Research on Data Standard System Based on Data Middle Platform

Lin Liu^{1,a,*}, Yongqing Liu¹, Yiwang Luo^{2,b}, Shen Lin²

¹¹State Grid Information & Telecommunication Group Co., Ltd., Beijing, 102211, China

²State Grid Info-Telecom Greate Power Science and Technology Co.Ltd., Fuzhou, 350003, Fujian Province, China

^a Email: Lin-Liu@Sgitg.Sgcc.com.Cn, ^b Email: Lyw9841@163.com

*Corresponding Author

Keywords: Data Center, Data Standar, Research on Technical Standard System

Abstract: Manual Management is the Bottleneck of Data Asset Sorting, and the Ability of Data in Taiwan Has Become One of the Important Construction Contents of Data Center. in the Existing System Data, the Table Association is Mainly Carried out Manually, and It is Easy to Produce Data Inconsistency. in This Paper, a Method of Technical Standard System for Grid Data Resource Service is Proposed. the Method of Vectorization Realizes the Accurate Data Middle Stage with Low Calculation Cost by Constructing Inverse Function, Which Can Record the Changes of Data in the Whole Technical Standard System and Identify Multiple Sources. the Tracking Performance Test on the Existing Data in the Data Center Shows That This Method is Robust to the Change of Data Volume and Number of Layers, and the Middle Time of Data Meets the Actual Application Requirements. Compared with Other Data Middle Stage Technology, This Method is More Suitable for the Technical Standard System of Large-Scale Business Data in Power Grid Industry.

1. Introduction

The Work of Power Enterprises to Improve the Value of Data Resources and the Effectiveness of Management Has Been in Progress, and with the Development of Information and Network Technology and the Increase of Total Data Resources Long, Data Resource Utilization and Data and Service Sharing Adaptively Adopt a Variety of the Mainstream Technology Used. from the Large Mainframe Mode Adopted in the Early Power System Automation. from Distributed Network to Regulatory Cloud [1-5], Distributed Computing. Big Data Technology Establishes Enterprise Level Massive Data Platform, Big Data Platform [6-8], Establishing a Unified Public Information Model Based on Object-Oriented Online Monitoring or Distribution Network. Domain Data Platform [9-11], to Some Extent, Eliminates the "Chimney" of Business System Data Effect [12], Promoting Information Sharing.

This Paper Studies the Construction of Power Enterprise Middle Platform Based on Cloud Computing Environment. Analyze the Available Cloud Services of Alicloud, and Realize Key Functions Such as Metadata Management, Data Resource Management, and Shared Service Release in the Enterprise's Middle and Taiwan, Combining the Requirements of Data Management and Service System Construction in the Middle and Taiwan. Use the Data Integration and Big Data Computing Services Provided by Alibaba Cloud Big Data Computing to Complete the Comprehensive Analysis and Statistics of Data in the Business System of Midrange Management [13]. the Result Data Formed by Big Data Analysis is Used as a New Data Resource to Supplement the Middle Office, as the Basic Data of New Business. Cloud Based Construction of Power Enterprises in the Middle of Taiwan, Substantially Improve the Integration and Sharing Ability of Data Resources, to Provide Assistance for Business Rapid Response to Changes [14].

2. Analysis of the Support Ability of the Data Center

Cloud Computing Uses Computer Clusters to Form Unified and Coordinated Computing

Copyright © (2019) Francis Academic Press, UK

Resources, Providing Users with on-Demand and Convenient Access to Shared Resource Pools (Computing Facilities, Storage Devices, Applications, Etc.) [15]. the Core Service Layer of Cloud Computing Platform Provides Services Including Infrastructure as a Service, Platform as a Service and Software as a Service. among Them, Iaas Provides Entity or Virtual Computing, Storage and Network Services, Paas Provides Application Deployment and Management Services, Saas Provides Applications, and Enterprises Can Rent Them for Information Processing [16-17]. for the Sake of High Security, the Cloud Computing Services That Power Enterprises Build on Are Usually Based on Private Cloud. the Service Provided by Private Cloud Adopting Public Cloud Technology Has the Same Service Capability. the Middle Platform Needs to Deploy Business System, Store Business Data, Standardize and Merge Business Data, Use Big Data Analysis Service of Cloud Platform to Analyze Data, Etc., and Use Services from Iaas, Paas to Saas.

2.1 Metadata Management

Metadata is the "Data" of Data. It Describes, Interprets and Locates Data and Information Resources, and Makes Them More Convenient to Retrieve, Use or Manage Data [18]. Metadata Has the Functions of Indicating Storage Location, Resource Searching and File Recording, Which is the Basis and Premise of Data Sharing and Exchange. as a Large-Scale System, the Power Enterprise Middle Platform Needs to Strengthen Metadata Management for the Overall Management of Various Resources in the Middle Platform. through Metadata as the Top-Level Abstract Control Layer, the Data and Services in the Middle Platform Are under All-Round Control. as a Data Resource and Service Sharing Center, in Principle, It Does Not Involve the Direct Support of Specific Business Functions. Therefore, the Metadata of Zhongtai Strengthens the Management of Technical Metadata (without Emphasis on Business and Management Metadata). Metadata Mainly Includes:

(1) Basic Data Mode. That is, the Mode Information of the Basic Data Formed after the Business Data is Imported and Integrated into the Middle Platform. There Are Various Forms of Data Resources Managed by Middle Office, Including Structured Data Managed by Relation or Object Mode, and a Large Number of Unstructured Data.

(2) Data Service Description, Including Two Parts: Basic Data Service and Analysis Result Data Service Description. the Basic Data Service of the Middle Platform is Based on Rdb, Web Service, Iec 61970, Iec 62541 and Other Standard Services (Object Data is Based on Opcua Unified Address Space Management and Supports a Series of Access Interfaces), and the Analysis Result Data Service is Mainly in the Form of Micro Service. the Analysis Result Data Can Be Accessed through the Platform Basic Data Service after Being Extracted and Supplemented as Part of the Basic Data.

(3) Data Analysis Logic and Task Description. the Data Analysis Logic is Established by Configuration, and the Big Data Analysis Engine or Data Processing Task Customized by Ecs Performs the Corresponding Analysis Task.

2.2 Business Data Integration Management

The data of Zhongtai comes from various business systems such as safety production management, power generation control, power grid dispatching, power marketing, etc. According to the way the business system provides data, integrate the source data into the middle platform for management. Relational business data is imported to RDS in the middle platform through data extraction and transformation. The object-oriented data is standardized according to the enterprise public information model, including the object including path, object association and so on. These business data are raw data, which are published through RDS and object data access services. Enter the relational business data of RDS in China and Taiwan. If it is within the scope of enterprise public information model, write the data standardization collection task running on ECS for object-oriented transformation. The business system data after the object-oriented transformation and the directly introduced object-oriented data are regarded as the whole to be processed. According to the multi-level matching mode of hierarchy, association and classification, the data of the multi-source system can be matched and associated, the standardized integration of the integrated data can be completed, and the association obstacles between the business system data can be eliminated.

In order to improve the quality and value of data, the existing data in the middle platform is provided to external applications through service sharing. At the same time, in the platform, large data engine or self-developed data analysis software is used for comprehensive analysis. The analysis results form a part of the data in the middle platform, and the metadata of the result data is refined and integrated into the basic metadata of the platform. Through this kind of analysis and supplement cycle, enrich the data content of the middle office management. The analysis results are directly provided by service mode, which also facilitates the application of statistical analysis data. After the completion of business data integration management, Central Taiwan provides three types of available data: ①business original data; ②global object data after integration; ③statistical analysis data.

2.3 Standard System

Power enterprises in Taiwan to provide a variety of international standards, industry standards and enterprise standards of shared services to meet all kinds of data access needs. Relational, object-oriented and unstructured data are equipped with various types of services. In addition to the database table provider inherent in RDS, all types of data are equipped with microservices and web services (OSB WEB).

Object data access supports the first generation of general data access (GDA) of IEC 61790 and OPC UA service as the second version of IEC 61970 CIS. GDA provides object access semantic support similar to SQL language, which is convenient for navigation and query by class batch and by association chain. OPC UA is the core standard of the industrial Internet communication layer. It manages the object's model, real-time, historical and event data in a unified address space, and can efficiently transmit data in binary encoding mode based on the secure channel. When the data access efficiency of the application end is very high, OPC UA service can be used preferentially.

3. Institutions Optimization Design

The analysis, statistics and processing logic of big data are realized by writing data analysis script. Establish data processing process and add data processing nodes in the process. Write maxcompute SQL statement in data processing node, and submit it to business process after passing the test. Maxcompute SQL syntax is similar to SQL, which can complete data query, cache and use built-in function calculation. For complex processing that cannot be completed by basic statements, you can write a user-defined function (UDF) to implement the function, and then integrate UDF into maxcompute SQL. After submitting maxcompute job, there will be dozens to several minutes of waiting time for queuing and scheduling. Therefore, big data analysis and calculation are generally not used for real-time or quasi real-time data analysis (analysis and calculation with high real-time requirements can be performed by programming on ECs), which is most suitable for derivative statistics and indicator data.

4. Data Service Publishing

The calculation results are stored in maxcompute. The data access API can be created by configuring the name, grouping, path, protocol, access method, access content, return type and other information. The generated API can be registered to the data service platform for unified management and release. Data service can publish API gateway directly. Data service and API gateway provide secure, stable and easy-to-use data open sharing services for applications. Besides supporting application access through data services, high-value calculation results are exported to designated data sources for further display or analysis. The configuration process of data export is similar to that of data import. Only the roles of source and target can be exchanged. Exporting the calculation result data to RDS can substantially add new data content to the enterprise's middle platform, and also facilitate the reuse of business intelligence and other database based data. Typically, the application of data resource management and control and index analysis can directly access the analysis results, with the visual display of quickbi, greatly improving the application

construction speed.

References

[1] Ali A, Sharma S. (2018). Content based image retrieval using feature extraction with machine learning. International Conference on Intelligent Computing & Control Systems.

[2] Yu Y T, Lin G H, Jiang H R, et al. (2013). Machine-learning-based hotspot detection using topological classification and critical feature extraction. Design Automation Conference.

[3] Zhu P, Isaacs J, Bo F, et al. (2017). Deep learning feature extraction for target recognition and classification in underwater sonar images. 2017 IEEE 56th Annual Conference on Decision and Control (CDC).

[4] Xinting Y, Jianping Q, Chunjiang Z, et al. (2007). Construction of information description language for vegetable traceability based on XML and its application to data exchange. Transactions of the Chinese Society of Agricultural Engineering, vol. 23, no. 11, pp. 201-205.

[5] Aung M M, Chang Y S. (2014). Traceability in a food supply chain: Safety and quality perspectives. Food Control, vol. 39, no. 1, pp. 172-184.

[6] Mittal D, Reddy A, Ramadurai G, et al. (2018). Training a deep learning architecture for vehicle detection using limited heterogeneous traffic data. 2018 10th International Conference on Communication Systems & Networks (COMSNETS). IEEE, pp. 589-294.

[7] Li Ting, Wang Ping, Yao Zhen, Guo Jun. (2019). application of distributed database technology in power enterprises . Science and Technology Bulletin, vol. 35, no. 06, pp. 114-117

[8] Collins S R, Navathe S, Mark L. (2002). XML schema mappings for heterogeneous database access. Information and Software Technology, vol. 44, no. 4, pp. 251-257.

[9] LeCun Y, Bengio Y, Hinton G. (2015). Deep learning. nature, vol. 521, no. 7553, pp. 436.

[10] Schmidhuber J. (2015). Deep learning in neural networks: An overview. Neural networks, vol. 61, pp. 85-117.

[11] Li C, Xie X, Huang Y, et al. (2015). Distributed data mining based on deep neural network for wireless sensor network. International Journal of Distributed Sensor Networks, vol. 11, no. 7, pp. 157453.

[12] Hinton, Geoffrey E., and Ruslan R. (2006). Salakhutdinov. "Reducing the dimensionality of data with neural networks." science 313.5786, pp. 504-507.

[13] Zeng H, Zhou H, Srivastava A, et al. (2018). Accurate, Efficient and Scalable Graph Embedding. arXiv preprint arXiv:1810.11899.

[14] Greff K, Srivastava R K, Koutník J, et al. (2016). LSTM: A search space odyssey. IEEE transactions on neural networks and learning systems, vol. 28, no. 10, pp. 2222-2232.

[15] Wang Y, Huang M, Zhao L. (2016). Attention-based LSTM for aspect-level sentiment classification. Proceedings of the 2016 conference on empirical methods in natural language processing, pp. 606-615.

[16] Wang J, Yang Y, Mao J, et al. (2016). Cnn-rnn: A unified framework for multi-label image classification. Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2285-2294.

[17] Bruna J, Zaremba W, Szlam A, et al. (2013). Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203.

[18] Yan S, Xiong Y, Lin D. (2018). Spatial temporal graph convolutional networks for skeletonbased action recognition. Thirty-Second AAAI Conference on Artificial Intelligence.