Application and Analysis of Big Data Technology in Smart Grid

View the article online for updates and enhancements.

Application and Analysis of Big Data Technology in Smart Grid

Shuchun Wang[1*], Xiaoguang Sun[1], Jianyu Geng[1], Yuan Han[1], Chunyong Zhang[1], and Weihua Zhang[1]

[1]State Grid Jilin Electric Power Company Limited Changchun Power Supply Company, Changchun, Jilin, 1300222, China

[*]Corresponding author's e-mail: 741983881@qq.com

Abstract. In order to make the modern power system more intelligent, various data acquisition equipment and information management systems are required, such as smart meters, remote terminal units (RTU), phasor measurement units(PMU), power distribution management systems(DMS), and energy management systems(EMS), customer management system(CMS), power generation management system(GMS) have been widely installed. These systems create large amounts of data, which is the main source of smart grid big data. Researching advanced data analysis technology is the most urgent task for the development of smart grid. It can make the best decision for the control and operation of the grid, and can improve the flexibility, safety, reliability and efficiency of the power system. This article analyses the cause of smart grid big data and the importance of data analysis, and discusses the development trend of big data technology. Selectively analysed the source and type of smart grid big data, summarized the characteristics of smart grid big data, gave various application scenarios, and introduced the smart grid big data analysis platform and the functions being developed.

1. Introduction

Smart grid is the direction and trend of power industry development. The smart grid is based on the physical grid and combines advanced information & communication technology with the operation technology of the physical power system to form a new generation of grid. It covers the six major links of generation, transmission, transformation, distribution, utilization and dispatching. It coordinates the needs and functions of the various participants in the power system and enables the system to operate on the premise of ensuring the reliability and stability of the entire system. More efficient, more economical, and more environmentally friendly. The smart grid is built on the basis of an integrated, high-speed two-way communication network, making full use of advanced sensor measurement technology, optimized control technology and decision support technology to achieve the goal of reliable, safe, economical, efficient and environmentally friendly power grid[1]. Strength, self-healing, compatibility, economy, integration and optimization are the main features of smart grids. The "intelligence" of the smart grid is based on a high degree of "observable" and "controllable". The basis of observation and control is to obtain panoramic real-time data that can truly reflect the operating state of the system and process the data quickly and analysis, transform the data into information as soon as possible, make judgments and predictions based on the information, and then convert the information into decision-making that can guide operation control[2-3].

In order to achieve the "observable" and "controllable" goals of the smart grid, numerous data collection and information management systems have been installed and deployed in the six links of the smart grid's transmission, transmission, transformation, distribution, use, and dispatch management.

The construction of these data collection and information management systems, especially the large-scale deployment of smart meters, and the widespread use of sensing and measurement technologies such as synchronous measurement units (PMUs) have led to the generation of large amounts of data in smart grids. The research results show that the amount of data generated in the smart grid is 4 orders of magnitude higher than that of the traditional grid. These data are diverse in structure, complex in source, and have typical "4V" characteristics, namely large volume, variety, low value and fast velocity. How to use these data to provide scientific decision-making for the development and operation control of the power grid is not only an urgent need for the development of smart grids, but also the only way to achieve the strength, self-healing, compatibility, economy, integration and optimization of smart grids.

This article first discusses the development trend of big data technology, and then focuses on the source, type, resolution and demand for communication channels of smart grid big data, and then summarizes the characteristics of smart grid big data, analyses the application scenarios of smart grid big data, introduces the technology platform and main functions of the smart grid big data analysis system being developed.

2. The development trend of big data technology

Big data technology is divided into two parts: general big data technology and professional big data application. General big data technologies include data processing, data storage, data analysis, and data display, etc., which are generally developed by professional information technology companies. Smart grid big data applications are professional big data applications that need to be jointly developed by professionals familiar with smart grid business and knowledge of smart grids and computer professionals who are proficient in general big data technology. For professional developers, universal big data technology provides a basic development platform for big data.

General big data technology presents the development trend of open source sharing replacing closed-door development and consulting services replacing software licensing. The open source software adopts the model of public development and free sharing, which saves the capital cost of startups, greatly reduces the threshold for entrepreneurship, and also enables innovative technology to be popularized as soon as possible. It plays a role in encouraging innovation for the whole people and is a typical manifestation of Internet thinking.

Open source software has become the mainstream trend of big data technology development. Taking database products as an example, according to incomplete statistics, more than 80% of startups established in the Silicon Valley region of the United States after 2000 use open source databases, even Google, Facebook, and other big companies. Companies also use open source software. Of course, these companies have also formed a strong open source software development team to participate in the development of open source software.

To a certain extent, the popularization and wide application of open source software are important factors for the rapid development of big data technology. In order to develop big data applications suitable for professional needs, there must be extensive participation of professional and technical personnel. First of all, big data applications must dig deep into professional scenarios to solve the problems that are difficult to solve because of the huge data scale and complex data types using the current traditional data processing technology, or the current data of a single business system is difficult to solve, and multi-source data must be merged Problems that the system can solve. Such business needs must be combined with the areas where big data technology is good, and closely combined with the actual needs of the business area to get it.

Secondly, after knowing the professional needs, a professional data model must be established. Different professional fields have different professional characteristics, and the data scale, type, structure, and mutual relationship are different. It is necessary to combine the professional characteristics and establish a professional data model that meets the professional characteristics. For

example, the data model of the retail industry is different from the data model of the banking industry, the data model of the power grid and the data model of the communications are not the same, and the data model of the aerospace and the aerospace cannot be the same.

Finally, we must select effective big data processing, storage, analysis, and display technologies based on professional needs and established data models to develop professional big data applications.

In order to develop a suitable big data application, one must understand the areas where big data technology is good.

Big data technology replaces sampling analysis with full data analysis. In the era of big data, the rapid development of computer hardware and software has enabled us to process more data, and sometimes even all relevant data, instead of relying on random sampling as before. Compared with the analysis of random sampling data, the use of full data can discover the details that could not be found before, so that we can see more clearly the details that cannot be explained by relying on random samples. For example, the big data project of the French power company, adhering to the concept of data asset management, integrates data resources such as smart meter data, power contract data, and grid structure data into an enterprise that can fully reflect the power company assets, users, and grid operation status. The big database regards it as an important asset management object, and establishes a professional management organization for data assets to transform the past disordered data resources into valuable data assets. With regard to such full data that can fully reflect the operating status of power companies, with the help of big data analysis and processing technology, market consumer groups can be accurately distinguished and positioned from multiple angles such as market positioning and consumer behaviour, so as to expand business areas and innovation for enterprises The profit model and the promotion of enterprise transformation and upgrading provide effective decision support and realize the value increase of data assets. The big data project of the French power company is a typical case of discovering the value of data assets based on full data analysis.

Big data is good at regularity analysis. Big data is good at discovering potential internal laws from numerous massive data with complex relationships, which is suitable for exploring the internal development trend of things, and to carry out early warning and prediction accordingly. The anti-theft software developed by the American C3 company is built on a multi-source big data platform that integrates smart meters, user information, work management, power outage management, geographic information, and meter manufacturer data. Through analysis of nearly 90 indicators, Machine learning methods quickly discover the laws of stealing electricity and identify possible acts of stealing electricity. This system has been put into use in the Baltimore Power Company in the United States, and the identification rate of theft of electricity has been greatly improved. In the first 6 months after the investment, more than 8,000 cases of theft of electricity were discovered, with a recognition rate of up to 90%. It also found 3,600 quality problems with the meter, with a recognition rate of up to 99%, creating huge economic benefits for the enterprise. As another example, in the past, it was often assumed that the load shape of the user was a double-peak type. Stanford University conducted a cluster analysis on the data of PG&E Power Company's smart meters and found that the double-peak load accounted for only 20%. This finding is of guiding significance for the accurate prediction of users, feeders and systems.

Big data emphasizes cross-domain analysis and is good at discovering interconnections between seemingly unrelated data sets. For example, the Italian Trentino Power Company based on a high-order Hilbert space mathematical model, using a random prediction algorithm, established a non-linear regression relationship between mobile communication data and grid power demand and crest, successfully predicting the grid within the next week Daily electricity demand and power crest. Its characteristic is that the model has less state space dimensions, so it can effectively use big data technology for fast and accurate analysis. This application successfully discovered the potential connection between mobile communication and power load data, and established the interrelationship between human behaviour data, mobile communication data and power consumption data. At first glance, these two types of data seem to have little relationship, but upon careful analysis, there is indeed an inseparable relationship between the two. This is mainly due to the popularity of smartphones. The

use frequency, location and method of smartphones deeply reflect human work behaviour and lifestyle, and are important data describing human behaviour. Human behaviour inevitably affects electricity demand, and it is not surprising that there is a potential relationship between the two. This point can also be corroborated by the relationship between the rural New Year's holiday of the rural population in my country and the increase in rural power load. The quantitative description of migrant workers returning home for the New Year can be obtained with the help of mobile phone movement laws.

Big data is good at fast and real-time analysis. Big data analysis and processing technologies represented by parallel computing, memory computing, supercomputing, etc. can effectively solve real-time analysis and decision-making problems. The most typical examples are the authorization processing of bank credit cards, the order processing of e-commerce companies, and the real-time and accurate recommendation of social media websites all apply the fast-real-time processing capability of big data. In the future, the smart grid dispatching operation control system based on big data will also be a place where big data can be processed in real time.

3. Multi-layer application of smart grid big data

The business value of smart grid big data analysis can be divided into the following four categories according to different service objects[4].

- Improve power grid intelligence: improve power grid operation reliability and improve grid planning accuracy.
- Improve asset intelligence: Improve asset utilization.
- Improve user intelligence: Understand users' electricity consumption behaviours, innovate service models, and provide users with customized services.
- Improve social intelligence: Provide quantitative services for energy saving and emission reduction in the whole society.

The smart grid big data analysis can be divided into the following four categories according to the correlation between the data generation time and utilization time.

- Descriptive analysis: analysis of historical data generated in the power grid, generally used to analyse what actually happened in the past, and generally used to generate various forms of reports.
- Diagnostic analysis: Online analysis of real-time data in the power grid to analyse why it happened.
- Predictive analysis: Analyse real-time data and historical data in the power grid to predict what will happen.
- Prospective analysis: Analyse real-time and historical data in the power grid, and predict possible future scenarios, analyse what should happen, how to make it happen, and the best countermeasures that should be taken when it happens.

The development of a successful big data analysis application is generally divided into the following three steps: definition, verification and delivery[5]. The process is shown in figure 1.

The definition phase is divided into 5 steps: 1) determine the analysis requirements; 2) define the output results; 3) determine the data processing and analysis control points; 4) determine the type of analysis required; 5) determine the degree of automation.

The verification phase is also divided into 5 steps: 1) define hypotheses; 2) determine the required data set; 3) determine the conditions, standards and scenarios; 4) prototype development and testing; 5) determine whether to meet the design requirements.

The delivery phase is also divided into 5 steps: 1) define operating rules, requirements and technical design specifications; 2) define control and error management methods; 3) establish industrial-grade solutions for testing, implementation and implementation; 4) implementation and verification analysis results; 5) improve and optimize the analysis process.
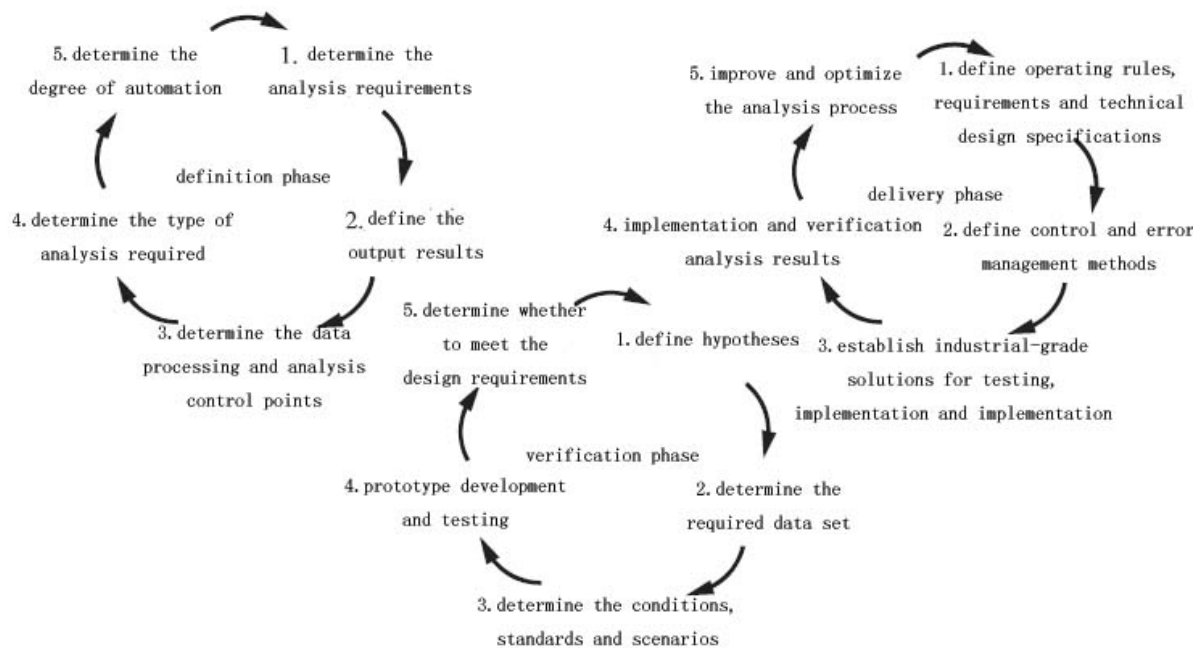
Figure 1. Definition, verification and delivery of big data analysis.

This data analysis development process is undoubtedly also applicable to the development process of smart grid big data analysis applications. The above process is indispensable in order to develop a suitable smart grid big data analysis application.

Duke Energy Corporation of the United States is using this method for smart grid big data research and development. So far, the first step has been completed and the second step is in progress. The following combines Duke Energy's report (Data Modelling and Analytics Initiative) to explain the process and results of Duke Energy's determination of smart grid analysis applications. Their specific practices have a good reference for the development of smart grid big data analysis applications in China.

First, in 2013, Duke Energy constructed an AMI system, transformers, distribution network sensors, distribution network equipment, blackout management systems, smart grid communication nodes, weather stations, billing systems, and society A dataset consisting of economic data. The data set is compiled based on the one-week data during the 2012 summer peak load period. This set of data includes 18 different data files in the format of text files(CSV) or spreadsheet data(XLS), and provides data element definitions and data mapping files. These documents are provided to 28 manufacturers in the field of big data in the United States, and these manufacturers are invited to provide feedback on possible data analysis application scenarios.

17 manufacturers returned the survey report as required. More than 150 application scenarios have been proposed, which can be divided into 12 types respectively[6]:

- Meter data analysis: meter event analysis, meter sample data error analysis, meter operation monitoring.
- User analysis: Use socio-economic data to group participants of the power company's demand-side response projects, carry out user load pattern analysis, and develop electricity price structures based on load levels and voltage levels.
- Distribution network analysis: feeder/phase load prediction, monitoring and management, equipment identification and management of voltage fluctuations on the feeder, real-time condition-based power corridor vegetation management.

- Power outage analysis: Use rapid power restoration to reduce the average number of system outages (SAIFI) and average system outage time(SAIDI), issue outage notifications and restore automation of power supply document processing, identify and handle short-term outage events.
- Income protection: use abnormal data analysis to realize income protection alarm, analyse and estimate electricity costs and electricity prices, track and recover losses due to electricity theft, and use billing data and economic information to predict illegal behaviour.
- Demand-side response: Develop demand-side response impact evaluation indicators, tap potential DR participants, and use real-time meter data to predict the amount of system load reduction due to the use of DR projects.
- Asset management analysis: Implement condition-based asset management for power distribution equipment, identify and replace overloaded transformers, and identify and monitor equipment loads and abnormal events.

4. Smart grid big data analysis

The commercial value of each type of data is different. The important thing is that the power grid company must fully understand the different commercial values contained in these different types of data, and use the appropriate technology to fully tap to meet the operational needs of the grid company. Figure 2 is the smart grid big data analysis platform developed in this paper based on the summary analysis[7].
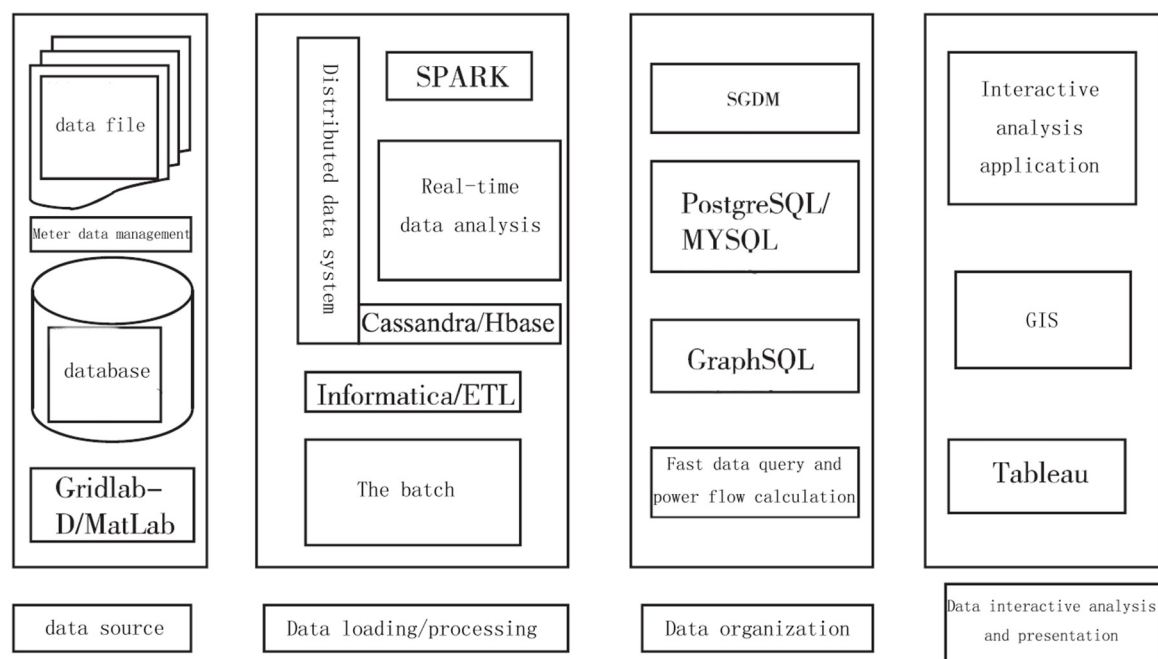

Figure 2. Big data analysis platform of smart grid.

The entire analysis system is divided into 3 parts: data loading and processing, data organization and storage, data interactive analysis and display.

In terms of data sources, smart meters, distribution automation/SCADA, network structure and parameters, meteorological data, etc. are currently considered.

4.1. Data loading and processing

Various types of data first enter the data loading/processing module, and the data is pre-processed and cleaned for the next analysis and calculation. This module includes a distributed Hadoop file system (HDFS), a cluster computing system based on in-memory computing Spark, and ETL tools. This module

can perform batch processing of computing tasks or process real-time streaming data. HDFS is a distributed file system with high fault tolerance and is designed to be deployed on inexpensive hardware. It provides high throughput to access application data, is suitable for applications with very large data sets, and provides standard streaming access (write once, read many times). Spark is a distributed parallel data processing framework that has developed rapidly in recent years. It can be used in conjunction with Hadoop to enhance the performance of Hadoop. Spark adds more advanced data processing capabilities such as memory caching, streaming data processing, graph data processing, machine learning, and SQL. Spark provides a comprehensive and unified framework for managing the needs of big data processing of various data sets and data sources (batch data or real-time streaming data) with different properties (text data, chart data, etc.).

While the data is being loaded, the memory-based cluster computing system Spark is used to analyse and process the data. The data from smart meters include voltage, current, active power, power factor, accumulated power, etc. The original data often generates abnormal data due to communication channel failure, smart meter failure, or other reasons. Therefore, during the data loading process, the original data needs to be pre-processed to find missing data and decide whether to use linear interpolation and other methods to supplement the data or delete a certain part of the missing data. Smart meters also record the important characteristics of distribution network operation. Therefore, based on smart meter data analysis, the voltage qualification rate, reactive power compensation quality, power supply reliability, and three-phase unbalance of the power grid can be refined for analysis and evaluation.

4.2. Data organization and storage

The historical data after data loading and processing is stored in the data warehouse; the processed stream data can be stored in the NoSQL database. This article uses PostgreSQL as a data warehouse, which is characterized by high reliability, supporting concurrent operations and mission-critical applications, open source, configurable, and high security. Cassandra/HBase is adopted as the NoSQL database. The main reason is that it is open source, supports large-scale distribution, is easy to expand, and is the most feature-rich in a non-relational database.

This article also developed the smart grid unified data model SGDM, which integrates data from multiple data sources according to asset information, network topology model, user information, and external data according to the top-down design concept, covering network operations and accounts Management, asset management, customer management, power outage management, work management, and weather models, and other business areas, combined with smart grid big data analysis and mining needs including KPIs, to support the operation and development of power companies, analyse user behaviour and service optimization, It provides decision-making consultation and support for three types of applications for the government and society. It has designed 48 business domains, defined logical models including entities and their relationships, and optimized physical models according to the third normal form principle.

Due to the natural physical shape of the power grid, the graphical model is an intuitive choice. A graph is a collection of nodes and edges, used to describe the relationship between nodes. Broadly speaking, graphs are ubiquitous in the real world; graph computing is of great value to the complex business environment and potential breakthrough technical solutions that are highly linked to big data today. Research on the power network as a graph has a long history of research. Through the application of the data model of the graph, power system analysis and complex data query can be combined for rapid analysis. Graph data management and computing technology are becoming one of the cornerstones of big data analysis.

4.3. Interactive analysis and visual display of data

In terms of interactive data analysis, using adaptive K-means algorithm, K-subspace analysis algorithm, shape clustering algorithm and other data mining technologies, we have developed electricity consumption behaviour analysis based on smart meter data, static load characteristic calculation, mid-to-long term electricity consumption With functions such as model evaluation and sensitivity analysis,

software such as demand response strategy evaluation and precise marketing strategy formulation has been developed.

The vast majority of resources in the smart grid are directly related to spatial location. Geographic Information System (GIS) has become an indispensable technology for building digital power grids, and has been widely used in the power industry. Develop a visual display of data cantered on geographic information maps, and closely integrate smart grid data with geographic information data, which can better achieve unified management and comprehensive display of various power grid resources and grid structures, which can help operation managers to understand the meaning of smart grid data expression more intuitively and accurately, and to understand the operating status of the system. Data visualization mainly relies on graphical means to clearly and effectively convey and communicate information.

This article selects Tableau visualization software, which has the following characteristics: high efficiency and smart functional features; beautiful operation view; convenient multiple data source connection function; perfect data integration function; online rapid data analysis function; support for geography Information charts can be used to create interactive visual human-machine interfaces and support interactive real-time display and analysis.

5. Conclusion

This paper analyses the background of smart grid big data generation and the necessity of studying smart grid big data analysis technology, discusses the development trend of big data technology, and focuses on the sources and types of smart grid big data, and then summarizes smart grid big data Characteristics, the system analyses the application scenarios of smart grid big data, and finally introduces the technical architecture and main functions of the smart grid big data analysis system being developed.

References

[1] Zhang P, Wu X, He J. (2015) Review on big data technology applied in active distribution network. Electric Power Construction, 36(1): 52-59.

[2] Yang W, Xu J. (2010) New thoughts of urban power network planning which building the urban power network structure system. In: 2010 International Conference on Mechanic Automation and Control Engineering. Wuhan. pp. 2966-2969.

[3] Yang H, Li G, Han J. (2009) Urban distribution network planning and analysis of existing problems. Inner Mongolia Electric Power, 27(5): 1-5.

[4] Lin C, Gang J, Hao T. (2010) A distribution network planning method based on the coordination of reliability and economy. Power System Technology, 11: 106-110.

[5] Wang C, Li P. (2010) Development and challenge of distributed generation, micro network and intelligent distribution network. Automation of Electric Power Systems, 34(02): 10-14+23.

[6] Liu J, Zhao S, Zhang X. (2012) The development of distribution automation in China and some suggestions. Automation of Electric Power Systems, 36(19): 6-10+21.

[7] Lu T, Gao C. (2014) A general model for optimal scheduling of battery charging and renewal network. Power System Technology, 38(10): 2700-2707.