Big Data Analysis Supports the Research and Pilot Application of Distribution Network Planning Technology

Jing Chen¹,*, Weiwei Shi¹, Ye Liu², Quan Chen², Pingfa Huang¹, Jiaying Li¹

¹State Grid Info-Telecom Great Power Science and Technology Co., LTD, Fujian, fuzhou 350003, China
²STATE Grid Jiangsu Electric Power co, LTD, JiangSu NanJing,210000

*Corresponding author email: jingchen@sgcc.com.cn

Abstract. The construction of the enterprise data center has met the automatic access of the massive data generated by the operation and management of the distribution network, and realized the integration and unification of the professional data of dispatching, operation inspection, and marketing, providing big data technology to support the planning and management of the distribution network. The application of new technical conditions such as broadband carrier has greatly improved the authenticity, reliability and timeliness of big data in the distribution network. Through the design of distribution network monitoring and early warning, diagnosis management, quality assessment, portrait research, power outage optimization, load forecasting and other intelligent application models, the accurate processing of distribution data is realized, and the state of the distribution network can be accurately evaluated. Provide technical support for grid structure optimization. Based on this research background, papers, articles have studied and proposed big data-based distribution network planning, planning auxiliary decision-making system. Based on unified technical principles, a unified planning platform is adopted to realize unified project construction. Practice shows that the plan, the auxiliary decision-making system has improved the quality of planning scheme preparation, improved the review effect, and raised the lean level of investment decision-making.

Keywords: Big data analysis. Distribution network planning, enterprise data center. Assistant Decision System.

1. Introduction
With the continuous improvement of social science and technology, the development of big data technology is becoming more and more mature, and it is widely used in various industries. On the other hand, with the large-scale promotion and use of new energy, the number of grid users is increasing day by day, and the grid structure is more complex. In order to scientifically plan the grid layout, improve the reliability, economy, and foresight of the distribution network, and enhance the level of distribution network planning and management, more and more power companies have introduced big data to analyze the characteristics of the grid load, and build load models to accurately Calculate the various data of power grid operation and management, ration…
arrange the operation of the distribution network, so as to realize user power consumption behaviour analysis, power saving, power consumption prediction, grid optimization, peak-shift dispatch and other business application scenarios, and realize big data technology. The technological expansion in the field of smart grids improves the level of intelligence and comprehensive benefits of the grid, and meets the requirements of the in-depth construction of smart grids for the integration and mining of multi-source data.

As the most important part of the power system, the distribution network is directly connected with users, which directly reflects the comprehensive operation level of the power grid. It is also a "big potential tapping company" for the economic operation of the power system in reducing losses and increasing efficiency. The plan provides reliable support for the construction of a green, strong and smart grid. At present, due to the relatively late start of the development of my country's distribution network, the phenomenon of "emphasizing construction and neglecting planning" is more common. The planning and design of the distribution network is clearly lagging behind the urban economic development and does not meet the requirements of the entire urban economic construction. Among them, problems such as unreasonable distribution network planning, insufficient professional coordination, heavy equipment overload, light no-load, and low efficiency of system fault diagnosis have become increasingly prominent. Therefore, in the era of big data, tapping the potential and increasing efficiency based on the massive data of the distribution network is of great significance to the improvement of the management efficiency and development quality of the distribution network [1].

2. Demand analysis

The internal data of power grid enterprises and external environmental information data jointly support the intelligent development of distribution networks. Among them, internal data mainly includes: (1) Distribution network equipment physical information parameters and organizational asset management information, (2) Distribution network operation data and topological relationships, real-time perception of the distribution network operation status, further monitoring, early warning, and diagnosis. Provide reliable support; external data mainly includes geographic information data and meteorological information data, laying a solid foundation for the study of the characteristics of the distribution network in the region, the analysis of new energy development, and disaster preparedness.

Table 1. Internal data and external environmental information data of power grid companies

| Data source       | Data Classification     | Data content                                                                 |
|-------------------|------------------------|-------------------------------------------------------------------------------|
| Internal data     | basic information      | City, maintenance team, nature of assets, equipment status                    |
|                   | Equipment type         | Substations, distribution lines, switching stations, ring network cabinets, branch boxes, power distribution (low voltage) switches, distribution transformers, user meters |
|                   | Topological relationship | Station line relationship, line change relationship, station-user relationship, branch, segment, contact |
|                   | Operating data         | Three-phase current, three-phase voltage, active power, reactive power, power factor, gateway power, switch information, power outage information, fault repair |
| External data     | Geographic information | Latitude and longitude, altitude, geographical features, geographical features |
|                   | Weather information    | Weather, air pressure, temperature, wind speed, wind direction, humidity, precipitation, irradiance, |

2.1. Functional positioning

The intelligent auxiliary decision-making platform for distribution network planning is based on providing systematic and comprehensive information support for distribution network planning.
management and design research work. It runs through the whole process of distribution network planning and covers the whole service of distribution network planning and design. The organization of data provides information-based, visualized, interactive and intelligent auxiliary decision-making [2].

On the one hand, it realizes the integration, update, sharing and management of data resources related to distribution network planning, and provides the aggregation and integration of panoramic data required for distribution network planning; on the other hand, it integrates massive distribution network planning business-related data. Carrying out relevant predictive analysis of distribution network, providing advanced auxiliary decision-making functions of distribution network, realizing distribution network investment benefit analysis, project optimization and intelligent comprehensive auxiliary decision-making, and becoming an auxiliary decision-making platform that comprehensively supports 110kV and below distribution network planning.

2.2. Data requirements
The convergence and integration of big data related to planning business is the basis for the realization of an intelligent auxiliary platform for distribution network planning. According to the business needs of distribution network planning and the needs of intelligent auxiliary decision-making, the platform needs to integrate grid information, equipment ledger, operation information, user information and external information, share and integrate business data of various professional lines, real-time monitoring data, external data, etc. Multi-source heterogeneous data.

2.3. Functional requirements
According to the functional positioning of the intelligent auxiliary decision-making platform for distribution network planning based on big data, the following main functions should be provided:

2.3.1. Realize the integration and data sharing of various information systems. On the basis of the original professional support systems, through the construction of an intelligent auxiliary platform for distribution network planning based on big data, data channels are opened up, and multi-source heterogeneous information such as power grids, equipment, resources, etc. are integrated to form intelligent distribution network planning and design Big database [3].

2.3.2. Support the completion of distribution network planning tasks. Through the construction of an intelligent auxiliary platform for distribution network planning based on big data, accurate query of information related to the power supply situation from 110kV to 380V is realized. Through electricity forecasting, load forecasting, business expansion analysis, topology analysis, grid non-compliance analysis, Functions such as investment project optimization and sorting provide decision support for distribution network planning, so as to ensure the scientific and feasibility of the planning scheme, improve the investment efficiency of the distribution network, and improve the level of lean management.

2.3.3. Promote the overall improvement of the company's planning level. Through the construction of an intelligent auxiliary platform for distribution network planning based on big data, the current problems of power supply companies such as low level of business data management and poor information sharing capabilities among various systems are solved, and planning, analysis and prediction and intelligent auxiliary decision-making capabilities are improved, and gradually become "Intelligent Brain Platform for Distribution Network Planning and Design".
3. Big data technology supports the technical solution of distribution network planning

3.1. Overall architecture
This project is implemented based on the overall architecture of the State Grid Corporation's full-service unified data center, as shown in Figure 1 below. According to the overall structure of the State Grid Corporation's full-service unified data center, it provides unified storage and management of structured and unstructured data, realizes the extraction, cleaning, storage and construction of multi-dimensional analysis models of structured data, and supports multi-dimensional analysis applications.

![Figure 1. Big data technology supports the distribution network system framework](image)

3.2. Data Architecture
According to the overall architecture requirements of the data analysis domain of the State Grid Corporation's unified data center, the technical architecture of this project is designed, as shown in Figure 2.
3.2.1. **Data access.** The data access work of this project is to connect all kinds of data from the data source to the analysis domain of the unified data center of the whole business for use in the calculation and analysis process. The types of access data include: structured data and collected measurement data.

3.2.2. **Data storage.** The project data warehouse uses a distributed structured database, and the collection and measurement data storage area use a distributed column database for storage.

3.2.3. **Data calculation.** The data calculation work involved in this project includes offline calculation and real-time calculation. Distributed calculation technology is mainly used to build data offline calculation engine and real-time calculation engine functions [4].

3.2.4. **Application display.** After the data mart integrates and processes the normal monitoring theme result data, it uses the display layer of the unified data center application for the whole business, uses the Tableau analysis component to configure the theme scene application, and develops a monitoring report automatic generation tool based on the display results.
4. Technical Route

4.1. High-speed carrier technology

4.1.1. Signal modulation technology. High-speed carrier communication uses orthogonal frequency division multiplexing (OFDM) to modulate the signal. The main idea is to divide a given channel into dozens or even hundreds of independent and different orthogonal sub-channels in the frequency domain of 2-34 MHz. Each sub-channel uses one sub-carrier for modulation, and each sub-carrier transmits data in parallel, which improves the frequency band utilization. At the same time, it can effectively resist multipath interference, and even when the distribution network is severely interfered, it can ensure high bandwidth transmission efficiency, thereby achieving high-speed and reliable data communication [5]. This technology uses the power line with the widest coverage as a high-speed communication carrier, and divides the frequency range of 2-34 MHz into 1,364 independent and different sub-frequency channels to transmit data at the same time, avoiding the narrow-band frequency range with the most interference sources on the power line. The power line carrier communication spectrum is shown in Figure 3 below.

![Power line carrier communication spectrum](image)

**Figure 3.** Power line carrier communication spectrum

During the transmission, the carrier device will continuously detect the interference status of each sub-channel. If sudden interference (such as harmonics) is found or the electromagnetic interference in some sub-channels is very serious, the device can intelligently adjust to transmission in other non-interference sub-channels to avoid communication interference. As shown in Figure 4.

![Carrier burst interference detection equipment](image)

**Figure 4.** Carrier burst interference detection equipment
4.1.2. **Communication protocol.** High-speed carrier-based data collection uses HPLC protocol to encapsulate the data. The protocol contains 64-bit protocol header information, and the rest is data information. The Q/GDW 1376.2 protocol is used for concurrent meter reading commands. The steps of the concentrator for downlink messages are as follows: 1. Configure high-frequency collection rules, determine the number of multiple frames of Q/GDW 1376.2 messages, and determine each Q/GDW 1376.2 message per frame Contains the number of meter protocol frames. 2. The assembly of the concentrator is completed and the Q/GDW 1376.2 downstream message is delivered. The format of the downlink data unit of concurrent meter reading is shown in Table 2:

| Data content          | Data Format | Number of bytes |
|-----------------------|-------------|-----------------|
| Protocol type         | BIN         | 1               |
| Keep                  | BIN         | 1               |
| Message length (L)    | BIN         | 2               |
| Message content (DATA)| BIN         | 1               |

Protocol type: 00H, transparent transmission; 01H, DL/T645-1997; 02H, DL/T645-2007; 03H, DL/T698.45; 04H-FFH: reserved; message length: the total length of the original message data of the protocol (L1+L2+…LN); Message content: original message data of the protocol. The format of concurrent meter reading message is shown in Figure 5 below:

**Figure 5.** Concurrent meter reading message format

When the protocol type is 01H or 02H (DL/T645), the message content allows multiple (not more than 13) DL/T645 messages; when the protocol type is 00H or 03H (DL/T698.45, etc.), the message the total length of the content does not exceed 2000 bytes. The concentrator continuously sends multiple AFN=F1-F1 meter reading frames (Q/GDW1376.2) downstream to the monitoring terminal, and when receiving a denial response from the monitoring terminal, it stops sending meter reading frames [6].

4.2. **Fault diagnosis technology for distribution network**

The distribution network fault prediction method is mainly divided into three steps, namely: (1) The data pre-processing part, which is composed of the analysis of the influence factors of the distribution network fault and the sample screening based on the particle swarm algorithm; (2) Feature variable selection Part, the preliminary feature selection of each relevant variable, and then use a special selection algorithm to re-screen to form the optimal feature variable set; (3) Construction of the distribution network fault prediction model based on the fault level division; classify the faults and use The improved particle swarm optimization algorithm optimizes the parameters of the support vector machine, and then the distribution network fault prediction model [7].
4.3. Automatic drawing technology of distribution network

The automatic drawing process of the distribution network one-line diagram can be divided into two parts: automatic layout and automatic wiring. The automatic layout process determines the coordinates of the graphic elements and the direction of the line to ensure that the one-line diagram layout is compact and beautiful and there is no overlap of graphic elements. The automatic wiring process is based on the automatic layout, calculates the outlet port of the primitive according to the coordinates and size of the primitive, and completes the automatic connection between the primitives.

Firstly, the line classification is performed, and the breadth-first search algorithm is used to traverse the tree nodes of the distribution network. Starting from the power supply node of the substation, the longest branch is found and the level of the branch node is marked as 0. Then visit the nodes on the 0-level branch in turn. If there is a T-connected branch on the node, traverse the branch to find the longest line on the branch as the first-level branch. Continue to recursively classify the line until the levels of all nodes have been marked [8].

Set as a rectangle, length and width are respectively \( l, W \); component center coordinate is \((x, y)\), pin \( P_y (j = 1, 2, ..., m_i)\); coordinate is \((x_j, y)\), \(x_j = x + \Delta x, y_j = y + \Delta y\), component is placed horizontally or vertically, for each pin, \(\Delta x, \Delta y\) is constant. The layout must achieve the following three goals: (1) The Manhattan distance \(d_{ij}\) between the constrained pins of the layout components meets the delay constraint; (2) The components with a large connection degree \(G(i, j)\) are as close as possible; (3) The minimum spacing between layout elements is \(\Delta S\). For the above three layout goals, we use the following energy function to describe them:

\[
E = E(v_1, v_2, ..., v_n)
\]  

\[
E = (A/2) \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left[ d_{ij} - \frac{\text{MAX} + \text{MIN}_{ij}}{2} \right] ^2 \left( \frac{\text{MAX} + \text{MIN}_{ij}}{2} \right) ^2
\]

5. Conclusions

With the advent of the era of big data and the development of big data application technology, obtaining effective data such as production data, management data, topography data, and water resources data of electric power enterprises under the platform of enterprise data sharing can extract accurate and valuable data. Data provides effective help for management efficiency and decision-making ability improvement. By integrating the information of various decentralized systems, standardizing data types, forming a rich and homogeneous big data sample, predicting the possibility of failure of equipment of different types, models, and states, and taking targeted protection for grid operation inspection The measures provide support, guarantee the safe operation of the power grid and the self-healing of the smart grid, further improve the equipment operation and management level, and provide a reliable data basis for scientific decision-making on operation and maintenance, which has huge economic value and social benefits.
References

[1] Cao, Z., Lin, J., Wan, C., Song, Y., Taylor, G., & Li, M. Hadoop-based framework for big data analysis of synchronised harmonics in active distribution network. Iet Generation Transmission & Distribution, 11(16) (2017) 3930-3937.

[2] Wang, G., Gunasekaran, A., & Ngai, E. W. T. Distribution network design with big data: model and analysis. Annals of Operations Research, 270(2) (2018) 539-551.

[3] Ahmadian, A., Sedghi, M., Fgaier, H., Mohammadi-Ivatloo, B., Golkar, M. A., & Elkamel, A. Pevs data mining based on factor analysis method for energy storage and dg planning in active distribution network: introducing s2s effect. Energy, 175(5) (2019) 265-277.

[4] Daghi, M., Sedghi, M., Ahmadian, A., & Aliakbar-Golkar, M. Factor analysis based optimal storage planning in active distribution network considering different battery technologies. Applied Energy, 183(12) (2016) 456-469.

[5] Liu, M., Qu, M., & Zhao, B. Research and citation analysis of data mining technology based on bayes algorithm. Mobile Networks & Applications, 22(3) (2017) 418-426.

[6] Sun, R., Ye, J., Tang, K., Zhang, K., Zhang, X., & Ren, Y. Big data aided vehicular network feature analysis and mobility models design. Mobile networks & applications, 23(6) (2018) 1487-1495.

[7] Morrison, S. M., Liu, C., Eleish, A., Prabhu, A., Li, C., & Ralph, J., et al. Network analysis of mineralogical systems. American Mineralogist, 102(8) (2017) 1588-1596.

[8] Wang, N., & Zhao, F. An assessment of the condition of distribution network equipment based on large data fuzzy decision-making. Energies, 13(1) (2020) 197-209.

[9] Lee, D., Kim, J., & Shin, J. Scale-free network analysis of big data for patent litigation cases in the united states. Journal of the Korean Physical Society, 70(4) (2017) 431-435.