Research About Big Data Platform of Electrical Power System

Dongmei Liu, Guomin Li, Ruixiang Fan, and Guang Guo(✉)

State Grid Information and Telecommunication Branch, Beijing 100761, China
43383047@qq.com

Abstract. Along with the construction of intelligent power grid and the continuous expansion of it, power systems produce large amounts of data, it is particularly important to integrate, analysis and process, and traditional data processing technology is difficult to meet the demand. The technology of big data injects new vitality to the development of intelligent power system, in the area of power system mastering the key technology of big data for the sustainable development of the electric power industry and the establishment of a strong smart grid is of great significance. Firstly, this paper does research about the key technology and processing scheme due to big data in power system. Secondly, it explores big data in power system based on cloud computing architecture.

Key words: Power system · Big data · Cloud computing

1 Introduction

Although the industry has some consensus, the definition of big data did not have a unified definition. McKinsey suggested "big data refers to the size beyond the typical data capture, storage, management and analysis scope of data collection"; Gartner said "big data need to deal with the new model that can have better decision-making, insight found mass force and the process optimization ability, high growth and diversification of information assets". In 2011, the Science pushed out the special issue name "Dealing With Data", deeply discussed challenges the Data deluge (DD) bring about, and pointed out that people would get more opportunities to play a great role of science and technology to promote social development if these huge amounts of Data could be more effectively organized and used [1].

As the energy supply system, economic development and human life depend on power system, this system also has the typical characteristics of big data. Power system is one of the most complex artificial systems, it has the characters of widely distributed geographical location, real-time balance of generating electricity, huge numbers of transmission energy, electricity transmission speed can reach, highly reliable communication scheduling, real-time operation and never stop, it expands in major failure immediately [2]. These characteristics determine the data that generated by power system in large numbers, rapid growth and type is rich, it is full compliance with all the features of big data, and it is typical big data.

J. Wan et al. (Eds.): Industrial IoT 2016, LNICST 173, pp. 36–43, 2016.
DOI: 10.1007/978-3-319-44350-8_4

Under the situation of the smart grid deeply developing, it is more and more informationalized, intelligent and digital, and thus has brought more data sources, such as smart meters of terminals from millions of households and firms.

2 Processing Scheme of Big Data in Electrical Power System

In cloud computing mode, the power system of human-computer interaction and big data processing are both in the depth of fusion state. Therefore, based on cloud computing is the core of data processing, the key must be based on the integration thoughts, all of the network data cooperative organization, on the basis of integration, to maximize scatter power system of the comparative advantage of resources integration as a whole [3].

In cloud computing mode, the key to process huge amounts of data lies in the construction of power system, various kinds of scattered resources system and various other assistive technology. From a macro perspective, the whole big data processing can be divided into mixed processing and management. Mixed management is the key to realize the wireless power system, limited, resource sharing, scattered data management and the management synergy mechanism, etc. [2]; Mixing processing is the key to the fusion of large power system data processing operation model and relevant auxiliary technology [4].

At the same time, we need to focus on the redundant data processing. Redundant data processing scheme has the following several ways:

(1) only receive a set of data in a redundant data processing, abandoned another set;
(2) transfer redundant data to the higher level of data receiving equipment;
(3) filter redundant data as the unit of source equipment, if a set of equipment had any abnormal data we would take another set instead;
(4) receive the redundant data filtered according to the point as the unit.

3 Key Technologies of Big Data in Electrical Power System

3.1 Integration and Management Techniques of Big Data in Electrical Power System

The electric power enterprises' data integration management technology is to merge data from two or more application systems, to create a more enterprise application process. From the point of view of integration, it is the different sources and format, characteristics and nature of the data on the logical or organic concentration on the storage medium, providing the system of storing a series of subject-oriented, integrated, relatively stable, reflects the historical changes of the data collection, thus supplying comprehensive data sharing to the system [5]. Electrical power enterprises integrated management technology came into being is to solve the problem of the various systems of the internal data redundancy and information island between electric power enterprises.

Integration and management techniques of big data in electrical power system, including relational and non-relational database technology, data fusion and integration technology, data extraction technology, filtering technology and data cleaning, etc. An important feature of big data is diversity, this means that the data source is extremely widespread, data type is extremely complex, this complex data environment brings great challenge to big data processing, firstly it must carry on the extraction and integration of data source of data, and extract entities and relationships after association and aggregation, then unified structure is used to store the data, the data are required for data integration and extracting cleaning, guarantee the quality and reliability of data. Big data storage management is an important technology in no database technology, it adopts the distributed data storage method, and removes the relational characteristic of the relational database, data storage is simplified and more flexible and it has good scalability, which solving a lot of data storage problem. Representative no database technologies are Google's Big Table and Amazons chateau marmot [6].

3.2 Data Analysis Technology of Big Data in Electrical Power System

The fundamental driving force of big data technology is to convert the signal into data, to analysis the data of information, and to extract the information of knowledge, so that it can contribute to decisions and actions. With the big data analysis technology, we can find out the potential from the huge amounts of data in power system modal and law, and provide decision support for decision makers. McKinsey argues that the key technology that can be used for big data analysis is the result of statistics and computer science and other disciplines, including correlation analysis, machine learning, data mining, pattern recognition, neural network and time series forecasting model, genetic algorithm and so on.

Big data research study is different from the traditional logical reasoning, it is a huge amount of data to do a statistical search, classification, comparison, clustering analysis and induction, therefore inherited some characteristics of statistical science, such as statistics the data correlation or relevance of attention, the so-called "relevance" means two or more variables between the values of a certain regularity. The purpose of "Correlation analysis" is to find out hidden in the data set of networks, general with support, credibility, reflect the correlation parameters, such as interest in degrees [7].

Big data correlation analysis method, based on large amounts of sample, does not use the method of random such shortcuts, and adopts the method of analysis of all data; Big data of simple algorithm is more effective than small data of complex algorithm, the result is faster, more accurate and less susceptible to interference, we believe that based on correlation analysis of prediction is the core of big data.

Big data technology does not pay attention to the causality but focuses on the correlation analysis method, it has brought a big change in scientific research way of thinking, the late Turing award winner Jim gray data intensive scientific research, "the fourth paradigm" proposed by the big data research three former paradigm (theoretical science, computing science and experimental science) in isolated, alone. It can be a kind of research paradigm is because of the research methods of different from the traditional research method based on mathematical model [8].

Big data analysis technology in electrical power system, in essence, belongs to the traditional data mining technology in the new development of huge amounts of data mining, but the characteristics of huge amounts, high-speed growth, diversity determine it not only contains structured data, but also includes semi-structured and non institutional data, so many of the traditional data mining method of processing small data is no longer practical. Data mining and machine learning algorithms of big data in the environment of power system can be researched from three aspects:

(1) From the view of the management of big data and sampling and feature selection, changing the big data to small data;
(2) Do research about clustering, classification algorithm of big data, such as least squares support vector machine (SVM) based on conjugate degree (further squares support vector machine, the LS - SVM), random extensible Fuzzy C - Means (FCM), etc.;
(3) To carry out the big data parallel algorithm, through the parallelization of the traditional data mining methods, application to the knowledge of big data mining, such as machine learning and knowledge discovery based on graphs.

3.3 Data Processing Technology of Big Data in Electrical Power System

Data processing technology of big data in electrical power system includes distributed computing technology, memory computing technology, streaming technology, etc., these three technologies that can applicable objects and solve the main problem are shown in Fig. 1. Distributed computing technology is to solve large-scale data distributed storage and processing. Memory computing technology is effective to solve data reading and processing of online real-time computing. Stream processing technology is in order to get real-time, the speed and scale of uncontrolled data.

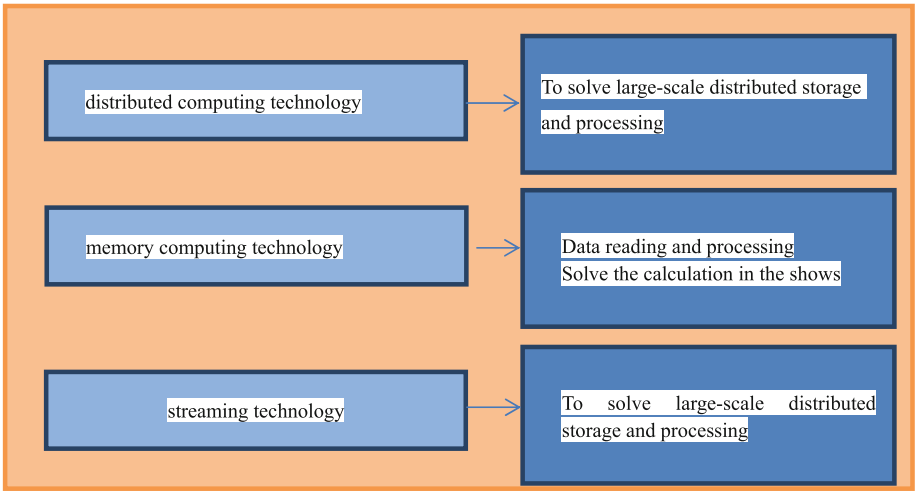

Fig. 1. Applicable object technology of data processing

In scattered data source of large-scale power system, we first use graphs programming model of distributed computing to block data that to be processed, then dividing into different Map missions, and the key-values stored in the local hard drive, finally we use the Reduce task to output the summary according to the key-value [9]. Memory computing technology puts all the data in the inner layer, it will overcome a lot of time consuming in the operations that disks read and write data, so the computing speed will increase greatly by several orders of magnitude. With the development of electric power industry, electric power system data quantity increasing, higher and higher to the requirement of real-time, applying data stream technology to power system can provide real-time basis for policy makers and meet the demand of real-time on-line analysis.

4 Electrical Power System Data Architecture Based on Cloud Computing

Cloud computing technology mainly has the following several aspects: data information network, service and customized and dynamic, etc., it has the characteristics that fully meet the power system's deployment and application of big data. According to the basic demand of large power system data applications, cloud computing technology can realize dynamic division or release of various resources, to realize the increase in the processing of resources, we can increase the resources available through proper matching to ensure that the resources are provided fast in the process of using elastic. If you no longer need to use the resources, it can realize the free release. Cloud computing technology configured dynamically, further implements the IT resources' continuous extension as needed (Fig. 2).

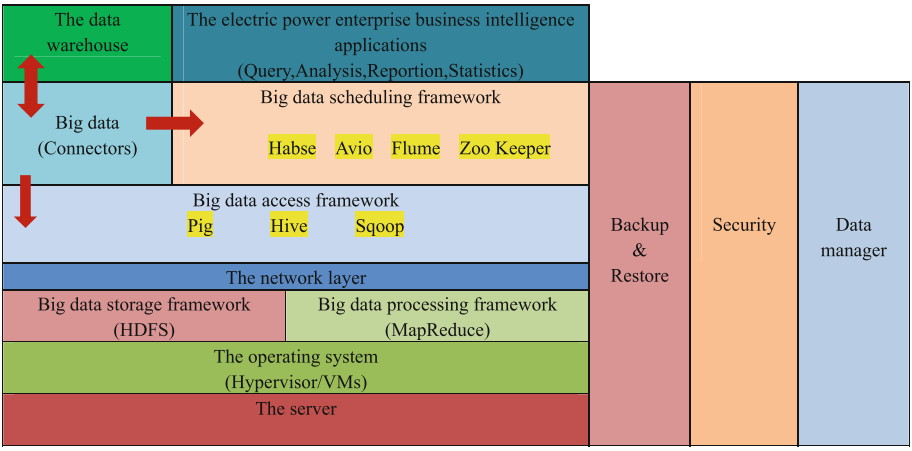

Fig. 2. The overall architecture of electrical power system data architecture

4.1 Network Architecture

By using cloud computing technology, to realize the change from traditional data physical model to the logical model, to realize the flexibility and agility of large power system in data processing through the construction of information network, to change the traditional IP network into SAN network, to change the big data processing network to carry on the further elaboration of layered, so that it can increase the elasticity of the network and flexible. Traditional physical network have been difficult to adapt to the requirement of cloud computing, virtualization, so by a variety of ways such as VXLAN, NVGER to change traditional physical network data flow type from two layers into three layers, to change the traditional physical layer into a logical layer in the network topology by increasing physical encapsulation at the same time, refining the classification management, so it can better meet the diverse needs from different types of needs from users [10].

Realizing the big data centralized control of the information network infrastructure. Through the automation network configuration of the scattered data, maximizing the efficiency of management and processing; Based on the data network virtualization processing to transit traditional physical network to the virtual abstract network, so that the application of the strategies of big data processing can be used to the workload. Comparing with the traditional processing way, this way has incomparable advantages: when the power load is deployed or migrated, the network configuration can be achieved automatically and adjusted to its settings.

4.2 Host Architecture

Host architecture plays a crucial role, it can be combined with bearing business characteristics to distinguish, to ensure that the system has reliable high-performance terminal, and effectively handle tightly coupled computing tasks through large-scale high-performance PC server or server cluster, such as large power system database, data mining, financial and marketing, etc.; Many general computing power system can provide low cost calculation and solution, especially in data processing of low hardware requirements of small and medium-sized application services business, this model not only can realize the processing of data, but also can effectively reduce the operation cost of the power industry through the adoption of high density and low cost intensive integration server cluster.

Host architectural pattern, on the other hand, has a fundamental change that from a traditional tower structure to a pool way. It can effectively achieve the host resources integration and optimization of the data processing system by adopting modern virtualization technology, modern virtualization technology and Rac One Node technology to realize database, such as composition and web level virtualization deployment and distributed among the Hadoop system, taking this model not only increases the speed of data processing, but also can effectively reduce the data processing system of all kinds of hardware failure of adverse effects on the business systems.

4.3 Storage Architecture

Storage system is the key to the whole system infrastructure and is the ultimate bearer of all the data in power systems. With the further development of various modern cloud computing, virtualization technology and big data technology, the traditional centralized storage has to be washed out gradually, and it is no longer the mainstream of data processing storage architecture. In order to deal with the access to huge amounts of data in electrical power system effectively, we must build the storage architecture with strong expansibility and scalability [11].

Cloud storage architectures based on cloud computing technology can effectively solve the problems that difficult to solve in the traditional architecture, it can store the data as cloud services. The key is to realize creation and distribution of huge amounts of data and gain the data through cloud service at the same time. The deployment of cloud storage technology mainly through the cluster or distributed file functions, the entire network of different types of storage devices in the system work together by related software, providing access to data storage and business, ensuring the safety of the whole system data meanwhile [12]. Through the distributed file system, such as object storage technology, providing scalable, extensible and strength data access function for all kinds of applications, meanwhile, as the adopted distributed technologies are based on standardized hardware technology, thus it can lower the cost of cloud storage and ops.

5 Conclusion

This paper expounds the big data processing scheme of big data platform in power system, and then expounds the key technology of four cores, namely the integrated management technology, analysis technology, processing technology, display technology. Comprehensive relational and non-relational database technology application should be considered and no database technologies should be emphasized on in big data integration management techniques; Carrying out the research of big data analysis techniques should be from big data's governance and sampling, large data's feature selection, big data smaller, data classification algorithm, parallel data mining, etc.; Big data processing technology should be taken into account according to the specific requirements of the application of distributed computing, memory, calculation, flow processing technology; Data display technology in electric power of big data can consider visualization technology, space information display technology and historical flow display technology. Finally, this paper illustrates the architecture of big data platform in power system, and gives overall executable frame of big data in power system, it has the reference value to the set up of data platform for electric power enterprises.

In the field of business, big data technologies have become more widely used and have created a huge commercial value, but its application in electric power system is just started, so combining technical advantage of big data and the application requirements of power system, exerting the value of big data in power system will bring new development opportunities for the smart grid construction. Power enterprises should firmly seize the opportunity, developing power big data technology from the aspects of data policy, talent training, key technology research and development, etc.

References

1. Shilong, L.: Grid integrated electrical equipment on-line monitoring platform based on the design of data acquisition system. Power Electr. **29**(2), 29–31 (2009)
2. Dewen, W.: Center infrastructure and key technologies based on the power of cloud computing data. Autom. Electr. Power Syst. **10**(11), 146–169 (2012)
3. Bingshuai, G., Jinlin, W., Xue, L.: A business related data collection method. J. xi'an Univ. Electron. Sci. Technol. **40**(2), 66–73 (2013)
4. McKinsey & Company. Big data: the next frontier for innovation, competition, and productivity **37**(1), 1−28 (2011). McKinsey Global Institute, New York
5. Wu, X., Zhu, X., Wu, G., et al.: Datamining with big data. IEEE Trans. Knowl. Data Eng. **26**(1), 97–107 (2014)
6. Kim, B.J.: A classifier for big data. In: Proceedings of the 6th International Conference on Convergence and Hybrid Information Technology. ACM, Daejeon (2012)
7. Kwon, T.H., Kwak, J.H., Kim, K.: A study on the establishment of policies for the activation of a big data industry and prioritization of policies: lessons from Korea. Technol. Forecast. Soc. Change **33**(20), 88–93 (2015)
8. Hansen, D.: Powering business analytics with big data and real-time using data integration. Database Trends Appl. **2**(1), 17–19 (2013)
9. Abdrabou, A., Gaouda, A.M.: Understanding power system behavior through mining archived operational data considerations for packet delivery reliability over polling-based wireless networks in smart grids. Comput. Electr. Eng. **18**(20), 118–119 (2015)
10. Avram, M.G.: Advantages and challenges of adopting cloud computing from an enterprise perspective. Procedia Technol. **27**(31), 199–201 (2014)
11. Divyakant, A., Philip, B., Elisa, B., et al.: Challenges and opportunities with big data. Proc. VLDB Endowment **5**(12), 2032–2033 (2012)
12. Shin, S.-J., Woo, J., Rachuri, S.: Predictive analytics model for power consumption in manufacturing. Procedia CIRP **29**(30), 191–194 (2014)