Research on Power Grid Data Asset Management Based on Big Data BI Analysis – The Role of Renewable Energy Technologies

Du Zhong\textsuperscript{1,*} and Yao Haiyan\textsuperscript{2}

\textsuperscript{1}Hangzhou Electric Power Design Institute Co., Ltd., Hangzhou, Zhejiang 311100, China
\textsuperscript{2}State Grid Zhejiang Electric Power Co., Ltd., Hangzhou, Zhejiang 311100, China
E-mail: neepu_0607@yeah.net
\*Corresponding Author

Received 26 August 2022; Accepted 21 November 2022; Publication 11 July 2023

Abstract

With the advancement of the industrial Internet of Things era, and the use of renewable energy technologies, smart grid has become the main form of power grid, and data has become a key link. The problem of data islands is common, and the data value, especially the commercial value, has yet to be tapped. This phenomenon is particularly prominent in power grid enterprises. Therefore, a data management model and method for power grid enterprises are proposed. The model aims at data asset management of power grid enterprises, reveals the process of extracting, modeling, analyzing and value mining of power grid data based on big data BI technology, and puts forward a practical and easy-to-land data asset management method, which really solves the problems of how to collect, store, manage and use data. Finally,

doi: 10.13052/spee1048-5236.4241
© 2023 River Publishers
taking the linkage between power marketing data and power grid planning in a certain area of Hangzhou, Zhejiang Province as a specific scenario, the BI analysis conclusion of relevant marketing data is displayed, which realizes the echo of conventional application and advanced application, reduces investment practically, improves economic benefits, and demonstrates the feasibility and scientficity of the method.

**Keywords:** Business intelligence, big data analysis, data assets, marketing data.

### 1 Introduction

With the advancement of the smart grid, the amount of data collection has grown rapidly, and the power supply company has stored a lot of grid data, which has been shelved for a long time and cannot be effectively used, resulting in a huge waste of resources. In addition, with the rise of the concept of data assets and the widespread use of big data business intelligence (BI) analysis in recent years, data mining and data applications have become the focus of power supply companies. Therefore, research on data asset management based on big data BI analysis plays a significant role in promoting rational data asset management and effectively supporting the rapid, stable and sustainable development of power supply companies.

BI was first proposed by the Gartner Group. This technology applies data warehouse technology, data analysis and processing technology, data deep mining and integration to realize data value management, and finally realizes data intuitive visualization through data display technology [1, 2]. The traditional BI data-based business decision-making technology integrates various types of scattered data, eliminates data islands, and finally realizes data synergy and value-added [3]. Driven by the era of intelligent information, BI has become a research hotspot and is widely used in the field of business data management. P.K. Abdussalm expounded the working principle of data mining technology and data processing in BI and used BI to optimize the business process of enterprises and realize the orderly management of enterprise data assets [4]. Lee Ki Kwang integrates big data technology and BI application characteristics, uses the BI platform to build a demand analysis system for analyzing infrastructure facilities, and better understands the overall process of enterprise data analysis [5]. Li Na uses BI to organically integrate the scattered information in the enterprise’s business activities, and builds the enterprise’s leadership decision-making system based on the report
system, which provides a reference for the development of the enterprise and market competition [6]. Zhai Shuying and others used business intelligence data processing technology to extract, integrate, and analyze network multi-source data, build a linear regression model based on physical data, and solve site selection problems with high quality [7].

With the advent of the Internet cloud era, power enterprise data presents the characteristics of large data volume, multiple data structures, and low value density. Big data analysis technology aggregates and integrates multi-source heterogeneous data within the smart grid to enhance the value application of power data [8, 9]. At present, business intelligence big data analysis and processing technology is still in the exploratory stage in power data management. Referring to cloud computing processing technology and combining the actual needs of power user big data analysis, Wang Dewen et al. developed a power user side big data parallel load forecasting prototype system to achieve accurate load forecasting [10]. Miloš Radenković et al. comprehensively analyze the internal data and information flow of smart grid, and use big data analysis and decision-making technology to build a unified and centralized data repository, which helps emerging power market operators to optimize data management [11]. Bo Hu made a practical exploration of the application of big data in the power grid industry by building the framework of the big data platform in the power grid industry, using big data technology to predict the power load level and assess the status of power transmission and transformation equipment [12].

The power grid makes sure that energy resources are used in the effective way, increases the capacity of the power supply, and improves the efficiency and dependability of power system operations. The spinning reserve, or reserve generation capacity, in each area is decreased by connecting the generating units together. Long-distance electricity transmission increases the humidity inside power lines, leading to large energy losses in the form of heat. The three basic parts of an electrical grid are generation, transmission, and distribution. Based on the big data BI analysis data management concept, Xun Ting et al. designed a new comprehensive intelligent analysis decision and visualization system for power grid regulation data to realize the overall control of power dispatching by business personnel [13]. For effective renewable energy assessment and energy conservation forecasting for the economy, AIEM has been recommended. The analysis’s objective was to assess, contrast, and develop a model combining AI and economic data pertinent to the economic prediction of renewable energy sources [14]. Aiming at the multi-source heterogeneous data extraction and storage problems in the construction of the electricity big
data platform, Zheng Haiyan and others adopted the data storage solution and big data technology of HBase and Hive cooperation to comprehensively improve the accuracy of load modeling [15]. Hao Ran et al. applied the distribution network big data to the analysis of users’ electricity consumption behavior, built a unified big data processing platform, and effectively processed the heterogeneous big data of the distribution network, which is helpful for the management of power data resources of power companies [16]. Zhao Lin et al. studied the multi-source heterogeneous power grid model ETL technology by analyzing the big data storage structure, and proposed a power grid model management and analysis framework based on big data technology applied to the dispatcher to realize data visualization display [17]. Liu Daoxin takes the domestic and foreign power industry, finance, economy and other data as the research object, explores the storage, processing and mining of big data, and realizes the value-added of the data life cycle [18].

At this stage, power grid data management is a research based on the concept of data assets. The representative ones are: Wei Junchao and others proposed that data become a key element of the digital economy in the context of big data, in order to improve the use and value mining of data by enterprises ability, put forward the concept of data asset management [19]. Jun Wang et al. expounded the necessity of data asset management in the context of big data era, and established a framework for power grid data asset management based on the value-added characteristics of power supply enterprise data assets throughout the life cycle [20]. Based on the concept of ubiquitous Internet of Things (IOT), Cui Jindong et al. built a data ecological management model of power supply enterprises, effectively evaluated the actual value of data assets of power supply enterprises, and provided theoretical support for decision makers of power supply enterprises to make intelligent decisions [21]. Li Feifei et al. proposed an enterprise data asset management model for the problems of unsound data management and control of power supply enterprises in the era of big data, and deeply excavated the potential value of power data, so as to maximize the economic benefits of data assets [22]. A standardized wide area network architecture for heuristic renewable demand energy optimization in smart cities (HRDEOSC) is described [23]. This design distributes the smart area domain.

At present, most of the research at home and abroad focuses on the theoretical research on the application of big data analysis and processing technology in power data management and value mining, but no scholars have substantially applied and practiced data asset management under big
In addition, in the actual power grid planning and design work, the internal power data of the power supply company has various forms, the phenomenon of power data islanding is serious, the data display platform is single, and the analysis and application depth is not enough. Therefore, on the basis of constructing a visual analysis framework of power grid data assets, the author optimizes the power grid data asset management model of power supply companies. With the help of the big data BI analysis platform, the application of data asset management under big data analysis is effectively realized, the data value is deeply excavated, the data is intuitively visualized, and finally the security and utilization rate of power grid data are improved.

2 Visual Analysis Framework of Power Grid Data Assets Based on Big Data BI

Big data analytics for the smart grid has the potential to transform a business-like utility, which is not well-known for its technical advancements [24]. Data sensors and other equipment are being incorporated into the delivery of utility services, generating an expanding overlap between utilities and data. Many data analytics apps and algorithms are being developed with a focus on data analysis for smart grid. Big data analytics play an important role in the smart grid by assisting to plan future operation and maintenance and to extract useful information from the previous data. In the information age, data assets account for an increasing proportion of corporate decision-making, and giving full play to the value of power grid data has become a strategic choice to achieve high-quality development of power supply companies. However, the large-scale access of users of different formats within the distribution network and the long-term shelving of power grid data by power supply companies have resulted in the failure to fully tap the value of power data, the inability to accurately predict user load growth, and the inability to accurately grasp the law of power changes. It will also make it difficult to accurately estimate the actual power consumption and load of users, and cannot provide an accurate reference for subsequent grid investment and construction. The power in kW must be multiplied by the number of hours the devices are used each day, week, or month is generally used to determine the actual power consumption. Therefore, it is of great significance to use big data BI to visualize the power grid data, deeply mine the intrinsic value of the data, and rationally use the data asset management of power supply companies [25, 26]. The big data visualization analysis framework of data assets is shown in Figure 1.
2.1 Power Grid Data Information Collection Based on Big Data Technology

With the development of intelligent interconnection of distribution network and the development of intelligent terminal DG, ESS, and V2G technologies in distribution network, the power grid architecture presents a complex
development trend. Power supply company data is complex, multi-source and heterogeneous. The data includes internal enterprise data such as business systems and archive data, and external data such as internet of things data, government data, and Internet data. For internal data, big data technology mainly realizes the unified collection of power grid data through the data collection methods of power business system information index, database collection and manual input based on ETL server; the data integration process known as ETL, or extract, transform, and load, merges data from several data sources into a single, consistent data store that is imported into a data warehouse or other destination system. ETL is used for a variety of tasks, including the extraction of data from previous systems, cleaning it to increase its quality, and loading it into a target database; for external data, big data technology mostly uses web crawler to capture useful information; In response to the real-time data of smart grid demand response, the core data set is extracted based on real-time streaming data processing of big data, and finally multi-dimensional data that meets the requirements of the physical data model is formed, and then smoothly loaded on the ODS.

2.2 Power Grid Data Preprocessing Based on Big Data Technology

The original data structure of the power supply company is complex and there are some duplicate data and invalid data. It is necessary to screen, abnormally process and effectively integrate the massive data sources. Data cleaning involves correcting or deleting inaccurate, damaged, improperly formatted, duplicate, or insufficient data from a dataset. There are several ways for data to be duplicated or incorrectly categorized when integrating different data sources. This procedure is essential and stressed since incorrect data can lead to incorrect business decisions, conclusions, and poor analysis, especially if enormous amounts of big data are available. Use big data cleaning technology to detect and eliminate data errors and inconsistencies in dimensions, and realize standardization and normalization of data formats, measurement units and data types. For data quality problems, cluster fitting is achieved by using similar eigenvalues. By normalizing and standardizing the data, it realizes abnormal data detection, vacant data filling, redundant data elimination and inconsistent data processing, and at the same time completes the projection of high-dimensional data to low-dimensional space. Data integrates data from multiple data sources and stores them uniformly, mainly to solve the problems of pattern matching, data redundancy and data value conflicts. Pattern
matching is used to assess whether the source files of high-level languages are syntactically proper. Additionally, it is used to locate and interchange a text or code’s matching pattern for another text or code. Pattern matching is used in some capacity by every program that offers search capabilities. And use data transformation technology to smooth, aggregate or generalize the data, and convert the data into a form suitable for processing and analysis. Data redundancy is when the same amount of information appears in several locations, whereas data inconsistency is when the same data exists in various forms in multiple tables. Unfortunately, data redundancy may lead to data inconsistency, which might provide a business information that is inaccurate or irrelevant. Data redundancy strengthens the backup by copying data to a different system, adding an additional degree of security. Data redundancy is frequently advantageous when it is included in disaster recovery plans.

2.3 Power Grid Data Analysis and Processing Based on Big Data Technology

In order to tap the potential value in the data, big data analysis technology is used to quickly extract and analyze key information, and the analysis results are presented through visualization technology. Big data analysis technology extracts user load information with the help of data warehouses, intelligently processes and physically models the pre-processed data, and gives full play to the full life cycle value of data. Big data analysis mainly includes three types: descriptive analysis, exploratory analysis and confirmatory analysis. The main methods include regression analysis, association analysis, classification analysis and cluster analysis. Descriptive analysis is to process and display data through charts; exploratory analysis is to explore the structure and laws of data through chart making, equation fitting and feature calculation; confirmatory analysis is to predict future development trends through quantitative and qualitative analysis. Build the relevant model and then validate the proposed model with the existing data.

2.4 Intuitive Display of Power Grid Data Based on Big Data Technology

The traditional data display platform is limited to the case where the amount of data is small and the data is weakly correlated. The static image of the table shall prevail, and it is difficult to adapt to the multi-dimensional, massive and dynamic characteristics of power load data. The big data visualization technology can display the excavated information more intuitively, and
realize data visualization by means of graphics, parallel coordinates, pixels and layers. Data visualization’s major objective is to make it faster to see patterns, trends, and outliers in big data sets. The terms information graphics, information visualization, and statistical graphics are frequently used interchangeably. Additionally, the main features of data visualization are accuracy, clarification, strength, and comprehensibility. Big data visualization can realize multi-view integration and explore data relationships in different dimensions; through the interaction and linkage of data views, data query in data views can be realized. The data acquisition system, is used to collect information from various sensors in a PV system, is an essential element of any monitoring system. The DAS provides data to the control center for processing and delivery after this data is digitalized for storage. The data visualization platform of the big data analysis system combines the GIS platform and the data acquisition management system, and finally realizes the integration and fusion of power grid data and graphics in the form of tables, dynamic images and 3D videos through human-computer interaction, and then displays the integrated data intuitively. Power supply companies can display the development trend of dynamic high-dimensional data according to the visualization platform, grasp the current load information, tap the potential value of power data, conduct situation assessment and macro display of uncertain change points in the power grid, and realize risk early warning.

3 A Specific Approach to Data Asset Management Based on BI Analysis of Big Data

3.1 Data Requirement Analysis Based on Big Data BI Analysis

Data asset management is the process of collecting, managing, employing, optimizing, and leveraging data assets to generate value. All kinds of data assets quickly become liabilities if they are not managed. When it comes to physical assets, this is obvious. When analysing the data needs of power supply companies for grid work, they should take their core business as the starting point, analyze the specific gaps between the current status of enterprise data management and the requirements of the data asset management concept, and focus on the specific data needed for grid work under big data BI analysis and the use of data. Second, Electricity supply company should analyze their own business processes and data flow of data management in power grid work, determine the core business data sets, clarify the sources and expected application scenarios of data involved in power
grid data asset management, and focus on solving the problems of manual data entry, data not effectively penetrated, poor data quality, non-synergy of objectives and non-standardized operation. Collect log data, database data and other data using big data collection technology to ensure the comprehensiveness of data information. By analyzing the industry data of power supply companies, we understand the industry data characteristics of power supply companies and clarify the characteristics and needs of each industry of users involved in power supply companies.

3.2 Data Extraction Modeling Based on BI Analysis with Big Data

In response to the specific needs of data under data asset management, we extract and model the core data sets for external data as well as internal data. Firstly, we integrate the access and screening of external data, so that the data can be unified and standardized. At the same time, we analyze the access methods of external data and unify the data interfaces, data formats and data types of external data. Using big data collection technology to collect the required data to optimize the modeling of internal data sources, data modeling and extraction of internal data, desensitization of the extracted data, and the desensitized data as a source of data asset management to achieve the purpose of eliminating redundant data and erroneous data and ensuring data accuracy. The process of collecting or extracting various forms of data from many sources, many of which may be irregularly arranged or entirely unstructured, is known as data extraction. The data extraction and modeling process under Big Data BI analysis is shown in Figure 2.

3.3 Data Asset Visualization Based on Big Data BI Analysis

When power supply companies perform visual analysis of grid data assets, they should first realize the organic integration of data in multiple information systems such as PMS and EMS. By integrating relevant information data scattered in each business, the integration of grid data and graphics based on GIS greatly facilitates the management and analysis of spatial data, while grid data is also more direct and interactive. A Geographic Information System (GIS Software) is a computer program that is used to store, retrieve, manage, display, and analyze many forms of geographic and spatial data. GIS software allows users to create maps and other visual representations of geographic information for study and presentation. GIS software may be divided into two
Figure 2  Data extraction and modeling process under big data BI analysis.
primary categories: open-source and commercial. Commercial GIS software requires some sort of paid licensing, such as a subscription model or a one-time permanent license. Everyone may use open-source GIS software for free. Secondly, the use of big data analysis technology for deep and multi-dimensional screening and mining of power grid data, according to the deep-level data information to diagnose and analyze the power grid. Finally, the grid data visualization is analyzed based on big data visualization technology. By combining external information management system, data interface and human-computer interaction as a means to finally realize the integration of grid data and graphics in multiple forms such as tables, dynamic images and 3D videos. In turn, the integrated data is displayed visually, and the visualized analysis results are presented for diagnostic analysis, and solutions are formulated in a targeted manner based on the diagnostic analysis results.

3.4 Data Management Strategy Based on BI Analysis of Big Data

In the process of proposing data asset management strategy methods, power supply companies should first conduct research on internal information ecological management strategy, and gradually apply new technologies such as big data, Internet of Things and cloud technology to power grid planning work, and realize the integration, fusion, mining and unified management of internal and external data. The Ecological Management System (EMS) is a mapping and database inventory system. The Ecological Management System will assist in achieving the objectives of a well-managed, sustainable park system and a healthy natural environment. Other communities that use tree inventories and other ecologically based inventories and map systems benefit from the EMS methodology. Three categories of ecological management are parasitism, mutualism, and commensalism. Secondly, research on specific methods of business collaboration within the industry, optimize and adjust the workflow of power grid business, and form an efficient and unified power grid business work system while developing new business and adjusting existing business. Finally, we will conduct research on the mining of external value-added business, actively collect and integrate external data, screen and integrate external data for the work needs of power supply companies, and form an independent information base for power grid planning and design data. Power supply companies should adapt to the new social formations and actively create methods and strategies for data asset management in the new situation. By carrying out new business, new profit growth points can be added. The operating mechanism of data asset management under big data BI analysis is shown in Figure 3.
Figure 3  Demonstration of the operational mechanism of data asset management with big data BI analysis.

4 Application of Data Asset Management Based on Big Data BI – Taking the Planning and Design Data of HZ Power Supply Company as an Example

4.1 Planning Data Collection and Processing Based on Industry Segmentation

Taking HZ Power Supply Company as an example, through its research, it was found that the company had previously invested blindly, with problems such as data on hold, inaccurate load forecasting, inability to accurately grasp power changes and poor business environment for electricity. In response to the above-mentioned data management status and problems, HZ Power Supply Company carried out data asset management work, successfully managed and applied the data, and then made effective guidance for planning and design work. Through analysis of the case, the data came from the development and planning department of HZ Power Supply Company, the power design institute and government departments, etc. Big data technology
was used to collect planning and design-related industry data and establish a temporary database, and the collected core data and external data were separately saved in the temporary database, with a total of 32579 pieces of data acquired.

We import the temporarily stored internal and external data into the big data BI software for data preprocessing. First, data cleaning is performed to remove duplicate data and invalid data. The filtered data is processed with unified standardization and unified data caliber, and the processed external data is stored in a separate database. The extracted core data sets are deep mined to build a suitable data structure and stored in a separate database. The process of combining data from several sources into a single, unified representation is known as data integration. Integration involves procedures like cleaning, ETL mapping, and transformation and begins with the ingestion process. Data integration and data transformation of the core data and processed external data, data smoothing using binning and clustering, followed by data aggregation and aggregation, and finally attribute construction to achieve deep data fusion. Data fusion is the collaborative analysis of many interconnected datasets that offer different perspectives on the same phenomena. In general, more precise inferences may be derived from the correlation and fusion of data from many sources than from the analysis of a single dataset. The merged data is desensitized, and the desensitized valid data is stored in the core database of the power grid business as the source of data asset management, and finally 26,550 valid data are obtained.

4.2 BI Analysis of Planning Data Based on Industry Segmentation

After the data preprocessing is completed, the BI software is used to perform a big data visualization analysis on the data, and through cluster analysis, the electricity consumption characteristics of users in different regions, industries and users with different contract capacities are revealed. On this basis, detailed analysis of user electricity consumption and load change trends was carried out to more accurately predict the load growth and electricity changes of new users over the next five years.

(1) Analysis of electricity consumption characteristics by industry type

By grouping industries with the same attributes as well as types in the data pre-processing, 11 industry types were obtained and the percentage of each
industry type is shown in Figure 4. Among the 11 types of industries, the construction industry, agriculture, animal husbandry, fishery and forestry, and public service and management organizations accounted for the largest proportion of the number of newspapers. The construction industry had the highest number of business applications, accounting for 43%, followed by agriculture, forestry and fishery services, accounting for 18%, the number of public service and management organizations accounted for 17%, while the remaining industries accounted for less than 5% of business data.

According to the classification results of industry types, the annual electricity consumption and the annual maximum load of different industry types in the past five years are obtained by visual analysis of the annual average electricity consumption and annual average maximum load of 11 industry types. The annual electricity consumption and annual maximum load for the last five years for different industry types are shown in Figures 5 and 6 respectively.

It can be seen that the annual electricity consumption and the annual maximum load of different industries are very different. The load rate of construction industry, public service and management organization, leasing and commercial service industry is relatively high, and the load fluctuation
Figure 5  Distribution of annual electricity consumption.

Figure 6  Annual maximum load distribution.
rate is large. The load change rate is small. The construction sector had the largest annual maximum load of the 11 industry types, but was at the bottom of all industries in terms of annual electricity consumption; instead, the wholesale and retail sector had the largest annual electricity consumption. In addition, electricity consumption and maximum load were generally low in the first two years for most industries and did not increase rapidly until the third year; whereas electricity consumption was higher in the first and second years for the wholesale and retail sector. The different electricity consumption characteristics of different industry types lead to significant differences in the actual data and trends in annual electricity consumption and annual maximum load between different industries, making load forecasting more difficult. Therefore, when making electricity consumption and load forecasts, the electricity consumption characteristics of different industry types should be taken into account and electricity consumption and load forecasting models for different industry types should be established separately.

In addition, it can be seen from the visualization data that the newly added business expanding system has passed the outbreak period, and the growth of electricity consumption and load has also entered a stable period; secondly, as the national economy has entered a new normal, the growth of electricity consumption has also returned to a stable mode. However, the electricity data of government agencies, public services and management organizations is still gradually rising with great potential. Therefore, it is necessary to communicate with the government in the future. The government is not only a manager, but can also provide massive electricity bills, business and data information. Secondly, the industry differences in electricity growth are becoming more and more obvious, and the differences in input-output benefits are obvious. The load and electricity generated by application and installation business in different industries show different growth characteristics. Textile, real estate and other industries have the highest electricity benefit ratios, while accommodation and catering industries, traditional industries and other industries have significantly lower benefit ratios than other industries. In the future, marketers should pay more attention to new industries (textiles, etc.) and the smart economy, because these industries have higher input-output ratios, future development prospects and profits are considerable.

(2) Analysis of contract capacity electricity consumption characteristics
All data are grouped according to different contract capacities, and carry out big data visualization analysis to calculate the average annual electricity
consumption in the past five years under different contract capacities. The electricity consumption in the first five years under different contract capacities is shown in Figure 7. It can be seen that electricity consumption increases with the increase of contracted capacity. In addition, the rate of change in electricity consumption between different years also increases. Among them, those with a capacity of less than 2000 KW, the annual electricity consumption is small and the change is relatively stable between years; while the capacity of more than 2000 KW, the annual electricity consumption is proportional to the contracted capacity, and the range of changes is large between years. In addition, the contract capacity of most of the installation business is below 2000 KW, therefore, the factor of contract capacity should be considered when forecasting electricity consumption and load.

As the national economy enters a new normal, electricity growth returns to a steady mode. For the future new application and installation business, we should maintain a stable mentality, and do not just seek the high growth of large users. In the future, the installed capacity of the single unit will be larger and larger, and the large-capacity installation will have a greater impact on the stability of the power grid. Although the number of times when the maximum reported installed capacity is above 2000 KW is stable, the maximum installed capacity of a single unit continues to increase, and the problem of stable operation of the power grid brought about by large-capacity installations is worth vigilant. In addition, the increase in the difficulty of large-capacity installation related work will not only bring difficulties to the scheduling department, but also bring related impacts to the operation and
maintenance department, development planning department, and marketing department.

(3) Analysis of electricity consumption characteristics in geographic regions

Using the combination of big data and geographic information system, the data is imported into the geographic information system through big data BI software, all businesses are marked in the geographic information system, and the geographic distribution map of business data is obtained, as shown in Figure 8. Through the visual analysis of the GIS data, it can be seen that most of the data are aggregated into clusters, and only a small part of the data is relatively scattered. When carrying out the installation business, the overall planning of the power grid should be considered, and the geographic information system should be visualized through effective management of data assets, so that the data can be displayed graphically, which is more intuitive. Businesses in different regions have large differences in electricity consumption and load, it is difficult to predict the load in densely planned areas, so different planning methods should be adopted. In addition, the degree of economic development in different regions is different, and the future planning and development plans are also very different, which will affect the electricity consumption and load to a large extent. When forecasting electricity consumption and load, different factors in the planning area should

![Figure 8 Geographical distribution of business data.](image-url)
be considered, the planning area should be divided into different grids, and electricity consumption and load should be more accurately predicted according to different grids. According to the difference in the quantity distribution of business data, economic development and power grid planning degree in different geographical regions, the distribution network region is divided into 42 grids, and the services are divided into their respective grids. It is convenient for the management of future power grid planning business data and the access of new installation services. Different planning and access measures are adopted for different grids, and the power consumption and load changes can be predicted more effectively and accurately. The grid division diagram of the distribution network is shown in Figure 9.

4.3 Data Asset Management Application Based on Big Data BI Analysis

According to the visual analysis of data assets under big data BI, the electricity consumption and load prediction model is constructed, and the user load and electricity consumption changes are more accurately grasped. From the visual analysis, it can be seen that the user’s electricity consumption and load are closely related to the industry type, contract capacity and regional division. When building an electricity consumption/load forecasting model, it should be clustered and grouped to achieve a more accurate forecast.
4.3.1 Routine analysis of data asset management based on big data BI

Cluster analysis is performed on the preprocessed data. First, the data of different industry types are clustered, and then the data of the same industry type is clustered again according to the capacity of 2000 kW. Use time series model and linear regression analysis to construct electricity consumption and load forecasting model per unit capacity by industry. The fitting degree of the power and load regression models of various industries is analyzed, and the linear regression equation is determined. The unit capacity electricity/load forecasting model is shown in Table 1. In addition, data mining is carried out according to geographic information, the electricity consumption and load are reviewed and diagnosed, the prediction model is evaluated according to the review and diagnosis results, and finally the electricity and load are corrected for the diagnosis information. Based on the existing business data, the electricity consumption of the new installation business and the results of the calculation formula of the load forecast model are evaluated to verify the correctness of the model.

Based on the industry’s name and the contract capacity size, the corresponding unit capacity electric quantity/load forecast model is selected. The x value is entered as the number of years to be forecasted (taking 1, 2, 3, 4, 5, etc. integers) to derive the unit capacity electric quantity/load forecast for the Xth year (variable Y). Newly reported forecasted electricity consumption = annual forecasted value of unit capacity electricity in the industry to which it belongs × newly reported contract capacity × capacity correction factor × grid correction factor; forecasted load value of newly reported load = annual forecasted value of unit capacity load in the industry to which it belongs × newly reported contract capacity × capacity correction factor × grid correction factor. The table of capacity correction coefficients is shown in Table 2, and the table of grid correction coefficients is shown in Table 3.

4.3.2 Advanced application of data asset management based on big data BI analysis

To verify the correctness of the data asset management application under big data BI analysis and the accuracy of the prediction model, we now take the new real estate user’s application for installation in the Nanyuan subdistrict of Hangzhou as an example. This user is powered by a double-line power supply and is scheduled to be connected in early 2020. The surrounding area is powered by the scenic G609 line and Chenyuan G619 line, and the
### Table 1  Unit capacity electricity/load forecasting model

<table>
<thead>
<tr>
<th>Serial Number</th>
<th>Industry Name</th>
<th>Unit Capacity Contract Capacity</th>
<th>Electric Quantity Forecast</th>
<th>Load Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Real estate (real estate development and management, etc)</td>
<td>&lt; 2000 kW</td>
<td>Y = (-6.54 + 85.118X - 28.358X^2 + 2.873X^3)</td>
<td>Y = (-0.027 + 0.047X)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 2000 kW</td>
<td>Ln(Y) = 6.605 - 0.909/X</td>
<td>Ln(Y) = (-0.824 - 3.231/X)</td>
</tr>
<tr>
<td>2</td>
<td>Textile industry (textile, printing and dyeing finishing, etc)</td>
<td>&lt; 2000 kW</td>
<td>Y = 1183.411 + 2857.954X - 1092.317X^2 + 114.733X^3</td>
<td>Ln(Y) = (-0.644 - 0.591/X)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 2000 kW</td>
<td>Ln(Y) = 8.032 - 2.12/X</td>
<td>Ln(Y) = (-0.829 - 0.808/X)</td>
</tr>
<tr>
<td>3</td>
<td>Industry (Manufacturing of other cast metal products and structural metal products.)</td>
<td>&lt; 2000 kW</td>
<td>Y = (-534.588 + 1300.184X - 402.292X^2 + 38.485X^3)</td>
<td>Y = (0.213 - 0.081X + 0.029X^2 - 0.003X^3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 2000 kW</td>
<td>Ln(Y) = 195.78 + 1060.371X - 389.029X^2 + 43.4X^3</td>
<td>Ln(Y) = (0.134 - 0.049X + 0.016X^2 - 0.002X^3)</td>
</tr>
<tr>
<td>4</td>
<td>Public service and management organizations (Public lighting, municipal facilities management, etc)</td>
<td>&lt; 2000 kW</td>
<td>Y = 119.21X - 63.01</td>
<td>Ln(Y) = (-2.122 - 0.783/X)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥ 2000 kW</td>
<td>Ln(Y) = 287.22 + 355.552X - 97.689X^2 + 8.058X^3</td>
<td>Y = (0.308X - 0.023)</td>
</tr>
</tbody>
</table>
Table 2  Capacity correction coefficient table

<table>
<thead>
<tr>
<th>Serial Number</th>
<th>Contract Capacity</th>
<th>Electric Quantity</th>
<th>Actual Electric Quantity</th>
<th>Electric Quantity Correction Coefficient</th>
<th>Load Actual Load</th>
<th>Actual Load Correction Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt;2000 kW</td>
<td>51.5912</td>
<td>56.2339</td>
<td>1.0900</td>
<td>0.1117</td>
<td>0.1262</td>
</tr>
<tr>
<td>2</td>
<td>≥2000 kW</td>
<td>51.5912</td>
<td>48.8396</td>
<td>0.9467</td>
<td>0.1117</td>
<td>0.1031</td>
</tr>
</tbody>
</table>

Table 3  Grid correction coefficient table

<table>
<thead>
<tr>
<th>Serial Number</th>
<th>Belonging Grid</th>
<th>Electric Quantity Value of Grid</th>
<th>Electric Quantity</th>
<th>Grid Load Actual Load</th>
<th>Electric Grid Load Correction Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HZH2-DH-01B</td>
<td>51.5912</td>
<td>51.5734</td>
<td>0.9997</td>
<td>0.1117</td>
</tr>
<tr>
<td>2</td>
<td>HZH2-DH-02B</td>
<td>51.5912</td>
<td>70.4938</td>
<td>1.3664</td>
<td>0.1117</td>
</tr>
<tr>
<td>3</td>
<td>HZH2-DH-03B</td>
<td>51.5912</td>
<td>51.7751</td>
<td>1.0036</td>
<td>0.1117</td>
</tr>
<tr>
<td>4</td>
<td>HZH2-DH-04B</td>
<td>51.5912</td>
<td>51.7575</td>
<td>0.9997</td>
<td>0.1117</td>
</tr>
<tr>
<td>5</td>
<td>HZH2-DH-05B</td>
<td>51.5912</td>
<td>51.6736</td>
<td>1.0016</td>
<td>0.1117</td>
</tr>
</tbody>
</table>

The line load is close to saturation. If its capacity is measured according to the conventional method, it needs to arrange to cut the load by 2.3 MW and finish the implementation in early 2020. Secondly, according to the industrial development of the surrounding land, there is a large commercial user who needs to be connected to electricity in 2021. To meet the users’ demand for electricity, there are also investments to be made in the renovation of the distribution network. After the commissioning of the 110 kV Highland Substation in 2021, a secondary investment will be required to improve the regional network, resulting in a waste of local grid investment.

This real estate user newly installed contract capacity is 15,000 kVA (15,000 kVA = 12000 KW), belonging to the real estate development industry, and also belongs to HZH2-LZ-04B grid, according to the contract capacity of more than 2000 KW real estate user load prediction model Ln(Y) = −0.824−3.231/X, bring X1 = 1 into the formula to get Y1 = 0.01734, Y2 = 0.0872, Y3 = 0.1494, Y4 = 0.1956, Y5 = 0.2298; according to the correction coefficient table, we find that the correction coefficient of load grid is 0.9850, the correction coefficient of load capacity is 0.9228, the maximum load= Yi*electricity consumption grid correction factor*electricity consumption capacity correction factor*the contract capacity, and each data is brought into the formula, that is, the maximum load in the first year = 0.01734*0.9850*0.9228*12000 = 189.14 kw, and the maximum load in other years is calculated in this way, that is, 551.14 kw.
in the second year, 1629.58 kw in the third year, 2133.49 kw in the fourth year, and 2207.4 kw in the fifth year. The forecast model for electricity consumption of real estate industry users with contract capacity above 2000 KW is \( \ln(Y) = 6.805 - 0.909/X \). According to the forecast model, \( X1 = 1, X2 = 2, X3 = 3, X4 = 4, X5 = 5 \) are brought into the equation to get \( Y1 = 363.58, Y2 = 572.78, Y3 = 666.47, Y4 = 718.92, Y5 = 752.35 \); according to the correction coefficient table, the correction coefficient of electricity consumption grid is 0.7974, and the correction coefficient of electricity consumption capacity is 0.9467, and the electricity consumption in the year \( i = Yi \times \text{correction coefficient of electricity consumption grid} \times \text{correction coefficient of electricity consumption capacity} \times \text{contract capacity} \). Bring the data into the formula, that is, the first year electricity consumption = 363.58*0.7974*0.9467*12000 = 3293592.31 kwh; other years electricity is calculated in the same way, that is, the second year is 5188689.7 kwh, the third year is 6037407.08 kwh, the fourth year is 6512540.25 kwh, the fifth year in Figure 10 shows the maximum load forecast for this customer in the next five years.

Through the observation of this user’s future five-year load and power forecast map, it is found that the real estate user’s load factor in the first and second years is not high, and the impact on the load of existing lines is limited, so it can be connected to the grid first but not arrange for line modification. At the same time, using the “online grid”, we carried out near- and mid-term load forecasting of grid parcels, optimized the grid-based planning by summarizing and collating the diagnostic conclusions of grid analysis, and settled the matter at one go in the project of Highland Substation 10 kV supporting line out, saving over $5 million in retrofitting.
costs for intermediate transition options. The cutover scheme before and after the optimization of the application of the new business expansion user model is shown in Figures 11 and 12.

5 Conclusion

In the process of smart grid planning in the era of big data, a large number of unstructured data sets are generated. How to tap the deep value of data inside and outside the enterprise and realize the efficient management and utilization of data assets is an urgent problem to be solved by contemporary power supply enterprises. The author proposes a data asset visualization architecture for power supply enterprises using big data technology and analyzes the operation mechanism of data asset management under big data. Taking the grid planning data asset management of Hangzhou Power Supply Company as an example, the author uses the big data BI analysis platform to realize data visualization and propose targeted countermeasures; and build the electricity and load forecasting model for the new installation business,
which provides a strong basis for the formulation of grid planning strategies as well as business development. To a certain extent, the author has realized the application of data asset management of power supply enterprises under big data analysis and improved the data asset management of power supply enterprises. However, there is still much room for improvement in data asset management and its application in power supply enterprises, and these are the directions for future research.

**Acknowledgments**

This work was supported by the National Natural Science Foundation of China [51607032].

**Conflicts of Interest**

The authors declare no conflict of interests.

**References**


**Biographies**

**Du Zhong**, born in November, 1979, is a senior engineer, majoring in electrical engineering. He is now engaged in electrical engineering and works for Hangzhou Electric Power Design Institute Co., Ltd.
Yao Haiyan, born in September, 1978, is a senior engineer, majoring in electrical engineering. He is now engaged in State Grid Zhejiang Electric Power Co., Ltd.