

Research and Design of Data Storage Scheme for Electric Power Big Data

Wenfeng Song^{1,a}, Wanqing Yang^{2,b*}, Jingzhao Luan², Xuepu Jiang², Yuedong Wang²

¹No.102, Zhongshan Road, State Grid Dalian Electric Power Supply Company, Dalian 116001, China

²No.200-7, Shengli East Road, State Grid Dalian Electric Power Supply Company Information & Telecommunication Branch, Dalian 116011, China

^aswf_dl@ln.sgcc.com.cn, ^bywq_dl@ln.sgcc.com.cn, yangchaoneu@sina.com

Keywords: Big data; Electric power data; Data storage; Storage architecture

Abstract. With the continued construction of information and smart grid for electric power sector, they are entering a big data era. The most immediate priority is to solve the big data storage for the power industry under massive amounts of data. The purpose of big data storage for electric power sector is not simple to show different patterns of data but also to provide guarantee for big data mining. The data storage models for electric power big data include the data model and the storage model. The paper states some optimal schemes for data storage in order to manage the electric power big data. At the same time, to improve the big data storage scheme further, some key techniques of index structure are introduced for electric power big data. The scheme for electric power big data storage and management is studied. Through combining big data with the electric power sector, storage architecture of power industry is designed.

Introduction

For the power industry, the data produced in each link related to the power system, such as operation conditions of the equipment, data of voltage stability, electric power enterprise operation and management, electricity price, the users and so on, together constitute the Electric Power Big Data, which can be simply summarized as the data generation side, transmission side, distribution side and electricity side.

From the power generation side, with the continuous development of the entire power industry, the demand for big data is also greatly demonstrated. First is the digital information construction of power plant [1]. The emergence of a large number of new energy sources produced in the electricity production process, such as wind and light, has changed the previous data of generation side which are relatively "static" when the energy structure dominated by thermal power, and the numerous data of production operation status, fault diagnosis, or the loss report have been preserved. So, fault diagnosis method based on data driven is gradually taken into the information construction of power plant [2], which uses historical data of power plants to solve the problems of production process control, fault diagnosis, optimize production and regulation of machine operation. In addition, as for the cooperation between different power plants, namely the matters of the operation condition monitoring, equipment management and control of the distributed power equipment, the electric power big data also proposed a better solution [3].

On the transmission side, with the smart grid construction carrying forward step by step which the State Grid Corporation adhering to, especially the pilot construction and promotion of the transmission condition monitoring system, the user information acquisition system, the distribution automation system and so on, together with the dispatch of power network operation mode, gateway power, protection, lightning and other systems, result in a large amount of real-time data and then precipitate to form the massive historical data. Taking the The-SCADA system of power grid "nervous system" of substation induction monitoring data for

example, a 66kV transformer substation has about two units, a transformer has twenty intervals and need to monitor the voltage, current, frequency and temperature index in a real time, measurement points in each interval of a standard, according to the sampling interval of 3 or 4 seconds, 108.27 MB of data will be generated every year ($1.03 \text{ TB} = 12 \text{ bytes / frame} * 0.3 \text{ frame / s} * 1 \text{ measurement points} * 86400 \text{ s / days} * 365 \text{ days}$); wide area measurement system (WAMS) has 1 measurement point and the sampling rate can reach 100 /s, according to the formula above, 52.03 GB is the annual data, a 66KV substation will produce more than 1TB data a year. With the China State Grid Corp as an example, the number of 66kV substation can reach to tens of thousands, the amount of data generated will reach several PBs, and the power system of 66kV substation is just one part of the whole grid system, there are 200kv、500kv、800kv and even much more larger 10kV/380V area with more intervals and so on. The pharos monitoring is part of the power grid monitoring, power grid is part of the power grid business, and its data is also very large [1].

On the distribution side, the electricity status of each region is collected, stored, analyzed, so that a lot of data will be generated. For example, the electricity distribution company requires the equipment provided with the function of communication and electricity meter with a certain requirement in order to obtain the user's data. These electricity meters send electricity real-time information at a frequency of once 5 minutes, there will be over 5TB data to a city of 10 million people produced in a month.

And for the electricity side, in the link of power operating, with the continuous development of the electric power system, highly flexible data driven power supply chain will gradually replace the traditional static electricity supply chain. Along with the payment methods for electricity costs of the network users getting more and more diverse, the Department in charge of the sale of electricity not only should store the data of users, but also should store the data of related payment information, a large amount of data thus produced.

Electric Power Big Data

Categories of Electric Power Big Data. Electric power big data are generated in various aspects of the power system. If according to the content of the data, it can be divided into 3 categories: monitoring data of power plant equipment or operation of the power grid; the second is the data of enterprise business and sales, such as the data of sales volume of electricity, price, user information power consumption and etc.; the third is the data of management of electric power enterprises and administration [4]. According to the data storage structure, it can be divided into structural data and unstructured data [5].

The structural data mainly refers to the data stored in the relational database, which can express the data with the table structure. Because of its convenient query management, a lot of data are structured data at present, such as the log table of operation condition of generation side machine and transmission side voltage, the regional electrical record of distribution side and the table of users' consumption records of electricity side.

The unstructured data mainly refers to the data that cannot adapt to changes of the length of the field stored in the relational database, including text information, pictures, sound, video, hypermedia and other information. In the electric power big data, each part of the surveillance video, format information of exchanged audio and picture are unstructured data. The massive manual recording text log data produced in the maintenance records of transmission, substation and distribution operations and equipment line in the power grid enterprises; the image data and spatial-temporal correlation data to record the condition at that time in power transmission and distribution operations and equipment line maintenance, fault repair work; surveillance video of producing, working, operations and sales locations in the power grid enterprise and equipment operation, all of these belong to the category of unstructured data [6].

Aims of Data Storage for Electric Power Big Data. The storage of electric power big data is the basis of analyzing and using, the data storage goals requires not only can present all forms of information including report, curve analysis map, columnar analysis chart, pie chart and so on, but also can provide the data base and

support for integrated query analysis tools of electric power data which combined data acquisition, data mining, data display, data management, data alarm, data sharing, and data security.

It will effectively integrate the existing resource; improve the use of enterprise production data, standardized production and information communication to realize the storage of electric power big data. It will integrate enterprise producing information data in a vivid, intuitive and visual way, thus further combining business management process, standardizing business operation process reasonably, and improving production and operation procedure of the enterprise, which makes the enterprise resources to the most reasonable configuration and the most reasonable application. It is undoubtedly that these realizations will greatly reduce the cost of enterprise management, the cost of human resources in enterprises and the production cost, improve the working quality and efficiency of enterprise and will support the development strategy of establishing the modern enterprise, which has a good economic benefit and management value.

Data Storage Scheme for Electric Power Big Data

Data Models. Data model of electric power big data includes associated data model, spatial geographic information model, time series model, structure data model and unstructured data model of text, picture, video and etc. Such as accident rush repair operation monitoring, every piece of data including the time attribute, geographical spatial property, large amounts of data coherence are described the accident to grab the car running track. In the GIS platform of power grid, power transmission lines and towers are stored in static coordinates of longitude and latitude. Taking the national power grid SG-186 marketing system as an example, each data records the user's payment, the user's payment is also stored in the database in accordance with the time series (month). To set the State Grid Corporation SG-ERP system equipment assets management module as an example, each data undertakes the management of primary key and foreign key, which means that equipment assets associated with the power company's property, and the power supply company associated with the property of the provincial power grid company [7]. The text data model is a large amount of text data recorded in the transmission specialty, distribution specialty, substation professional operation and maintenance record in the power network enterprise. Picture data model is embodied in the transmission, transmission and distribution, electrical professional photographs, equipment ledger photos. The video data model is a large number of office, business and equipment monitoring images in the power grid enterprise's office premises, business network, and production site.

Storage Models. The storage model is also called the physical model, which emphasized the structure of the system.

Data Partition Storage Model Based on Business Partition:

The electric power enterprise data can be divided into core business system data, important business system data, and general business system data in accordance with the business processing of data types, processing method, the frequency of traffic and system parameters [7,8].

Core business systems data refers to all data in the electric power enterprises and power business which involves a lot of system and gateway. These data records and monitors whether the electric power enterprise system operates normally, including business data of power plant productive management, business data of grid power transmission operative management, business data of financial management and management data of electric power project. The loss of the data will not only affect the normal operation of the enterprise itself, and even cause the cessation of all the related areas, which will bring immeasurable loss.

Important business system data refers to the data that is only related to electric power enterprises, the loss of these data may affect the company's normal operation, but it is easy to return to normal, its influence is far less than the loss of the core business data. These data include data on power marketing management, human resource management and material management.

General business system data refers to the system data not too related the enterprises' business, the loss of these data will also has certain effect on the normal operation of the system, but the degree is very low and the duration even damage degree has very high tolerance, including some data of coordination office business and administration business.

The advantages of data storage based on business partition for power data partitioning are highlighting the points, providing higher protection to the core business data, and controlling the loss in the controllable range.

Distributed Storage Model:

Distributed storage model is a kind of data-centered storage method, the design idea is to distribute the data storage to independent equipment and through the interconnection network, using multiple servers to share storage load to improve the reliability, availability and access efficiency of the system, which is easy to expand [4].

A distributed storage structure is usually composed of a master node and several data nodes. Master node is used to manage the storage space and adjust the main server of the access file; the data node is responsible for managing the data within the nodes.

Design of the Storage Architecture of Electric Power Big Data

The power data storage architecture is compose of the query layer, the translation, executive engine layer, data storage layer and resource management layer. The operation procedure is transferred from the top down to the bottom layer, and reaches the bottom device through compilation, execution, and the data management. This paper is mainly concerned with the two lower layers, that is, data storage and management. [9, 10].

The Query Layer. The query layer mainly receives query request of power data from the user or the system, namely through the SQL statement or the complex inquiry interface, calls the related statement to carry on the operation. Data analysis tool set is mainly used for the existing data for the analysis of the operation, such as drawing the equipment operation indicators of the trend picture for a certain period of and analyzing the health status of equipment; presenting an electricity trend picture in a certain area and analyzing power distribution in the next time period.

The Translation Layer. The function of this layer is to translate the operating requests received from the upper level to the machine language through the compiler or the parser, and pass it to the next layer. The implementation of the program generator is used for the data stored in the distributed memory referred to above, specifying the relevant parts of the data storage when the operation involves data distributed on different servers. The optimizer is the component to optimize on the condition that the query is not very efficient and further enhance the query efficiency. Cache management is used for performance improvement based on the caching as above mentioned.

Executive Engine Layer. Transaction management is mainly for crud operation and data management, analysis engine service is used for electric power data query processing and related information mining, parallel execution engine is to divide an operation into several independent operations of parallel execution, summing up the results after the execution completed to accelerate query speed.

Data Storage Layer. Data storage layer is a physical model facing to data storage. Data partition management is management for data storage model, divided into horizontal partition management and vertical partition management to shorten the response time by adapting to the operation of accessing to a record or different operation of the same attribute records. Log management is the management of the database operation file for the record transaction. Restore the backups is used to recover data when data disaster happens. The existence of multi-level index is also to accelerate the operation.

Resource Management Layer. Resource management layer inclines to the management of the underlying device that relies on data storage. The management of communication service is the management of

communication between the servers which storing the power data and provides the guarantee for the normal use of data. The distributed service framework is mainly to provide good frame service for the distributed storage and cloud storage. Multi-tenant management is a kind of technology of software architecture based on multi-tenant technology to realize using the same system or software programs under the multi-user environment and ensure isolation of data between users at the same time.

Summary

Resource management layer inclines to the management of the underlying device that relies on data storage. The management of communication service is the management of communication between the servers which storing the power data and provides the guarantee for the normal use of data. The distributed service framework is mainly to provide good frame service for the distributed storage and cloud storage. Multi-tenant management is a kind of technology of software architecture based on multi-tenant technology to realize using the same system or software programs under the multi-user environment and ensure isolation of data between users at the same time.

Acknowledgement

The authors would like to thank Ning Bo, Yang Chao for their scientific collaboration in this research work. This work is supported partly by the Fundamental Research Funds for the State Grid Liaoning Electric Power Supply Co. Ltd. (Grant No.2014YF-29, Grant No.2015YF-67).

References

- [1] Shi Jie, Ai Jian. The smart grid implementation of some key technology problems research [J]. Power System Protection and Control, 2009, (19):1-4.
- [2] Zhang Hongmei. Mass data storage method in smart grid research [D]. Liaoning University, 2012.
- [3] Song Baoyao. Support large-scale smart grid data storage solution [J]. Computer Application, 2012.
- [4] Yang Guangzhen. Zhongshan electric power data storage system solutions [J]. Electric Power Information, 2004, (5).
- [5] Song Xiaofang, Hu Chaoyang, Xue Feng etc. After large-scale wind power connected to the smart grid adaptive scheduling technology support system research [J]. East China Electric Power, 2011, (9).
- [6] Li Ming. Intelligent data collection method and the implementation in the power system [D]. Northeast Forestry University, 2006.
- [7] Zhang Wenyi, Xiang Lianzhi, Wang Xiaofang. Big data partition management model and its application research [J]. Journal of Harbin Engineering University, 2014.
- [8] Wang Mingjun. Smart grid hot issues discussed in this paper [J]. Power System Technology, 2009, (18).
- [9] Xiao Han, Hu Guangping. Large amount of data storage structure of the database design research [J]. Group Technology & Production Modernization, 2011, 28(2).
- [10] Song Y, Zhou G, Zhu Y. Present status and challenges of big data processing in smart grid [J]. Power System Technology, 2013, 37(4): 927-935.