Construction Method and Typical Application of Data Analysis Service for Power Grid Enterprises Based on Data Middle Platform

Li Wenjuan[1], Liu Shi[1], Zhang Fan[1], Yang Zhi[1], Wang Honggang[1], Hu Xishuang[1], and Zhou Wenjin[2(✉)]

[1] Big Data Center of State Grid Corporation of China, Beijing 100052, China
[2] Beijing State Grid Xintong Accenture Information Technology Co., Ltd., Beijing 510030, China
kqx019@sina.com

Abstract. With the promotion of the construction of the power Internet of things of State Grid Corporation of China, the company's business is developing towards lean and digital, and the range of data demand is expanding. The company requires the construction of enterprise data center to provide various data analysis services, and support the big data application of important business efficiently and flexibly. The data analysis service is based on the unified construction of data center, It provides service retrieval and view to users in Web mode, and provides data analysis service to external business system and business platform through unified data analysis service in data platform to support rapid construction of business analysis scenario application.

Keywords: Data analysis Service · Data Center · Power Grid Enterprise · Power Data

1 Introduction

According to the company's strategic deployment, the data center of State Grid Corporation of China has been initially completed, and the data analysis service will be uniformly constructed through the data center to support cross department and cross level data sharing and analysis applications. In order to effectively manage the company's data assets and avoid the problems such as high threshold of data analysis service application, difficulty in data understanding and service acquisition, it is urgent to build enterprise level data analysis service construction standard and construction ability based on data platform, provide unified data analysis service for all specialties, improve data management level, and create a benign ecological environment for data assets [1]. At present, data center has become an indispensable technical support platform for power grid enterprises in production, operation, management and decision-making. With the continuous expansion of business, the scale of the enterprise is growing every year. In the process of operation, there are more and more failures, some of which even make system

administrators unprepared [2]. After a fault occurs, it is often necessary to analyze the log to locate the problem. Log is an important tool for analyzing user behavior, monitoring system running status and locating system fault. And through log analysis can also find system vulnerabilities, optimize system design, help to quickly locate system faults, find the root cause of the problem [3].

However, at present, the development of the power industry is relatively backward, facing many serious problems, such as the old and lagging power grid construction, the lack of scientific basis for the allocation of power resources, and the sluggish infrastructure construction of the power market. Among them, the lack of reactive power supply is the main factor leading to power grid blackout, voltage fluctuation, harmonic distortion, serious loss and other problems, while power failure and power grid loss are the main factors restricting the growth of power grid enterprises. Therefore, through the local compensation of facts and other reactive power compensation equipment, it can reduce the network circulating current, reduce the network loss, improve the voltage and frequency stability, reduce the power grid failure probability, and improve the scale efficiency of power grid enterprises. In the future, with the gradual improvement of the power market and the introduction of new energy into the grid, power grid enterprises need to allocate reactive power resources more finely to improve the input-output efficiency of power grid enterprises [4].

In the era of electric power data, the resource allocation method of data aided decision-making is one of the necessary links in the upgrading and transformation of power grid enterprises. Power grid enterprises need to start from the demand of the power market, abandon the previous blind grid investment and construction methods, excavate the potential value information of power data, grasp the trend of regional industry and industry economic change, so as to form a refined power grid planning and construction strategy. In the rapid development and reform of the power industry, power grid enterprises need to study the operation efficiency optimization of power grid enterprises from the perspective of data, establish an effective incentive evaluation mechanism, accurately judge the power economic development trend, optimize the grid structure and resource allocation strategy, and drive the refined development of grid enterprises through data. This has important theoretical and practical significance for the stable supply of electric energy in the power grid industry and the steady improvement of the economic benefits of the power grid [5].

2 Metadata of Data Analysis Service

2.1 Classification of Data Analysis Services

Using the hybrid classification method, the data analysis services are classified in the following four aspects:

(1) Service form: classify the data analysis service according to different data analysis service forms;
(2) Construction unit: according to the different units of construction data analysis service, the data analysis service is classified;

(3) Business category: classify data analysis services according to different business categories involved in analysis services;
(4) Service release time: data analysis services are classified according to the year, month and day of the analysis service release [6].

Among them, there are three forms of data analysis services, which can be supplemented and improved according to needs.

Result dataset: it provides the query, subscription, push and other services of the result dataset, and supports the service consumers to quickly call the required result data through keyword, combined query and other forms.

Algorithm model: provide specific data analysis model and algorithm for service users to call according to their needs.

Analysis scene: provide accessible big data analysis scene page of finished product visualization.

2.2 Naming and Coding Rules of Data Analysis Service

The naming rules of data analysis service should meet the following requirements:

(1) A normalized full name of no more than 40 characters should be given;
(2) It should be named according to the format of "time + place + analysis object + analysis method", and the time and place can be default;
(3) Standard Chinese characters, numbers and English characters should be used instead of dialects, slang and obscure words;
(4) It should be unique and unambiguous within the scope of the unit;
(5) When the name of the self built data analysis service of each unit is the same or similar to that of the unified data analysis service of the headquarters, it shall be merged, renamed or revoked according to the specific rules and application conditions to ensure the uniqueness of the naming of the data analysis service.

Data analysis service code should include construction unit, service form, business category, creation time, serial number, etc.

Construction unit: representing the construction unit of data analysis service, adopting the company code in the human resource master data of State Grid Corporation of China, with a total of eight digits, including headquarters, provincial companies and directly affiliated units;

Service form: code according to the data analysis service form, with two digits in total;

Business category: code according to the first level business category and the second level business category;

Release time: the time when the analysis service is published, coded according to the data format of year, month and day;

Serial number: a six digit serial number generated automatically by the system when the service is published to ensure the uniqueness of the code [7].

2.3 Data Analysis Service Construction Method

(1) Service demand

The data analysis service requirements are managed in a unified way. The service demanders fill in the service requirements according to the demand reporting requirements, and enter the service construction link after being approved by the service manager.

The implementation method of data requirement reporting is to realize the management of new demand management, demand analysis summary, demand review management and other main functional components by expanding and enhancing development based on the current demand management situation of data analysis service.

New requirement: through enhanced development, the function of new requirement submission is realized to meet the requirement submission. After saving, the system will automatically record the information of the submitting personnel and the unique information code of the requirement number.

Requirement analysis and summary: the function of query, analysis and summary of requirements according to different dimensions is realized through enhanced development. Among them, the analysis function is to preliminarily screen and classify the submitted requirements, and supports the filling in of single or batch demand analysis opinions; the summary function is to classify and gather the qualified requirements according to the summary conditions.

Requirements review: the approval information can be filled in and returned for modification through enhanced development. The approval information should be filled in by category, recording the necessity, feasibility, risk and implementation details of the requirements in the requirement review; for the requirements that have not been approved, the historical records should be kept at the same time.

(2) Service Construction

The core of data analysis service construction includes result data set, algorithm model service and analysis scenario construction. Results the data set construction is based on the data platform sharing layer and analysis layer business data, based on the data platform technology component ability, through data processing, service development and other processes to complete the service construction; algorithm model service construction is based on the algorithm model component construction algorithm model service; business analysis scenario construction is based on the data platform self-service analysis component through the business After the business analysis scenario is formed by logical processing, the service is published [8].

2.4 Service Construction

Data analysis type: Based on the big data component of Alibaba cloud platform, access structured and collected measurement data. The structured data is output to online application area (ads, RDS, DRDS) after maxcompute data cleaning and summarizing, and the measurement data is stored in TS. Then according to different service types, the construction of services is completed.

Construction process of data analysis service: relying on the technical system of data middle platform, the analysis layer of data platform is constructed by using dimension model method to create public data layer and realize data exchange and sharing; open data application layer to meet personalized business application requirements, enrich data application of each business, and release data value; refine data service layer, construct public data service and support number According to the sharing of services and the rapid construction of data applications.

2.5 Service Launch

Data analysis service online is the process management of sharing and opening the data registered in the service API gateway. The data analysis service can be launched only after passing the security audit and confirming that it has been desensitized or does not contain sensitive information. Data analysis service users can query.

3 Typical Scenario Application of Data Analysis Service

Based on the support demand analysis of five special applications, including financial multidimensional, digital audit, material intelligent supply chain, integrated line loss and online state grid, the common demand characteristics are extracted from data demand (data type, demand frequency), data exchange, computing power and tools, data interaction mode, etc., and the application scenarios are classified as structured offline analysis and on-the-spot inspection Inquiry, acquisition, measurement and analysis, ad hoc query, real-time calculation and processing scenarios.

Ad hoc query scenario of structured offline analysis: the scenario focuses on structured data, and meets the ad hoc query requirements of applications after data access, data integration, analysis and calculation, scene display, etc. The off-line analysis and ad hoc query scenario of acquisition and measurement focuses on the acquisition and measurement data. After data access, analysis and calculation, data service and other links, it provides offline data query for applications to meet the ad hoc query needs of such applications. The real-time computing processing scenario is based on structured, measured and unstructured data, and provides real-time calculation results for applications after real-time data warehousing and real-time calculation.

4   Data Analysis of Bench Test

4.1   Functional Test Bug Distribution Analysis
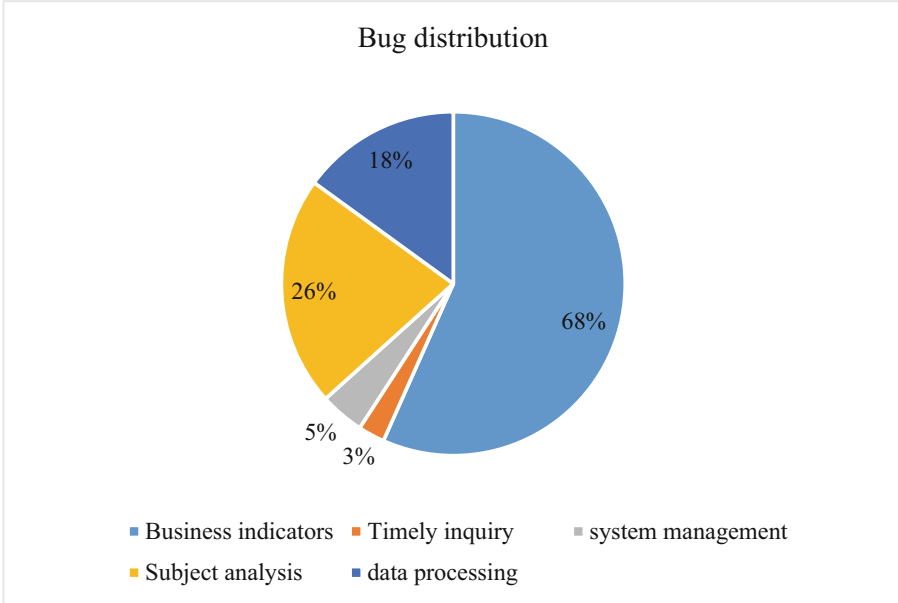

Fig. 1. Bug Distribution

A total of 387 test cases were designed in the functional test, of which 387 were effective cases, and 856 bugs were found. Among them, 96.9% were minor and general bugs, and there were no fatal bugs. All bugs have been fixed and closed. In the data accuracy test, 50 test cases were designed, 50 were effective cases, and 58 bugs were found. Among them, 100% were mild and general bugs, and there were no fatal bugs. All bugs have been fixed and closed. The function of the system meets the requirements of users (Fig. 1).

4.2   System Login Scenario Results and Analysis

Based on the different pressure of the design system, the compression performance of the platform is tested. According to the data in Table 1, the processing results of the platform system for concurrent login of users meet the test requirements, and there is no platform crash or obvious performance defects due to a large number of concurrent logins.

Table 1. System Login Scenario Results and Analysis Test

Scene	Execute Script	Number of Concurrent Users	Loading mode	Concurrency Srategy		Results of Enforcement		
				time interval	Synchronization Point Settings	Success	Fail	Average Response time
System login	System login	20	Start loading 10 people and load 2 people every 10 s	None	5	20	0	Satisfy
System login	System login	100	Start loading 20 people and load 5 people every 20 s	None	5	100	0	Satisfy
System login	System login	50	One time load	None	5	48	2	Satisfy
System login	System login	100	One time load	None	5	99	1	Satisfy
System login	System login	300	One time load	None	5	298	2	Satisfy
System login	System login	500	One time load	None	5	500	0	Satisfy

4.3   Performance Test of Log Analysis System

Through the analysis of the above test results, the log analysis system can meet the requirements of both performance and function (Fig. 2).

The former company analyzes the demand of operation and maintenance log. In terms of function, it realizes the fast and automatic analysis of operation and mainte-nance logs. It improves the efficiency of operation and maintenance personnel to deal with application system failure, and effectively ensures the efficient operation of the company's application system. In terms of performance, the system fully meets the existing daily log processing requirements. According to the growth rate of log in recent half a year, the computing power and storage capacity of the analysis system can meet the needs of log analysis and processing in the next two years.
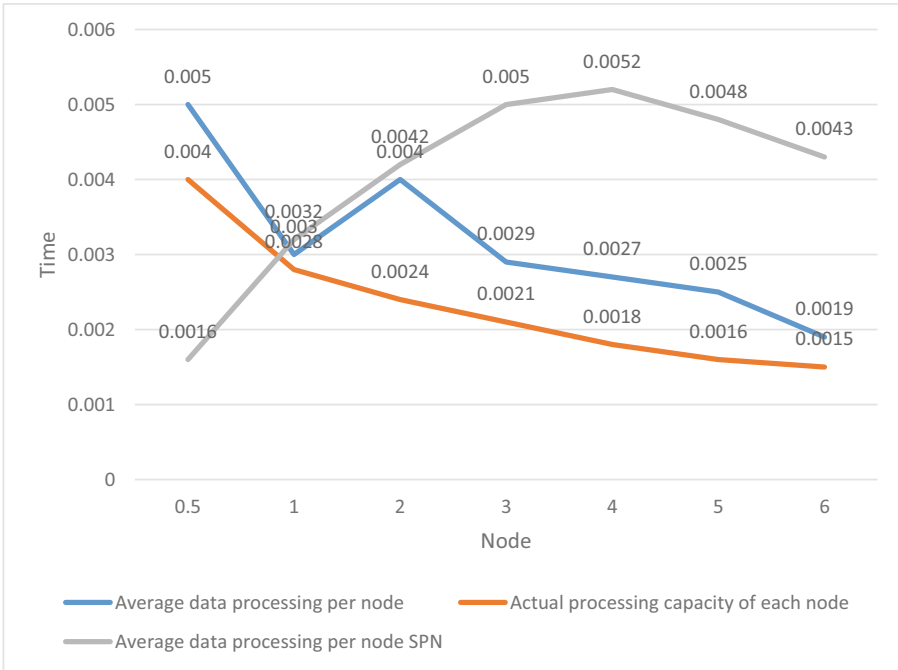
Fig. 2. Performance Test of Log Analysis System

5   Conclusion

The core value of data center is to precipitate common and reusable data assets. Through the rapid construction and iteration of data analysis application, it can realize multi-dimensional analysis and business exploration of business data, and simultaneously improve the intelligent ability of business application, so as to meet the requirements of providing agile and open data analysis and sharing for all disciplines, grass-roots units and external partners of SGCC The demand for service. With the continuous development of digital technology, the information level of power grid enterprises continues to promote, the power grid business continues to expand, data analysis services will also continue to improve with business expansion. The construction method proposed in this paper relies on the existing business status to make data analysis services have sustainable growth to face the analysis services generated by future business expansion. In the future, the construction of data analysis service still needs to be based on business needs, and the construction of data analysis service in this paper also needs to be continuously optimized and improved according to the actual work needs.

References

1. Lee, S.U., Soh, M., Ryu, V., et al.: Analysis of the health insurance review and assessment service data from 2011 to 2015. Int. J. Ment. Heal. Syst. **12**(1), 9–10 (2018)

2. Jing, L.I., Xiao-Hua, S., Yong-Ge, B., et al.: Research on data analysis and optimization in library self-service printing service— a case of shanghai Jiaotong university library. J. Acad. Libr. Inf. Sci. **11**(3), 19–21 (2016)
3. Walker, A.J., Curtis, H.J., Croker, R., et al.: Measuring the impact of an open online prescribing data analysis service on clinical practice: a cohort study in NHS England data (preprint). J. Med. Internet Res. **21**(1), 58–61 (2018)
4. Liu, L.G., ZLiu, H.X., Chen, W., Liu, Z.F.: Analysis of tripartite asymmetric evolutionary game among wind power enterprises, thermal power enterprises and power grid enterprises under new energy resources integrated. SCIENTIA SINICA Technologica **45**(12), 1297–1303 (2015). https://doi.org/10.1360/N092015-00244
5. Vergis, J.M., Vanderwall, D.E., Roberts, J.M., et al.: Unlocking the power of data. Lc Gc North America **33**(4), 1–2 (2015)
6. Badawi, O., Brennan, T., Celi, L.A., et al.: Metadata correction: making big data useful for health care: a summary of the inaugural MIT critical data conference. JMIR Med. Inform. **2**(2), 22–23 (2015)
7. de Montjoye, Y.-A., et al.: Unique in the shopping mall: on the reidentifiability of credit card metadata. Science **347**(6221), 536–539 (2015). https://doi.org/10.1126/science.1256297
8. Hong, S.D., Kim, J.T.: A construction of online luminance analysis service website. **12**(3), 45–48 (2017)
9. Zhang, B., Yu, L., Feng, Y., et al.: Application of workflow technology for big data analysis service. Appl. Sci. **8**(4), 1–5 (2018)