dimensional collaborative design [3]. All the related new technology has changed the working method and has improved the efficiency of computing. At present, with the development of the Internet of Things, the cloud computing, the mobile Internet and other new technology, the digital level of the water industry has greatly improved. Besides, a series of growth of the accumulation for data resources and the formation of a massive data source has declared the advent of water data era.

Big data technology is used to collect and analyze huge amounts of data. This technology performs a comprehensive analysis and processing for all the data resources, and then provides a strong support effect for the specific objectives and tasks, which is different from the random analysis method[4]. At present, the construction of water conservancy information in China has made initial achievements in information acquisition and processing and the construction of water conservancy information data storage has also started. It is the most needed task that how to effectively use large data technology to organize and manage water conservancy information data, and how to construct water conservancy information data sharing system to improve the efficiency of data management and utilization.

2. Related theory and research status

2.1 The big data thchnology

Big data has become the commen focus for the govement, academia and business since it was first put in Nature in 2008[5]. The big data technology has change our life and driven the revolution of our country, enterprise and society[6]. Figure 1 shows the evolution of big data technology. In 2012, the America published the “ Big Data Research and Development Initiative” and started the program

“Big data research and development plan” which intended to strengthen the research of the collection methods for massive data and to find the way to enhance the ability of analysing and fast extracting complicated data. There are about two major characteristic of big data technology. Figure 1 shows the evolution of big data technology.

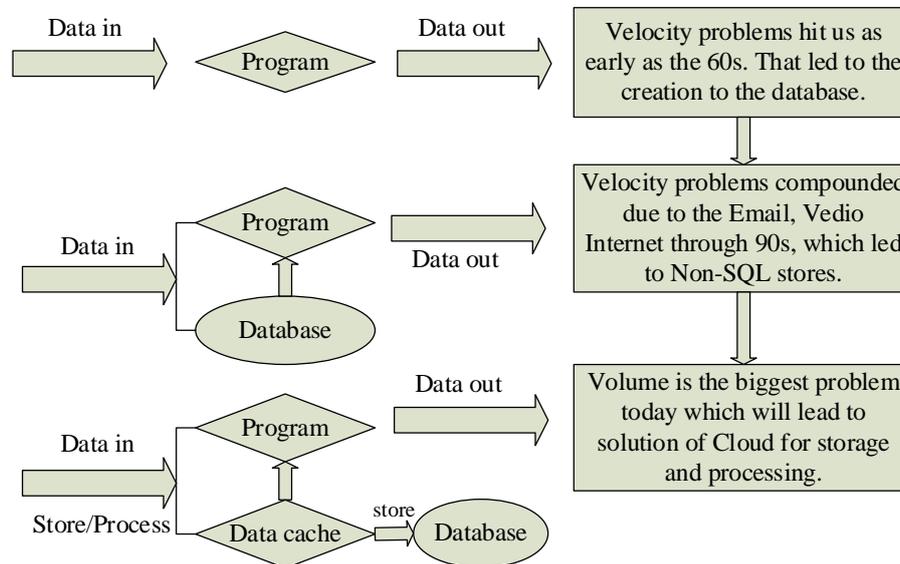


Figure 1. The evolution of Big Data technology.

The first feature is the complexity of big data and mainly reflects on the complexity of its type and structure. First of all, the complexity of the data type is changing with the changes of the Internet of Things technology progress, increasingly rich data types and data collection, which indicates that the new data storage and processing technology also need to continue to innovate. On the other hand, the traditional data processing methods can not meet the requirements of modern multi-type multi-temporal data fusion, such as the integration of object geographic information and its attribute information, and the combination of different spatial and temporal information of the same object. And then the second complexity of the big data reflects in its structure. The current research focus is concerned with the basic analysis of unstructured data, however, due to the diversification of information collection methods, the traditional way of storage, analysis and mining of the commonly used unstructured data, like images, video and other networks storage, must be changed. At present, a lot of research about the unstructured data analysis has done at home and abroad. For example, Google developed the non-relational database to store unstructured data to achieve a variety of distributed architecture. In China, research and development of unstructured processing has also been largely invested and achieved quite good results [7]. For example, Baidu and Ali build a cloud service platform, and many industries carry out the Internet+ action.

The second feature of big data is its uncertainty. The first is the uncertainty of the data, referring to the different sizes and dimensions of the uncertainty. The reason is that the different phases corresponding to the requirements of original data acquisition and processing for the granularity, system or user application is different. Specifically, in the face of massive, multidimensional, multi-type uncertainty data, there are new ways to solve the big data collection and storage, modeling, and other aspects of the uncertainty. Then is the uncertainty of model. In the modeling of uncertain data and system design, the early most commonly used model is the "possible world model". But in practical applications, the model will usually be simplified to facilitate the rapid use of data. At present, the most widely used in the modeling field of uncertain data is the probability graph model. It combines the knowledge of probability theory and graph theory, and provides a simplified representation of mathematics for complex reasoning and computation.

2.2 Practice and problem of water conservancy information sharing

The hot spots of water conservancy information construction mainly focus on drought and flood warning, water ecological monitoring, water conservancy facilities monitoring, water resources dispatching, water conservancy e - government and so on. At the end of the 20th century, Chinese water conservancy scientific research workers put forward the concept and idea of water environment information [8]. Therefore, the raise of digital river basin made the information technology and 3S technology more widely applied to water industry. In the strong support of country, all local water conservancy information management units invested in a large number of resources to build flood control and drought information system in the software and hardware. They both equipped with high-performance computers, servers, high-speed network, but also have built a number of flood control drought management software [9]. The water conservancy information data can be divided into two categories: the engineering attribute data and the water quality data. The data of the project attribute data are static data, which usually do not change with time. This mainly relates the engineering objects in construction of water information sharing system. The physical data mainly includes the data generated during the hydrological situation, water resources, water environment and management research, and the data change with time. A large part of the water information management system established in the water industry is spatial data.

Database technology is one of the important tools to realize water informationization. It can effectively integrate various data into a unified environment, and plays an important role in storing various information for statistics, query and analysis. At present, the major river basin agencies rely on the database technology to establish their own water data center, and their internal system information sharing has been basically realized.

Water conservancy information construction started late in China, but the development is rapid. Computer technology, communication technology and network technology have been widely applied in the water system. Besides, the remote sensing technology and telemetry technology have promoted the transmission of hydrological basic information. The sharing of information data is the most direct display of water conservancy information construction, and also the most effective way to use. However, there are also some problem in the sharing of water information with the development, mainly reflecting in the following several aspects. First, the early water conservancy information construction management is more dispersed. The integration of the built and unbuilt information data center debris is the current primary task of water conservancy information construction. The lagging infrastructure mainly reflected in the backward and inefficient data collection means, which affected the quantity and accuracy of water conservancy information. Second, the informationization of the grassroots department is weak. The sharing of resources and the acquisition ability is limited and the utilization rate of hardware resources is low which casue difficulty in the maintenance and resource integration as well as low efficiency of organization. Third, the scope of application of water conservancy information construction in various places and departments is narrow. The information storage is the core of water conservancy information construction, but there is no unified storage technology standard, which leads to disagreement in nationwide. The demand for intensive applications, such as water conservancy census, hydrological water quality and other basic data is needed to improve continuedly. And, to improve the water conservancy information system security and defense capabilities has become an unavoidable problem for water sector.

3. The construction of the water information sharing system

The premise of water conservancy information sharing system construction is the integration of resources, infrastructure, water business applications and other resources for the overall planning. With the deepening demand for water resources and information technology, the problems, such as the adaptability and sustainability of the original design and partial construction achievements, have become increasingly prominent. Carrying out water conservancy information project after the assessment will provide a scientific basis for the planning and adjustment of the macro investment policy[10, 11].

3.1 Construction of the sharing platform

With the current water conservancy information construction requirements of the continuous improvement, big data technology is used to design three-dimensional digital collaborative platform. The content of the platform includes heterogeneous database integration, unified management of resources, resource sharing and access management. The shared technology is used to achieve resource sharing and application integration, the establishment of data sharing platform to achieve data sharing from the application of direct access to non-proprietary database through the shared platform is an creative idea. Shared water information platform consists of two parts: first is the computing platform. This platform includes computing resources, storage resources and network resources. Second is the processing platform, including analysis, management and application. Water information collection includes water information, rain information and other content, and also, the first line collection of hydrological stations and water monitoring stations and other fixed terminal equipment. The mobile information collection terminal is set as the core of the computing platform. Network communication includes water resources special network, satellite network and other services. Figure 2 is the four-tier structure for big data resources of water resources industry. In addition, the data application stage uses a unified information-based data storage platform to provide users with industry-specific information services. Infrastructure construction and platform services has gradually achieved virtualization, and the post-evaluation work has been performed on the base of past construction of water conservancy information database. Therefore, the unified database type and information sharing framework is constructed. For the convenience of use and maintenance, open source systems, Nimbus, is used to build water information cloud platform.

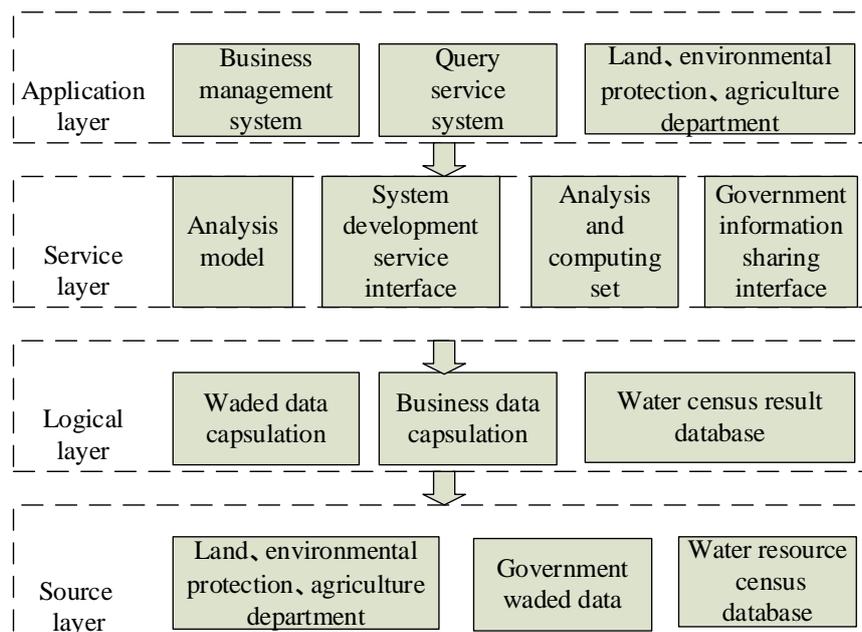


Figure 2. The four-tier structure for big data resources of water resources industry,

3.2 Technical route of big data for water information sharing

The value of big data is to share and service, and the water resources of this project should be based on the basic service platform of water conservancy data. Besides, the resource technology should be explored, and the data should be distributed heterogeneously and efficiently. The technical route of water conservancy information sharing system construction using big data technology should be based on the resource layer, and be relied on the logic layer for technical support. As shown in Figure 3, the standard application interface is designed baesd the service layer and application layer. The logical layer serves as the support layer of data development, which provides diversified application for data heterogeneous processing, composition, encapsulation, delivery and sharing. It is a support layer of data development, which provides secure, reliable and continuously updated storage capacity for data. The service layer is also used as the support layer of data development, but also focuses on creating a dynamic and scalable service form of the application structure of water

resources large data to meet the basic application of the practitioner to the large data business. The application layer is the support layer of the intermediate interface, reflecting the value of large data of water conservancy. Therefore, the users can be in accordance with the standard platform interface and application interface to develop the system interface.

3.3 Database and data storage standard construction

The contents of the standard sharing system database include the basic data (including the surveying and mapping data, hydrological and water resources information, geographic information data, modern water network information), business data (such as the water industry knowledge data, historical basis collection data, hydraulic building information), document data (like the survey results data, project data, design results file data). And then a unified library management based on the collection, summary and statistics will build an easy using standard base library, as shown in Figure 3.

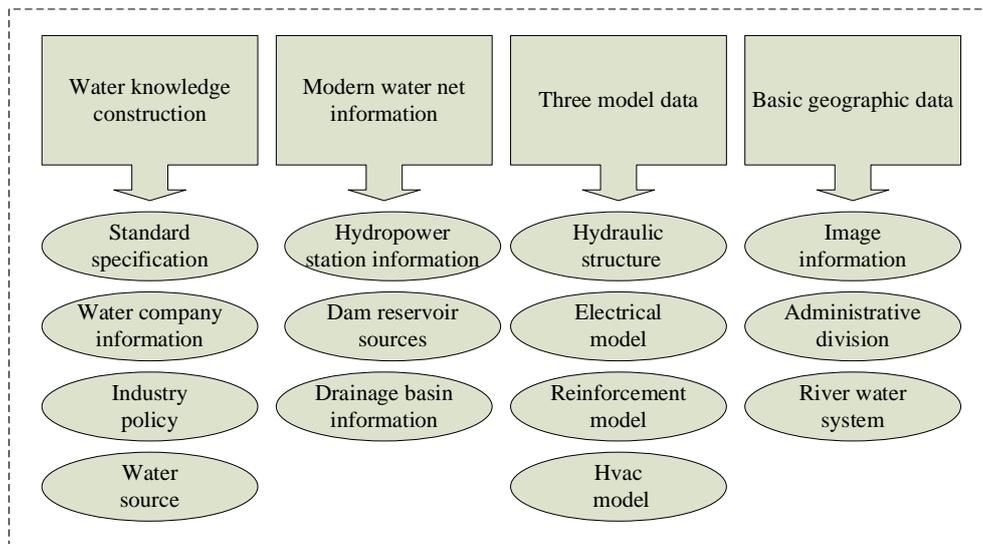


Figure 3. The underlying database model for the construction and management.

As we all know, data element is a group of data can be identified and defined as the basic unit, and plays an important role in the water information data sharing system. The combination of water conservancy information and spatial information is an important part of water conservancy informatization. So the construction of hydrological information storage standard has a guiding role in water informationization. Table 1 shows the storage methods and costs. The NAS method is relatively cheap[12]. Therefore, the water conservancy information sharing system is based on the DBMS database and

NAS method is used to store information in order to effectively organize and manage the hydrological information. In addition, with the advent of GIS products, combining the hydrological spatial information into big data technology will make the hydrological data storage more convenient and efficient.

Table 1. Storage methods and costs

Storage	Dollars/Giga Byte/year	Estimated Cost/year(\$)
SAN	4.00	1000000
NAS	0.65	162500
Cloud	0.312	78000

3.4 Sharing system management construction

The construction of the integrated management information sharing system is based on the "Quality, Environment, Occupational Health and Safety Management Handbook", and related system documents [13]. For the purpose of collaborative management and efficient application, the organization of experienced staff to form a development team to participate in the completion of the

production, operation, management and other work system platforms. The whole management system include the office management, technical quality management, planning and design project management, surveying and mapping project management, digital archives and other daily work integrated management. At the same time, this system will be preliminary into the field of big data management when achieving the support of the above modules to run the sharing data.

Summary

Water resources information sharing system will greatly accelerate the development of water conservancy and related industries in China, a variety of water information can be accessed through the water resources information sharing platform. Big data technology have been promoted and applied in water conservancy industry scientific computing and collaborative design as a new data application model. When constructing the water conservancy information sharing system based on the big data technology, the corresponding policies and regulations should also be developed, so that the water information sharing will be lawful. In addition, it is necessary to develop a meta-data standard adapted to describe the scientific water conservancy data. The water information can be provided the users with multi-directional service on the shared platform.

Acknowledgement

This work is supported by Social Science Project of Jiangxi Province and Journal of Jiangxi Resources Institute Quarterly.

References

- [1] Chen M, Mao S, Liu Y. Big data: A survey[J]. *Mobile Networks and Applications*, 2014, 19(2):171-209.
- [2] Liu Qingtao, Cui Ruiling, Geng Dingrui. Discussion on construction of water information cloud[J]. *Water Resources Information*, 2012(2):5-9.
- [3] Vickers A. Handbook of water use and conservation[M]. WaterPlow Press, 2001.
- [4] Marz N, Warren J. Big Data: Principles and best practices of scalable realtime data systems[M]. Manning Publications Co., 2015.
- [5] Big Data[J]. *Nature*. 2008,455(7209):1-136.
- [6] Dealing with data[J]. *Science*, 2011, 331(6081):639-806.
- [7] Erl T, Khattak W, Buhler P. Big Data Fundamentals: Concepts, Drivers & Techniques[M]. Prentice Hall Press, 2016.
- [8] Fang S, Xu L, Pei H, et al. An integrated approach to snowmelt flood forecasting in water resource management[J]. *IEEE transactions on industrial informatics*, 2014, 10(1): 548-558.
- [9] Curry E, Degeler V, Clifford E, et al. Linked water data for water information management[J]. 2014.
- [10] He Hong, Liu Cong, Gan Haoxin, et al. Study on the post- project appraisal of water informatization project[J]. *Renmin Zhujiang*, 2013(1): 68-71.
- [11] Chen Lan, Zhan Quanzhong. Review and prospection of water information system construction[J]. *Journal of China Hydrology*, 2011(6):67-69.
- [12] Cloud and Edge Computing for an IoT-Based Smart Grid.
<https://dzone.com/articles/cloud-computing-and-edge-computing-for-an-iot-base>
- [13] Calder I R. Blue revolution: integrated land and water resource management[M]. Routledge, 2005.