Design and Research of Big Data Platform Framework for Power Enterprises

Rui Ma\textsuperscript{1,2,3}, Weibing Li\textsuperscript{4}, Nengwu Ma\textsuperscript{1,2,3}, Xiong Zhang\textsuperscript{1,2,3} and Hang Zhang\textsuperscript{1,2,3}

\textsuperscript{1}Changjiang Survey Planning Design And Research Co., Ltd., Wuhan, Hubei, 430010, China
\textsuperscript{2}Changjiang Spatial Information Technology Engineering Co., Ltd., Wuhan, Hubei, 430010, China
\textsuperscript{3}Hubei Province Engineering Technology Research Center for Water Resources Information Perception and Big Data, Wuhan, Hubei, 430010, China
\textsuperscript{4}China Yangtze Power Co., Ltd., Yichang, Hubei, 443002, China

*Corresponding author’s e-mail: 58122712@qq.com

Abstract. Big data technology is one of the most fashionable computer science and technology. Through the rational use of big data technology, traditional enterprises can accelerate the promotion and transformation of business informatization and intelligence. After long-term accumulation and precipitation, the power industry has accumulated a large number of structured and unstructured data of production, operation and management. Based on the actual demand of the electric power industry and the characteristics of the existing information system and mass data, this paper studied the big data platform framework system of electric power enterprises, and carried out the big data framework design from the overall architecture, technical architecture, data architecture, logical deployment architecture and other aspects. The overall design idea conforms to the development concept of "big data +" and "Internet +", and has a good technical foundation and application support, which can be widely used in the design and construction of power big data platform.

1. Introduction

In Big Data: A Revolution That Will Transform How We Live, Work and Think\cite{1}, compiled by Viktor mayer-schonberger and Kenneth Cukier, Big Data refers to the analysis and processing of all Data without the shortcut of random analysis (sampling survey). Big data has 5V characteristics: Volume, Velocity, Variety, Value, Veracity\cite{2}. Research on big data has been studied earlier in foreign countries, and there are also some mature solutions or software, including Apache Hadoop\cite{3}, Google file system (GFS), Google MapReduce, Google BigTable, Google Chubby\cite{4} and so on.

With the continuous advancement of information construction, the information system is constantly increasing, and the application and promotion degree is constantly deepening, the electric power enterprises have accumulated a large number of structured and unstructured data of production, operation and management. Big data in electric power industry, domestic also carried out a large number of research and practice of application of Zhu Hongbin\cite{5} to study the electric power data security management, Zhao Chuan\cite{6} to study the electric power data network security analysis, He Anhong\cite{7} of the electric power enterprise data environment research on operations management, Li
Jialiang[8] studied large material management data environment, but are considered a big data applications or research, there is no systematic research on the design of the big data platform. Based on the actual needs of power enterprises in the basin, this paper aims at the problems such as insufficient data sharing among systems, data resource redundancy and dispersion, and the urgent need for big data technology to mine the potential power of the company’s data value, and carries out the construction and design of big data platform from the overall architecture, technical architecture, data architecture, deployment architecture and other aspects. Through the design of the big data platform architecture for the construction of high reliability, high security, high performance, high expansion of the big data platform to provide technical support.

2. Overall architecture design
According to the actual demand of power industry management units and production units for data management, processing and application, this paper makes an overall design of the system architecture and functional composition of big data platform by combining with the demand of data governance and control system, security guarantee system and platform operation and maintenance system, as shown in figure 1.

![Figure 1. Overall architecture](image)

Data acquisition realizes the collection of data of each system of the power department and its subordinate power plants and ladder dispatching center. Collection data types include: management information system and external system of relational data, time series data and high-frequency waveform data, equipment operation of events or log data, document drawing data, audio and video data, and other types of semi-structured and unstructured data.

Data storage function realization of electric headquarters and subordinate units of the system data stored centrally, for relational data, temporal data and semi-structured data, unstructured data storage capacity, as well as to all kinds of data on demand (generally is one of the largest thermal data for the latest or usage) placed in the data cache, provides rapid access data and combining with the data in real-time message middleware or quasi real-time push.

The data storage includes the basic data storage and the subject data storage, in which the basic data storage is used to store the collected original data, the subject data storage is based on the architecture of the data center, and the data storage is carried out by the subject to achieve the storage of data warehouse, data lake or graph data.

Data computing provides a variety of computing engines such as SQL computing, batch computing, streaming computing, parallel computing and graph computing, and realizes real-time and non-real-time data computing for the data of the system.

Data analysis USES analysis tools, statistical analysis, data mining, deep learning and machine learning algorithms to analyze the data stored in the big data center. Specific functions include...
algorithm model library, algorithm engine, statistical report, mining and analysis, and self-service data exploration.

Data service provides data according to the business domain through the data access interface. The specific functions include data access, data search, computation invocation, service management and service monitoring.

Data asset management realizes metadata management, measurement data management (for production data), master data management, coding management, data model management, data standard management, data label management, data-catalog management, data quality management, data blood relationship management, data value management and other functions.

Data security management is the basic function of big data platform, including data encryption, data desensitization, data right verification, data audit and data authority control, etc., to ensure the security of system data.

Platform management realizes the functions of big data platform such as operation status view, configuration information management, log information view, authority management and system audit.

3. Technical architecture design

For big data platform overall architecture of the system function and the function demand, based on the mainstream technology route and the overall optimal principle, the paper USES the many kinds of big data technologies, to systematic and effective combination of various technology components, meet the demand of big data platform overall function, forming the system of the technical architecture is shown in figure 2.

Relational database supports transaction management, lock mechanism, data integrity check, create and use index, support SQL query, use cache technology, with monitoring management and backup recovery function.

As a branch of the development of DataBase system, Real Time DataBase is generated by combining DataBase technology with Real Time processing technology. It can directly collect and obtain all kinds of data in the operation process of an enterprise in Real Time and convert them into effective public information for all kinds of businesses. Real-time database products can be selected including PI, eDOS, and PGIM products.

ETL tool has script management, transformation management, task management, mapping management, job management, connection management and process management functions; Support data extraction from different data sources, including business system database, data files and data.
interface/communication protocol, etc., to meet the requirements of various business scenarios; For data extraction, cleaning, conversion and loading into the big data platform. Parallel computing platform is used to store management information data and form data warehouse, which supports data analysis and mining.

The big data computing platform adopts the big data platform based on the distributed storage computing mode to form the complete data storage, processing, calculation and analysis functions. It mainly includes functional modules such as distributed storage, distributed resource management, distributed computing engine, data acquisition and data analysis, as shown in figure 3.

![Figure 3. Big data computing platform](image)

Big data computing platform is the most core part of the architecture, including distributed file system, distributed store management, distributed resource management, nosql database, sequential database, distributed computing engine, time-series database, data acquire, data analysis, data collection, platform management, and so on. Among them, distributed store management includes search engine, data store; distributed computing engine includes job scheduling, dag compute, streaming process, distributed computing framework, memory compute; data acquire includes sql query, data cache; data analysis includes data mining, graph process, r language; data collection includes message queue, database etl, log collect tool; platform management includes software management, configure management, fault management, perform management, safety management, backup management, users management.

The big data platform provides visual business intelligence tools for designing and building the OLAP Cube model.

The report platform supports tables, charts, perspectives, drill-down reports and decision reports, etc., so as to realize the production and operation of various comprehensive reports.

There are two ways to integrate third-party algorithm model library and power big data platform, as shown in figure 4.

![Figure 4. Integrate third-party algorithm model library and power big data platform.](image)
Method 1 is to upload the relevant data stored on the power big data platform to the external algorithm platform after data desensitization. The external algorithm platform reads the uploaded data. After the relevant algorithm is used to process the data, the processing results are transmitted back to the power big data platform through the interface.

In method 2, the external platform provides a separate algorithm library, which is deployed on the power big data platform. The algorithm library uses the power big data platform to calculate resources, reads the corresponding data for calculation, and provides the calculation result query service through the data service.

Data visualization tools provide rich charts and styles, dynamic decorative effects, fast style conversion; Complete data analysis and visual page design with full drag-and-drop and graphical configuration; To realize the visual transformation of data, to form clear and understandable images and charts, to intuitively convey key aspects and features, and to achieve in-depth insight into a large and complex data set.

The data asset management platform and data service platform are built on the basis of container and microservice platform. The data service platform can provide data services to the outside world through the micro-service platform and the SOA information integration platform, and it can also provide its own special interface, file transfer and other data service interfaces.

4. Data architecture design

Data types in power enterprises mainly include relational data, measuring point data (including routine time series data and high-frequency waveform data), event log data, audio and video data, document drawing data and other semi-structured/unstructured data. These data are generated by professional systems such as management information system, power production system and video monitoring system, which are uniformly collected into the big data platform for centralized storage, management and application, as shown in figure 5.

5. Logical deployment architecture design

According to the characteristics of power unit composition and business development, as well as the needs of data processing and application, the deployment of big data platform can adopt the two-level deployment mode of management unit and production unit to realize the hierarchical and hierarchical management of enterprise data resources. The aggregation and distribution of component resources and data resources of management unit and production unit platform are realized through business rule filtering and distribution task scheduling mechanism, as shown in figure 6.
Power plant data include monitoring, stability control, phase angle measurement, equipment monitoring, dam safety detection, electrical energy measurement, water situation measurement, protection information, traveling wave ranging, fault recording, etc. The data of cascade control center include monitoring, stability control, phase angle measurement and reservoir operation, meteorological forecast, geological early warning, electric energy measurement, protection information, fault recording and so on. After being processed by the data acquisition and processing unit, the data are stored in the real-time database of zone II, and then transmitted to the real-time database of zone III through the isolation gate. Finally, the real-time database in zone III of the production unit is transferred to the big data platform of the production unit, and then to the big data platform of the management unit through the big data platform of the production unit.

After the data of the production unit is transmitted to the real-time database of zone III through the isolation gate, the data can be transferred to the real-time database of the management unit first, and then extracted into the big data platform of the management unit by the ETL tool of the management unit. The management unit can send down the data to each production unit according to the business needs.

6. Conclusion
Based on the actual demand of the electric power industry and the characteristics of the existing information system and mass data, this paper carries out the construction and design of the big data platform for electric power enterprises from the aspects of the overall architecture, technical architecture, data architecture and logical deployment architecture. The construction of the big data platform will provide unified big data sharing and analysis capabilities, carry out forward-looking forecast and analysis on all kinds of businesses, provide unified decision analysis support for users at all levels, and improve the ability of data sharing and flow.

The design of big data platform framework for electric power enterprises conforms to the development concept of "big data +" and "Internet +". It has a good application foundation and can be widely used in the design and construction of electric power big data platform. In the next step, we will combine with the use of cloud computing, Internet of things, artificial intelligence, mobile applications, chain blocks, such as a new generation of information technology, on the basis of the existing big data platform, carry out independently controllable energy information platform design and construction, speed up the integration of information technology and the electric power industry, improve the clean energy planning, construction, operation, management informatization and intelligent level.

7. References
[1] Viktor Mayer-Schonberger, and Kenneth Cukier, “Big Data: A Revolution That Will Transform How We Live, Work and Think,” Hodder & Stoughton, 2013
Acknowledgments
This study is supported by The National Key Research and Development Program of China: Dynamic simulation and intelligent decision of watershed water resources management driven by big data (No.2018YFC0407904)