Research and Application of Application System Architecture Based on Big Data

Hongyu Li*, Ling Gao and Xiaojun Yang

Information and engineering school, Yunnan technology and business university, Kunming 650000, Yunnan, China

*Corresponding author e-mail: lihongyu@ynsxy.edu.cn

Abstract. With the advent of the era of big data, traditional application systems have been unable to meet the actual needs of the development of various industries. Building an application system architecture based on big data has become a top priority for many enterprises and organizations. In this context, based on the previous research, this paper proposes an application system architecture model based on big data and based on the application system design principle. The model is divided into data source layer, data storage layer, data acquisition and processing layer, data analysis layer, and data application layer. After a detailed description of the functions and characteristics of each layer, an application example is given for further analysis, hoping to provide some reference for each industry in building a big data application system architecture.

Keywords: Big Data, Application System, Architecture, Application

1. Background of the Study

1.1. Literature Review

With the rapid development of Internet technology, users have changed from simple data users to data producers, and the popularity of mobile terminals provides the possibility of data collection (Bland A R et al., 2009). At every moment, a large number of new data are constantly generated, and the scale of data has risen substantially. At the same time, data types are heterogeneous. These data include not only the user's evaluation of goods, but also the user's travel records, specific geographical location and so on. According to incomplete statistics, the global Internet users' unused traffic is 1.3EB. Take Sina Weibo, a popular microblog in China, as an example. According to the statistics of the official microblog platform, the daily login volume is 340,000, and the daily return volume is 350,000. The number of active Facebook users promoting more popular Facebook has exceeded 1 billion. Under this background, the concept of big data has been formally put forward. The so-called "big data" refers to the huge amount of Yang Hua data assets that need to use new information processing technology to truly have the ability of discovery, decision-making and process optimization. The real meaning of big data is not to master massive data information, but to process these data information professionally and extract truly effective and valuable data [1-3].
In view of the huge value hidden behind the big data, the government and enterprises urgently need an efficient application architecture to serve as internal reference, so as to deal with the massive data resources collected. At present, distributed parallel computing is a mature way of large data processing. Chen Li puts forward a reference model of application architecture based on hierarchical design. The model is divided into data source layer, data acquisition and processing layer, application layer, analysis layer and application layer. A specific application case of WEB online system combined with big data application is given (Chen, 2014). Zhong Yuan put forward that Apple, Microsoft, Huawei and other well-known international enterprises have begun to focus on the construction of their own ecosystem under the background of big data (Zhong, 2014). Zhang Bowen, Yan Chunyu and others have sorted out the specific business needs of large data in power grid companies, proposed specific functions to meet various needs, and completed the design of early warning system architecture (Zhang et al., 2016). Wu Ningbo focused on the specific implementation process of big data application system in auditing, combining pioneering R language with process language, and reforming the existing information processing mode (Wu, 2014). Li Qiang, Zhao Chenjie and others have designed a Hadoop-based analysis system based on informational and experiential teaching behavior data (Li et al., 2018). Guo Jiguang and Huang Sheng proposed to explore and analyze the architecture design and operation of intelligence analysis system in military field based on big data technology (Guo and Huang, 2017).

1.2. Purpose of the Study
With the continuous improvement of big data technology, people's daily life is more and more closely related to its practice. Generally speaking, big data system is a new type of system which is different from traditional application system. Behind it is a huge network database resource. Through efficient processing mode, data value can be fully exploited to provide users with better service (Guan and Zhan, 2017). By combing the existing literature, we can find that although the research on the architecture of large data application system has been gradually on the right track at this stage, overall, the research is still in the initial stage, and the architecture of large data application system designed is far from practical application. Based on this background, this paper is based on the actual needs of large data application system, and analyses the large number in detail. According to the specific principles that should be followed in the design of application system, a systematic architecture of large data application system is constructed in order to provide some reference for the design and use of subsequent system architecture [4-8].

2. Architecture Design of Application System Based On Big Data
With the rise of social systems such as Sina Weibo and Ding Talk, a large number of structured and unstructured data are generated. One of the important functions of big data is to locate the data generated by these social systems, thus providing a distributed data processing platform for further mining and analysis of massive data. The big data application system is built on the Hadoop platform to collect and organize Internet user data from third-party interfaces through direct or indirect methods, providing possibilities for the development and deployment of large-scale data analysis systems.

2.1. Analysis of Application System Requirements Based On Big Data
At this stage, the main application scenarios of big data application system research have the following categories. The first is the brand value model. By establishing a customer's value model for a certain type of brand, collecting user evaluations from social platforms such as Weibo, and using distributed neural network systems to analyze user evaluations and calculate the brand value. Secondly, it monitors the corporate image, selects the keywords corresponding to the enterprise, searches on the portals such as Weibo and Tianya, obtains the real evaluation of the customers, uses the sentiment analysis system to evaluate the corporate image, and provides data reference for the evaluation of the advertising effect of the enterprise. Make an image promotion strategy targeted at the company (Li, 2016). Once again, it is a hot topic mining, using community mining and other means to deeply
explore popular keywords, so as to dynamically track hot events and topics, and provide evidence for social, government, and enterprise research institutions to formulate relevant policies (Liu, 2018).

Finally, the video recommendation, in this session, the application system first uses the user history on-demand data to obtain the data association mode displayed and thickened during the user's viewing process, thereby utilizing the existing user preferences to accurately recommend the user to be interested video [9-11].

2.2. Architecture Design Principles of Large Data Application System

Big data challenges traditional application architectures, but it also brings new problems. For example, the original analysis method can not meet the requirements of accurate mining of valid data. In the process of enterprise application, big data application system needs to be further improved in reliability, availability, fault tolerance, security and privacy (Zhu and Lu, 2017). From the above analysis, we can see that in the process of designing the architecture of large data application system, we need to follow the following principles: The first principle is that we need to meet the 3V requirements of large data, that is, to be able to process, load and analyze large data capacity level. We need to improve the compatibility of data types, have the ability to load and process different data at the same time, and the processing speed should be adequately guaranteed. The second point is to take full account of the actual needs of the application subject. The architecture of the large data application system should have security, openness, reliability and so on. The last point is that we must meet the requirements of data processing and analysis of the original format, and be able to analyze the complex original format.

2.3. Architecture Model of Large Data Application System

At present, the generation, processing and organization of large data are mainly realized by distributed storage technology. Among all kinds of distributed technologies, Hadoop technology has won the trust of many experts because of its accuracy and high speed. Based on the analysis of the existing large data application architecture and its sale, this paper innovatively proposes a large data application model architecture which has nothing to do with specific products and can realize interoperability and integration with enterprise information architecture, as shown in Figure 1.

![Figure 1. Large Data Application Model Architecture](image)
2.3.1. Data Source Layer. For traditional data sources, strict definitions are often made at the initial stage. Large data sources are not constrained by this condition. At this stage, the data source layer will collect all relevant information as much as possible. In the preliminary analysis, due to the strong data acquisition ability in the collection stage, the data source layer will get a lot of cluttered information. Not only does the data structure present diversified characteristics, but its data sources are also different. According to the characteristics of data types, it can be divided into three categories. The one type is fully structured data, which can be presented in a pre-defined format. Take MySQL as an example, the table and column data have been defined in advance, just insert the data according to the definition. The second type is unstructured data. Common text, video and audio data have no well-defined structure, and the content is very different. The last one type is semi-structured data, which has certain logic and format, but also has a large number of invalid data. Valuable data need to be further excavated and analyzed.

2.3.2. Data Storage Layer. Different data formats and sources have different storage requirements. For example, relational database is suitable for storage of fully structured data, because its management model highly pursues correctness and consistency, other types of data can not meet this requirement. Semi-structured data is suitable for NoSQL database, because this database is easy to store without fixed data structure, and has good read-write performance. In a large amount of data and information, it can greatly improve the availability of the architecture. In the specific storage link, the mainstream Hadoop distributed file system technology is adopted. In order to ensure the accuracy and consistency of data, the model of "input once, read many times" is adopted. Provide uniform file naming rules to improve data access throughput.

2.3.3. Data Collection and Processing Layer. In order to fully guarantee the requirement of 3V information resources in the era of big data, a distributed parallel architecture is used to parse, and the reliability and extensibility of large data application system are guaranteed. Faced with different processing needs, the implementation of differential technology processing strategy to improve response speed. For example, when dealing with offline demand, the time requirement is not high. Hadoop technology framework with high latency can be used, such as offline statistical analysis, machine analysis and so on. For online real-time requirements, a streaming analysis architecture is needed. Especially for financial products and Internet service products, the real-time requirement of data is very high. In order to ensure the normal use of users, a large number of data feedback and analysis must be carried out in seconds. In addition, data integration tools are used to migrate data to decision analysis system and form a complete database. Structured data should be stored using Apache Sqoop tools.

2.3.4. Data Analysis Layer. The preliminary results from the data collection and processing layer will be transmitted to the data analysis layer in the first time. Preliminary results will be loaded into data marts, traditional ODS and data warehouses. Further data mining and analysis work will be carried out on these platforms. In view of the differences of data mining tasks, task analysis can also be divided into many types. Classification analysis, sequence analysis, Association analysis, model prediction analysis based on neural network or decision tree is common.

2.3.5. Application Layer. The data analysis layer will be equipped with advanced analysis and intelligent display software, and the resulting analysis structure has strong visualization characteristics. All kinds of data analysis results need to generate reports directly for further queries. In this level, the specific form can refer to the current market more mature large data analysis software, such as flight prediction, video guesses you like, and so on. In the application layer, there are also service management module and system security module. These two modules will cover all the mechanism policies related to service management and security. The design of big data application system needs to be constructed from five security dimensions: network, data, system, physics and management. In
addition, the corresponding technology and rules should also be formulated in time, such as access rights, security review mechanism, identity authentication mechanism and so on.

3. Design and Operation of Application System Architecture Based on Big Data

The significance of designing and running large data application system architecture based on large data is to provide platform support for distributed data mining system. Limited to space, this paper takes the brand value model as an example to introduce the specific operation of the framework of the big data application system designed. The experiment chooses the iPhone XR with more topics on microblog as the research object. Firstly, according to the UGC content and brand characteristics of the iPhone XR, the brand value model for the iPhone XR is established by using Woodeuff idea, and the Hadoop distributed technology is used to model, so that the model can predict the attributes of the corresponding application layer according to UGC content. In this process, the idea of BP neural network and artificial intelligence domain knowledge are introduced for model training, and the specific establishment process is shown in Figure 2.

![Figure 2. Brand Value Modeling Process](image)

First, collect the evaluation of iPhone XR from Weibo. The evaluation includes various attributes in the model definition, such as price, signal, screen, etc. This process is mainly carried out at the data source layer. After the evaluation is completed, the data is stored in each of the pre-specified memories to complete the distributed storage. Second, since the collected information cannot be directly used for modeling, preprocessing is required, such as digitization, normalization, data denoising, and removal of duplicate content. This process is mainly carried out in the data acquisition and out of the layer. Again, use the BP algorithm to model, then bring the data in and then enter the training phase. The purpose of the training is to make the actual analysis results obtained by the machine fit the reality. To achieve this goal, you must ensure a sufficient number of experiments. When the results obtained have good convergence characteristics, the experimental error is measured until the demand is met. This process is mainly carried out at the data analysis layer. After the model learning is completed, the prediction result stage is entered, and the evaluation of each attribute value in the sample is performed to obtain a generalized sample. Next, the target attributes are calculated and summarized. The experiment selects 5G experimental samples for model learning. After the end of the study, the 100G content statistics are obtained, and finally the target attributes of Iphone XR are obtained. The results show that the target attributes of the Iphone XR brand are positive, among which the exchange attribute value is the highest and the value attribute value is low.

4. Conclusion

The construction and popularization of big data application system is the actual demand of the development of all walks of life in the information society. In this context, we must constantly
improve the data storage capacity and improve the current large data architecture system. In the specific application, we should fully combine the characteristics of various industries, around the design principles of the large data application system, build a personalized and customized system architecture, and constantly improve the stability and reliability of the data system. Through the optimization and reorganization of the existing large data application system design module, the overall benefit of the data application system is improved. It should be pointed out that the architecture design of big data application system is still in its infancy, and there is still much room for information sharing and application in various industries. The specific methods need to be further explored and innovated.

Reference
[1] Bland A R, et al. (2009). Implementation and Testing of the OPT Model As a Teaching Strategy in An Undergraduate Psychiatric nursing Course. Nursing Education Perspectives, 30(1), 14.
[2] Chen Li. (2014). Research and Application of Application System Architecture based on Big Data. Software Industry and Engineering, 5(5), 33-38.
[3] Zhong Y. (2014). Research on Enterprise Ecosystem Architecture Based on Big Data Background. Network Security Technology and Application, 14(6), 208-209.
[4] Zhang B.W, Qi C.Y, Bi J.G, et al. (2016). Research on Architecture of State Alarm System for Power Transmission and Transformation Equipment Based on Big Data. Electric Power Information and Communication Technology, 14(12), 30-36.
[5] Wu N.B. (2014). Research on Reconstruction Strategy of Audit Information System Architecture Based on Big Data. Computer CD Software and Application, 2014, 2(17), 264-264.
[6] Li Q, Zhao C.J., Luo X.L. (2018). Analysis and Design of Academic Analysis System Based on Big Data Application Technology. Software Engineering, 21 (05), 38-41.
[7] Guo J.G., Huang S. (2017). Research on Military Intelligence Analysis and Service System Architecture Based on Big Data. Journal of China Academy of Electronics Science, 12(4),389-393,413
[8] Guan F.F., Zhan P.Z. (2017). University Education Big Data Application Research and Future System Architecture Outlook. Jiangsu Communications, 24(04), 39-41.
[9] Li D.K. (2016). Research and Application of Intelligent Public Security System Based on Big Data. Journal of Network and Information Security, 2(12). 63-68.
[10] Liu W.J. (2018). Research on Information System Command and Control Based on Big Data Technology. Electronic quality, 2018, 379(10), 16-18.
[11] Zhu J.L., Lu S.D. (2017). Application Research on Management Accounting System of Communication Enterprises Based on Big Data. Friends of Accounting, 34(2), 16-19.