The Application of Big Data Technology in the Power Dispatch Data Center of China South Grid

Qian Feng, Lin Qingnong, Gu Quan,
Research Institute NR Electric Co. Ltd, Nan Jing, China
qianf@nrec.com

***Abstract*—With the development of the smart grid business management, the requirements of the information sharing and collaboration in power grid are increasing. An efficient application analysis system is needed to integrate and manage a variety of information in the multi-level, large, and complex power grid, including grid model, data, graphics and etc. An integrated power grid operation data center based on the research of big data is presented. The data center has been applied to a real engineering in China southern grid, and received good results. Big data technology widely used in the power industry will bring revolution, and lead the development of smart grid to a new stage.**

***Index Terms*-- Big data, Integrated data center, Smart grid**

I. Introduction

Smart grid is a new type of power grid, which highly integrated the information technology, computer technology, communications technology and existing transmission, distribution infrastructure [1]. In recent years, smart grid has become the main direction of development of electric power system reform. It improves grid technology level, energy utilization efficiency and power grid reliability. It also promotes energy conservation, the use of new energy, the optimal allocation of resources and etc. Smart grid is a social project, which represents the future direction of development of power grid enterprises.

"Big Data" is a new concept in recent years, which captures, discovers and analyzes a large number, type and source of complex high-speed data. The terminology, "Big Data", can be traced back to the open source project "Nutch" in Apache Software Foundation. At the time, "Big Data" is used to describe the need to update the network search index for batch processing or analysis of large data sets [3]. In 2008, Nature published a special issue on "Big Data", which described the challenges brought about huge amounts of data from many aspects of the network economics, supercomputing, internet technology, biomedicine, environmental science. In 2011, Science published the special issue "Dealing with data", which deeply discussed the challenges of the data deluge. It noted that if more effective organization and use of these huge amounts of data, people would get more opportunities to promote science and social development [2-7].

With the development of smart grid, the digitizing, intelligence of power system is evolving. That brings more data sources. The numbers of data generated by the running power system are huge, and the data are fast-growing with rich type. Because the power system running date is fully complied with all the characteristics of big data, it is the typical type of big data. [8-11]

Big data technology applied in the smart grid is the requirement in power industry intelligence development. An integrated power grid operation data center is presented to meet the data sharing and unified processing. The date centre provides comprehensive data services to support the business application needs and maintain system good scalability. It integrates variety of information technology, including data warehousing, business intelligence, SOA architecture, ETL, data federation and etc.

II. RESEARCH & REALIZATION OF BIG DATA KEY TECHNOLOGY

The realization and application of big date key technology in China southern grid is "Integrated power grid operation data center". It is the application center to deal with the big data uniformly. Research and realization of the center are as follows. They are hierarchical modeling, source modeling, data services across safety zone, object register center and service register center.

A. Hierarchical Modeling

This template has been tailored for output on US letter-sized paper.

The hierarchical modeling technology of integrated power grid operation data center uses the primary power grid model as a core. The models of traditional energy management system and wide-area measurement system are built on it. Peripheral monitoring systems include water dispatching system, desulphurization monitoring system, power equipments status monitoring system, power quality monitoring system, environmental monitoring system and etc. They are supplementary systems for real time running monitoring system. Dispatching management includes dispatching operation management, network operation mode management, relay management, grid communications management and etc. The relationship between power grid model and panoramic information model in hierarchical modeling is shown in Fig. 1.

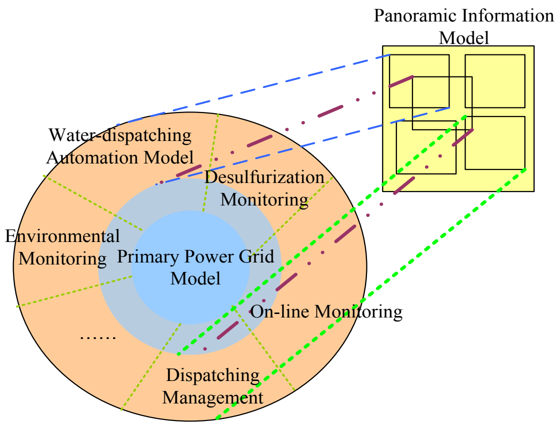

Figure 1. Relationship between Power Grid Model and Panoramic Information Model in Hierarchical Modeling

Based on the information model in integrated power grid operation data center, the modeling process is divided into three parts, including information modeling, horizontal level modeling, and longitudinal model merger.

1) *Information modeling*

The unified information model is created according to IEC 61970 CIM and IEC 61968 CIM. Based on the class "PowerSystemResource" described in the latest version of the IEC 61970 CIM and the actual situation of the power grid, the primary power equipment models are created. The models include the topology model, the dynamic model, the auxiliary device model and the construction equipment model supporting power lines. Measurement point model is created based on the class "Measurement" in CIM.

The increments of information model in business system are considered in hierarchy modeling, and the incremental models of each system are determined by the business system data. For example, when the relay information management system is integrated into the panoramic model, protect setting model is need to be established to describe relay equipments because the protect settings are import information in the relay information management system although the protect settings model are not built on IEC CIM. Another example, the model of desulfurization devices and information are also need to be built in panoramic information model.

2) *Horizontal level modeling*

Based on the unified information model, the hierarchical modeling begins from the power primary models which are derived from source modeling (energy management system, distribution automation system, substation automation system). Core power models and related graphics are built in the integrated power grid operation data center, and the object property and assets in various related business system are gradually added. Involving business systems include, but are not limited to energy management system, distribution automation system, wide-area pharos measurement system and etc.

Horizontal merger requires the conversion in different system mode, and has many processing steps, such as specifying data sources and the order of modeling. Based on the source model of the energy management system or substation automation system, a primary power grid model is constructed, and the objects and complementary attributes in other business system are merged.

In the process of horizontal model merging, various business data are needed to be aggregated. Most descriptions of those data are different with the panorama information model and must be standardized. The process of standardization is completed by conversion software. It includes data extraction, data transformation and date merging into the model. Individual conversion tasks for different business systems are established to achieve an orderly conversion task according to the execution order set. Business system can provide data though standard interface, and also can output model files such as BPA files. Conversion software in the integrated power operation data center is responsible for adapting different supplying mode. After horizontal merger, primary models from multiple systems and sharing data from business systems are integrated normatively.

3) *Longitudinal model merger*

Longitudinal model merger means obtaining the data model needed by this level from adjacent levels in integrated power grid operation data center.

How to deal with huge amount of data must be considered, and if the traditional TASE2 or 104 is used, the bottlenecks of capacity, performance and hard maintenance workload are hard to break. Therefore, the data center combines the regional, provincial and local dispatching model, and uses JMS message server. Based on that, a SCADA data transmission and collection system is designed to combine three-level dispatching systems. By combining regional, provincial and local models with unified coding technology, the whole power grid objects and measurement points are coded uniquely. So the longitudinal measurement point models in regional and provincial system are maintenance-free. Based on JMS message server, the longitudinal data transmission channels are established to solve the bottleneck problem by distributed acquisition technology, parallel processing technology and etc. System analog values and status values are designed as 400,000 points, and normal acquisition points are 800,000.

B. Source Modeling

Source modeling of the integrated smart grid operation data center means that the grid models in data center are built based on all source systems managed by the center. Source modeling technology consists of two parts, one is primary & secondary equipments source modeling and the other is application source modeling.

1) *Source system*

IEC 61850 standard unifies the substation description in the SAS. By complete SCL files, digital and intelligent substation information models can be used as the source system in primary and secondary power grid model required by master station.

Grid model described by SCL files can be correspond to the EMS model defined by IEC 61970. Descriptions in two areas are standardized, and after obtaining the source data from substation though data access interface, IEC 61970 model can be transformed from IEC 61850 model.

2) Source modeling based on substation model

When the integrated grid operational data center receives SCADA date from intelligent gateway directly, source modeling is realized by exchanging date and files with intelligent gateways as following.

(1) SCD files and picture files of primary & secondary equipment model are created in substation.

(2) SCD files are sent to the high-level master station (integrated operation data center or energy management system), and analyzed in it. The object GID generator in the register center of the master station is called to generate GID for each object, and the results will be returned to the substation.

(3) Object GID information files are received in substation and merged with SCD file. Through that way, modeling is completed in substation.

(4) For any requests from master station, intelligence gateway in substation will send the model files including GID information and picture files to the client uniformly. Intelligence gateway will communicate with the front-end system in the master station by IEC61850, and send real-time acquisition data. So source modeling is realized in substation and shared with master station.

3) Source modeling based both on EMS and substation model

If the integrated power grid operation data center is not responsible for building models for EMS, it usually does not face SAS directly and will use the primary & secondary power equipment model built by EMS. Because SAS does not provide the line model and just only provides the internal models in substation or power plants, the source of substation models should be EMS.

The key technologies of source modeling in power system resource are the conversion from SCD model to CIM model, the object identity management and incremental model maintenance. When the grid models in master station are established by SAS, SAS mainly generates SCD files, imports transferring information tables into communication system, and automatically configures transferring tables according to those information tables. The SCD model check, the process of SCD files converting to CIM models, and automatic generation of measurement and control points in model conversion are completed in master station. Model information and SCD files are transferred to the integrated power grid operation data center by FTP file transfer service or OSB. Conversion software can realize the conversion from SCD models to CIM models.

To solve the problem of sharing graphics between substation and master station, SVG format wiring diagrams complied with IEC 61970 in substation are need to be provided in the way of auto-generation or manual drawing in substation. That can realize graphically sharing and reduce the modeling work load in master station.

Through the coding service in the data center, object identity management will establish a unique code for all kinds of objects based on defined coding rules.

Incremental model maintenance is based on the object data provided by the source system. The object data are compared with the data in current data center and the difference information are obtained to update the optimization model. Incremental model maintenance is an important part of the master system model maintenance.

4) *Source modeling for physical devices (assets)*

The power grid model based on power system resource is the core model in integrated power operation data center. In addition, the physical device (asset) model described in the production management system is an integral part in comprehensive data analysis and application. That must be considered in source modeling.

C. Data Services across Safety Zone

The concept of safety zone is existed in china power grid uniquely. The power grid monitoring system is running in the safety zone I, DTS is running in the safety zone II and WEB is running in the safety zone III. Those safety zones are isolated by isolating devices, and then the real-time running system will not be destroyed by mistake operation. So the exchange technology for data services across the safety zone is needed. Based on “Mirroring" technology, client bus can call data services and application services on the other side of the zone across the isolation devices. By this way, the transparent transferring mode of mode data and picture files across safety zones is realized. The technologies of data services across safety zone are “service mirroring” and “data mirroring”.

1) Service mirroring

"Service mirroring" technology based on OSB is used to realize sharing data and service across safety zones. Service call and data sharing across safety zones include four steps as follows. They are service agent, request across safety zone, query implementation and data return.

(1) Service agent

After deploying mirroring service on OSB, it will firstly determine the model is in panoramic model database in this safety zone or not, according to the client’s query information. If the data in panoramic model database is in this safety zone, the data request is directly routed to data service. If the data in panoramic model database is in the other safety zone, the data request is routed to a special communications program I, and then the program I will process the data request.

(2) Request across safety zone

After getting the data request, the communication program I will form the packet from data request as packet files, and send them across the isolation devices to the communication program II in the opposite side of the safety zone. Communication program II in the opposite side of the safety zone will parse the files and send them to the OSB to simulate the client's request.

(3) Query implementation

After the OSB in the opposite side of the safety zone receives the data request packets, they will be sent to the data service in its own safety zone. Data service in this safety zone will parse the packets to obtain the corresponding data and return to the OSB, and the OSB will return the data to the communication program II.

(4) Data return

After receiving data result packets, the communication program II will form the date files and return to the communication program I in the opposite side of the safety zone across the isolation devices. Communication program I will parse the files, simulate the service request in its own safety zone and send to the OSB. At last, OSB will send the data to client.

2) *Data mirroring*

Through the “service mirroring” based on "data mirroring", service exchange is realized across safety zone. One of safety protection principles is to minimize the flow of data from safety zone III to safety zone II.

By using "data mirroring" technology, the supporting data for service are established according to the different demand in different safety zone. The data required to build the service in safety zone III come from safety zone II. That means the date flow from safety zone II to safety zone III is forward. In safety zone III, the data center processes those data for all kinds of services in a standardized way. Then the applications in each safety zone can obtain support services directly from its own zone. That is same as across zone service.

D. Object Register Center

The object registry center in integrated smart grid operation data center is a unified global object management tool, and it is the technical assurance for global object naming and coding. Object registry center is a system to realize global naming, which uses automatic method to ensure global ID (GID) and global naming unique. Object registry center can be used by multiple applications and multi-level dispatching centers.

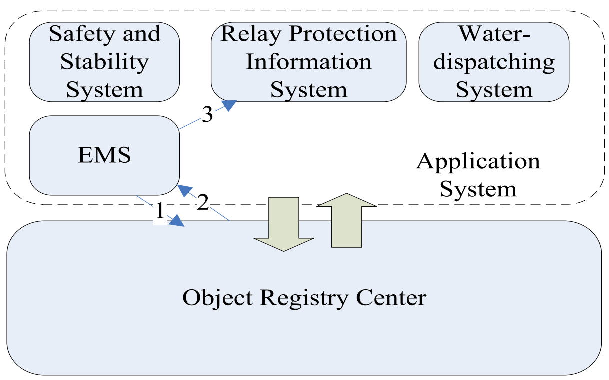

Figure 2. System Framework of Object Registry Center

Fig. 2 is the system framework of object registry center. Each business system exchanges information with the object registration center. If EMS queries an object data in relay protection information system, there is one way. Firstly, EMS requests object registry center to query object GID through the global name. Secondly, object registry center returns the corresponding GID. Finally, by the GID, EMS will query the relevant data from relay protection information system.

In a word, power system object registry center is a unified global object management facility to provide object registration, query, renaming, cancelling and other related services.

E. Service Register Center

Based on all kinds of OSB data services and applications, service register center is designed in the integrated smart grid operation data center to release, use and manage various services in a unified way. Service register center achieves service registration, query, locating, cancelling and other functions. It serves as a bridge between the service publisher and users to provide support for service publishing and client calling. The center also achieves standard data transmission and service calling method across business systems, modules, dispatching centers and stations, to ensure all types of operation data and business services can be exchanged and shared flexibly in whole power grid.

Service registry centers are established in safety zone II and safety zone III. For multi-level dispatching systems, the service registry centers are established in the middle-level dispatching system in order to achieve data longitudinal interconnection.

III. PRACTICAL APPLICATION OF BIG DATA TECHNOLOGY

Integrated smart grid operation data center was mainly developed and constructed by NARI-RELAYS Electric Co., Ltd. It has been realized in dispatching center in Southern Grid Company (SGC). SGC is responsible for investing, constructing and managing five province power grids in China south area, including Guangdong, Guangxi, Yunnan, Guizhou and Hainan province.

SGC integrated operation data center has realized the standardization construction of data center in dispatching system by establishing three platform tools including the unified data maintenance tool, the data integration management tool and the date comprehensive query tool. At the same time, it provides standard data access interfaces and services not only to meet the demand for current systematic data integration, but also to provide a unified data integration method for future power grid analysis application data and different business system data. Grid operation and dispatching management data include various types, such as power system information model, graphics, real-time/historical operating data, events, alarms, statistical analysis and management. Data center achieves the complete coverage and sharing of those data by data collection, integration and storage from time, space, objects, and other dimensions. In that way, the consistency of data models and graphics is ensured.

Through OSB and operation service buses in different-level power grid, the data integration platform in SGC integrated operation data center makes data exchange with decision management center, analysis module and other centers. It provides the unified data collection, panoramic data modeling, panoramic data validation, data integration and other services.

Logical framework of data integration platform in SGC integrated operation data center is shown in Fig. 3. It includes data source layer, data acquisition layer, data storage layer, ETL service layer, operation and maintenance, application services, data validation management, metadata management and etc.

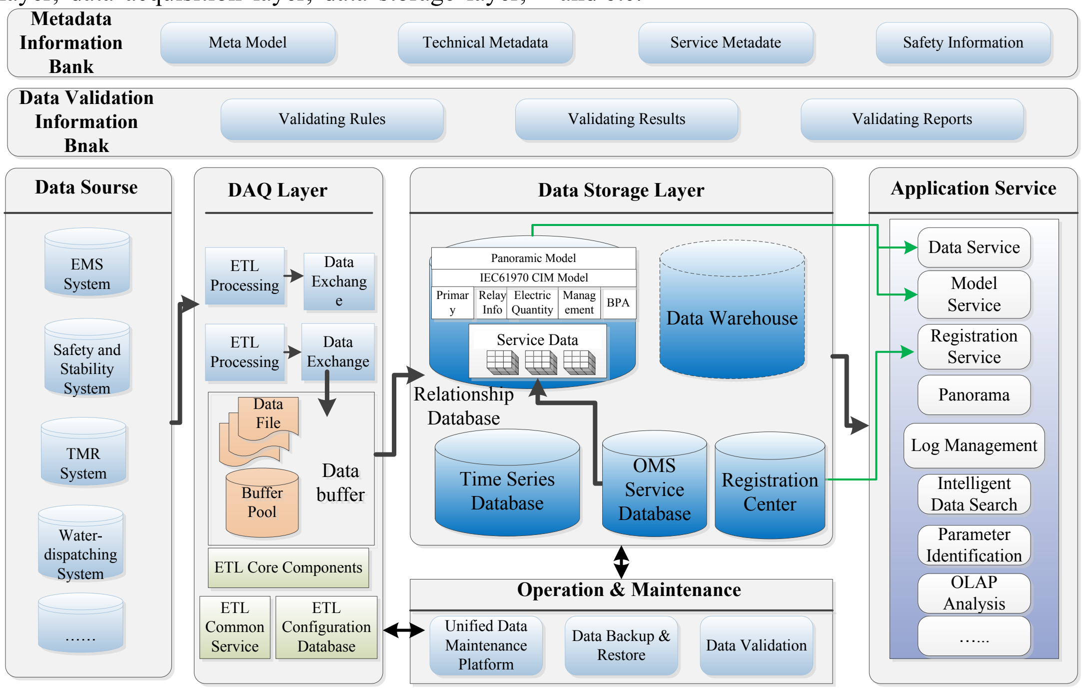

Figure 3. Logical Framework of Data Integration Platform

The key service of data center for prototype system is to realize the interfaces complied with IEC 61970 CIS. Now in China, the component interface specifications still in used are Universal Data Access (GDA), high-speed data access (HSDA), timing data access (TSDA) and general event and subscription (GES). In addition, because IEC 61970 CIS interface has upgraded, IEC 62541 (OPC UA) interface is used as next-generation interface standard. After deploying those services, SGC integrated operation data center already has a complete, satisfactory data service interface complying with IEC 61970.

Practice shows that the performance of CIS and OPC UA service published by CIS server in SGC integrated operation data center can meet the actual needs of the project. Service performance indicators are shown in Table I.

TABLE I. PERFORMANCE OF CIS AND OPC UA SERVICE

Service	**Performance (Best access mode)**
CIS GDA Query object by class	Get object description≥ 4,000/s Support concurrent access≥ 20
CIS HSDA Query real-time data	Get object description≥ 2,000/s Support concurrent access≥ 20
CIS TSDA Query historical data	Get object description≥ 12,000/s Support concurrent access≥ 20
OPC UA Browsing service	Get object description≥ 4,000/s Support concurrent access≥ 20
OPC UA Query real-time data	Get object description≥ 20,000/s Support concurrent access≥ 20
OPC UA Query historical data	Get object description≥ 12,000/s Support concurrent access≥ 20

IV. CONCULSION AND PROSPECT

The construction of SGC integrated operation data center has realized and verified the big data key technologies, including hierarchical modeling, source modeling, data services across safety zone, object register center, service register center, longitudinal data exchange and others. The integrated power operation data center has laid a good foundation for the big data technology applied in Chinese smart grid. In a word, the integrated power operation data center will be popularized.

In the next step, cloud platform and cloud compute are focus. Due to the generation of massive data in power system, if big data and cloud platform are combined, the actual system will has the advantages of low cost, high reliability, parallel analysis and others. It will make the system more practical and lead the development of smart grid to a new stage.

REFERENCES

[1] Peng Xiaosheng, Deng Diyuan, Cheng Shijie, Wen Jinyu, et al, "Key Technologies of Electric Power Big Data and Its Application Prospects in Smart Grid", *Proceedings of the CSEE,* vol. 35(3), pp.503-511, 2015.

[2] Birney E, "The making of ENCODE: lessons for big-data projects", *Nature*, vol. 489, pp.49-51, 2012.

[3] Gartner, (2014, Aug. 17), "Top ten strategic technology trend for 2012". Available: http://www.gartner.com.

[4] Zhang Dongxia, Miao Xin, Liu Liping, et al, "Research on development strategy for smart grid big data", *Proceedings of the CSEE,* vol. 35(1), pp.2-12, 2015.

[5] Zhang Suxiang, Zhao Binzhen, Wang Fengyu, et al, "Short-term power load forecasting based on big data", *Proceedings of the CSEE*, vol. 35(1), pp.37-42, 2015.
[6] UN Global Pulse, (2014, Aug. 10), "Big data for development: challenges & opportunities". Available:
[7] http://www.unglobalpulse.org/projects/Bigdata Development.
[8] Li Guojie, Chen Xueqi, "Research status and scientific thinking of big data", *Bulletin of the Chinese Academy of Sciences,* vol.6, pp.647-657, 2012.
[9] Meng Xiaofeng, Ci Xiang, "Big data management: concepts, techniques and challenges", *Journal of Computer Research and Development,* vol. 50(1), pp.146-169, 2013.
[10] Li Guibing, Luo Hong, "Intelligent data analysis under the background of big data", *Science & Technology Information,* vol. 30, pp.11-12, 2013.
[11] Zhang Wenliang, Liu Zhuangzhi, Wang Mingjun, et al, "Research status and development trend of smart grid", *Power System Technology,* vol. 33(13), pp.1-11, 2009.
[12] Sun Bolin, "Big data technology and its application in power industry", *Electric Age*, vol. 8, pp.18-23,2013.