

Research on Big Data Technology Applied in Smart Grid

Bei He^{1, a}, Fuli Yang, Ke Zheng, Gang Li, Xuedan Tao^{2, b} Wang Yao^{3, c}

¹ State Grid of Chongqing Electric Power CO. Electric Power Research Institute, Chongqing China

² State Grid of Chongqing Electric Power CO. Electric Power Research Institute, Chongqing China

^a18680887087@163.com, ^b cqepzk@sina.com

Keywords: Smart grid, Big data, Big data technology, Application

Abstract. With the big data and smart grid more and more valued, the power big data will become an inevitable trend. The power big data has a complex structure and various kinds, with the characteristics of dispersion, diversity and complexity. The traditional data processing method can not meet the requirements of smart grid development. Therefore, the application of big data technology in smart grid is very essential.

1 Introduction

The smart grid is the intelligent of the grid (smart power), also known as "Grid 2.0", which is build on an integrated, high-speed two-way communication network, based on advanced sensing and measurement technology, advanced equipment technology, advanced control methods and advanced decision to support system technology. It can improve energy efficiency, reduce the impact on the environment, improve the safety and reliability of power supply, reduce the power loss of the transmission network, not only inherit the traditional power grid, also develop the traditional power grid, more in line with the demand for modern society on the power system, which is the vision of future power development.

Big data is a widely mentioned concept in recent years, refers to a collection of data can not be captured, managed, and processed with conventional software tools for a certain period of time, which is massive, high-growth-rate's and diversified information assets, needs new processing mode in order to have a stronger decision-making ability, insight, discovery and process optimization capabilities.

With the overall construction of the intelligent, digital, information technology power grid[8], the power industry data shows a explosive growth trend. Power system in the operation will produce a huge amount of data information, data growth is very fast and data types are more. The characteristics of power data are consistent with that of big data, so the smart grid is also one of the important technical applications of big data. Since 2012, domestic and foreign universities and research institutions, IT companies, power companies have carried out intelligent big data research and engineering applications. China opened in 2013, "the first year of big data," set off a research boom on the power big data.

This paper introduces the concept of big power data, and the key technologies of big data application in smart grid are expounded from four aspects: data integration technology, data storage technology, data processing technology and data analysis technology.

2 The concept of power big data

Power big data is the practice in the power industry of the big data concept, technology and methods. Power big data relates to the various links including power generation, transmission, substation, distribution, and electricity scheduling, is cross-unit, cross-professional, cross-business data analysis and mining, and data visualization. Power big data is composed of structured data and unstructured data, with the application of smart grid construction and Internet of Things, unstructured data shows a rapid growth momentum, which will greatly exceed the number of

structured data. Power big data has the characteristics of multiple sources, spatial distribution, time multi-scale and real-time interaction. Its development prospect is broad and the related processing technology has been improved in recent years.

3 Key technology of power big data.

According to the information processing flow, the application of big data in the smart grid can be divided into six aspects: data acquisition, data cleaning, data storage and processing, data analysis, data interpretation and data application. The key technologies include data integration technology, data processing technology, data storage technology and data analysis technology [11].

3.1 data integration technology

To deal with decentralized, diverse, complex power big data, it is necessary to logically or physically organically concentrate the data of different origins, formats, and properties, to store a series of data sets which are subject-oriented, integrated, relatively stable, and reflecting the historical changes, so as to provide a comprehensive data sharing for system. At present, the commonly used data integration model includes federated, middleware model and data warehouse.

ETL (Extract-Transform-Load) is the main solution for enterprise data integration, which is the process of extracting, cleaning and converting the business system data before loading it into the data warehouse. In general, do data cleaning from the business system to ODS, filter out the non-compliant data; do data converting from ODS to DW, unify the same type of data of different business systems, carry out polymerization in accordance with the data warehouse size, and perform the calculation of some business rules; then load the converted data into the destination data source.

Data integration is a key link in the application of power big data. Power big data integration involves many applications that differ in terms of real-time requirements, data size, and data types, therefore, there is a need for a combination of various factors in power data integration. In integration technology, single technology may be difficult to achieve, combining a variety of technologies to achieve is needed.

3.2 data processing technology

Power big data has many types of applications, needs to use different data processing technology based on different business needs[9]. we divide power big data processing technology into distributed computing, memory computing, stream processing, etc.

(1) distributed computing technology

Distributed computing is a new way of calculating, studies how to divide a large computational task into multiple small computational tasks that can be distributed to a number of machines to calculate, and then summarize the results.

Google's MapReduce programming model is a typical distributed computing technology, the main idea is to split the source data into blocks, process blocks in different Map task areas, and store key values to the local hard drive, then through the program of Reduce function aggregate the results and output the final one.

Distributed technology is suitable for dealing with large-scale decentralized data sources in the field of information collection of power systems.

(2) memory computing technology

Memory computing technology is the computing technology to put all data on the inner layer to operate, which overcomes the large time consumption on disk read and write operation, the calculation speed is improved by several orders of magnitude. Memory computing can be divided into three categories: 1. distributed cache (Memcached / Redis); 2. memory data / computing grid (IMDG / IMCG, GemFire / Hazelcast / GridGain); 3. distributed memory database (IMDB, VoltDB / Impala). Memory computing technology is suitable for efficient reading of power data and processing online real-time calculations.

(3) stream processing technology

The processing mode of the stream processing treats the data set as a stream and processes it immediately when the new data arrives, and returns the corresponding result. The data flow itself has the characteristics of continuous, fast and large scale, so it is necessary to analyze the latest data as quickly as possible and get the results. Stream processing technology can handle real-time, speed-and-scale-uncontrolled data, applying it to power systems can provide immediate basis for decision-makers and meet real-time online analysis needs.

3.3 data storage technology

Power big data is from the log, historical data, user behavior records, etc, not only contains structured data, also contains more semi-structured data and unstructured data. And the traditional database stores structured data, can not meet the storage requirements of big data, so new data storage technology routes are put forward. This paper will introduce the three most typical data storage technology routes [7].

The first is the new database cluster of MPP architecture, focusing on the power industry big data, using shared nothing architecture, to achieve large-scale data storage by distributing data to multiple nodes, to improve query performance by parallel query processing. The note information exchange is achieved through the node internetwork.

The new database based on the column storage + MPP architecture can realize the storage and management of the structured data on the PB level, whose operating environment is mostly low cost PC Sever, has high performance and high scalability. And it's widely used in the enterprise analysis application field.

The second is based on Hadoop ecosystem technology extension and encapsulation, around the big data technology derived from Hadoop, to achieve semi-structured and unstructured data storage and management. At present, the most typical application is to achieve the internet big data storage analysis support through expanding and packaging Hadoop, which contains dozens of NoSQL technologies, and is further subdivided.

The third is Big Data Appliance. It is a combination of software and hardware designed for the analysis and processing of big data, which is constituted by a set of integrated servers, storage devices, operating systems, database management systems, and specially pre-installed and optimized softwares for data query, processing and analysis. High-performance Big Data Appliance has good stability and vertical scalability.

The new MPP database will be used in conjunction with the Hadoop ecosystem. On the one hand, MPP will be used to handle PB-level, high-quality structured data, and provide powerful SQL and OLTP services. On the other hand, Hadoop is used to implement semi- and unstructured data processing. The hybrid model will be the futural trend of big data storage and management .

3.4 data analysis technology

Data analysis is the core of big data processing, with the help of analytical techniques, the potential modality and law can be found from the massive data of power systems, to provide decision support for decision-makers. Many traditional small data analysis algorithms are no longer applicable in the big data environment because of the big data characteristics of large amount, large changes, complex and diverse[2], new data analysis algorithms or improving the existing algorithms are needed.

Data mining involves a lot of disciplines and methods, there are classification, clustering, association rules and machine learning methods.

(1) Classification: is to find out the common characteristics of a set of data objects in the database and classify them into different classes according to the classification pattern. Including decision tree, Bayesian, neural network and other algorithms.

(2) Cluster analysis: is to divide a no category marked sample set to several categories by a certain criterion, to make the similarity within classes as large as possible and the similarity between classes as small as possible. Including division method algorithm, level method algorithm, network method algorithm, density method algorithm and so on.

(3) Association rules: divided into two stages: the first stage is to find all the frequent itemsets from the massive raw data; the second stage is to generate association rules from these frequent itemsets. Including Apriori algorithm, FP-tree algorithm, DHC algorithm and so on .

(4) Machine learning methods: to study how computers simulate or achieve human learning behavior, in order to obtain new knowledge or skills and reorganize the existing knowledge structure to continuously improve their own performance. Including induction learning, analytic learning, analogy learning, genetic algorithm, joint learning, enhanced learning, depth study and so on.

Data mining technology is applied to the smart grid, which is helpful to discover the useful information in the power data, so as to further discover the higher value through data analysis. It plays an important role in dynamic safety assessment of power system, load economic dispatch, load forecasting and user feature extraction, fault diagnosis and so on.

4 Conclusion

This paper briefly analyzes the current development of smart grid and big data, pointing out that the development of power big data is an inevitable trend, elaborates the four key power big data technologies, data integration technology, data processing technology, data storage technology and data analysis technology, and their applications in power systems. It can be seen that big data technology in smart grid will play an increasingly important role with the construction of smart grid more and more perfect, so combining big data technology advantages and power system application requirements, playing the value of power big data, will become the trend of future power grid development. At present, the smart grid big data is still in its infancy, we should continue to study and fully promote the development of power big data technology.

References

- [1] ZHANG Genzhou. Application of big data in the field of smart grid [J]. Grid and clean energy, 2016, 32(6):114-117(in Chinese).
- [2] CHEN Xingying, ZHANG Xiaohua, QU Feng, etc. Application of Data Mining in Power Systems[J]. Journal of Electric Power Science and Technology, 2007, 22(3):51-56(in Chinese).
- [3] TAO Yuan. Application of Big Data in Smart Grid [J]. Guangxi Electric Power Industry, 2015(3):14-15(in Chinese).
- [4] Katal A, Wazid M, Goudar RH. Big data: issues, tric Power Construction,2015,36(10):1 1-19· challenges.tools and good praectices[C]//2013 Sixth International Conference on Contemporary Computing.Noida,India:IEEE,2013:404-409.
- [5] LI Junhui. Research on application of big data in smart grid [J]. Construction Engineering Technology & Design, 2017(5) (in Chinese).
- [6] SONG Jing, TANG Jing, XIAO Feng. The development status and analysis of smart grid at Home and Abroad[J]. Electrotechnics Electric, 2010(3):1-4(in Chinese).
- [7] Three of the most typical big data storage technology routes[EB/OL]. <http://www.cnblogs.com/liangxiaofeng/p/5166795.html>(in Chinese).
- [8] CAI Defu, CAO Kan, TANG Zeyang, etc. Application of big data technology in power systems[J]. Hubei Electric Power,2006, 40(6):22-26(in Chinese).
- [9] PENG Xiaosheng, DENG Diyuan, CHENG Shijie, etc. Key technology of power big data for smart grid applications[J]. Proceedings of the CSEE, 2015, 35(3):503-511(in Chinese).
- [10]ZHANG Dongxia, MIAO Xin, LIU Liping, etc. Research on development strategy for smart

grid big data[J].Proceedings of the CSEE,2015,35(1):2-12(in Chinese).

- [11]MENG Xiangjun, JI Zhixiang, YANG Yi. Research on big data platform and its key technologies in smart grid[J].Distribution&Utilization,2015(8):19-24(in Chinese).