Study on Optimization of monitoring and early warning system for hydropower plant equipment safety implementation based on big data Era

Wu Xiaodong, Zhang Wenyuan, Fu Jun, Tang Shan, Dai Jianjun, Xu Tao

Yingxiuwan Hydropower General Plant of State Grid Sichuan Electric Power Company, Wenchuan, Sichuan, 611830, China

E-mail: wux_dong9236@sina.com

Keywords: big data; cloud computing; industrial internet; intelligent Hydropower Station

Abstract: Combining with the latest progress of information technology and the new industrial revolution process, and fully utilizing the big data and cloud computing technologies widely used by major technology companies, under the background of speeding up the construction of power industry internet, this paper analyses the current situation of data utilization of traditional hydropower stations, and puts forward the application of big data of cloud platform. The scheme of building intelligent hydropower station can greatly improve the data integration and analysis ability of hydropower station, avoid data waste, fully mine information from massive data, study equipment status and trend as far as possible, improve the level of system intelligence and self-diagnosis function, and enhance the efficiency and reliability of equipment operation.

1. Current status of data utilization in traditional hydropower stations

At present, the hydropower's stations in operation basically adopt layered and distributed system layout, and rely on computer monitoring system to collect and store data centrally and carry out calculation and processing. In the next layer, there are many small systems, such as governor, excitation, protection, auxiliary machinery, oil, water and gas, which are responsible for measuring or calculating their respective measuring point data and transmitting them to the monitoring system. There are data barriers. If the same data is used in different systems, they need to be collected separately. Thus, there may be deviation in data consistency. In addition, due to the current status of equipment manufacturing technology in the industry and the constraints of relevant regulations and specifications, most of the data acquisition in various systems of hydropower stations is still transmitted through analog quantities, and their transmission efficiency, sharing degree, etc. All of them are low, far from meeting the needs of data interconnection and rapid real-time sharing. At present, the main functions of hydropower station monitoring system are still concentrated on the real-time control level that is, relying on real-time.

After the relevant logical operation of data, the corresponding control and adjustment tasks are completed. The historical data are only stored, and the deep mining analysis and correlation operations are very few. This results in the massive data sleeping in the hard disk. It does not play any role in promoting the management of power plant equipment and economic operation, and does not reflect the value of data assets.

2. Technical Scheme of Data Processing Platform for Hydropower Station

2.1 Data Storage Computing Layer

2.1.1 Data storage

Data storage mainly completes the fusion storage of multi-source data. Data storage is mainly based on distributed file system (HDFS) and column database (HBase, etc.), supplemented by memory database to realize the output storage of streaming computing, and through relational database to realize the storage and management of metadata, user, permission configuration and other data. Data Computing mainly includes offline computing services, real-time computing

  DOI: 10.25236/aisct.2019.069

services, distributed database services for massive structured data processing, etc.

The industrial data management platform based on big data implements data extraction from the edge, data and model downloading, data cleaning and quality discrimination, data asset management, and provides different database storage according to different types of data.Such as real-time database, relational database, unstructured data, and unified identification of data.In view of the heterogeneity and massive scale of industrial data, the appropriate storage mode is selected according to each data characteristic, and the database technology (SQL/NoSQL/...) is adopted. / Graph Relations/Distributed Indexing/Memory Data/File Data, etc.) to achieve a unified data storage solution to meet the requirements of heterogeneous, massive data with high concurrency, high throughput, low latency and high availability of data in industrial scenarios

2.1.2 Data Processing

Data processing refers to the deep reprocessing of data in data management system by real-time and batch analysis, combined with professional industry operators and AI artificial intelligence, to generate derivative data and index data, and to enhance the analysis significance of data in data management system.

The processed data can bring more professional analysis and decision-making utility to the upper application, greatly reduce the development complexity of the upper application, and conform to the concept of "separation of number and use".

2.2 Data Service Layer

Data service layer mainly provides data resource service and data computing service for large data platform. Through standardized open access to data services, applications and data are separated. Service layer adopts micro-service architecture as a whole. Data resources and data computing should be encapsulated into SOA and RESTful architectures to provide services to the outside world. Common RMI, JMS, HTTP, Hessian, thrift and other protocols can be implemented. Based on standardized interfaces, API encapsulation of data access and data computing methods can be realized, so as to facilitate the value-added development and utilization of upper application, reflect data value, and support the machine. The data service layer focuses on the platform and services of large data application development and data analysis and mining. The whole system is based on the underlying distributed batch processing framework, distributed stream data real-time framework and machine learning algorithm library and other related infrastructure. It should provide various application development suites to meet various development needs. Data service layer includes data development service, data algorithm service and data visualization service.

Based on the design concept of "separation of data and usage" of the platform and the framework of micro-service, data service is encapsulated and data management and utilization are realized,including data writing, query, quality management, extraction, cleaning, model association, application invocation and other data services, to fully meet the data invocation, storage, so as to achieve data separation by means of services.At the same time, data services can be deployed independently, and different types of databases can be adapted according to business needs. Data processing and applications based on databases are all applicable to corresponding data through relevant data services.

2.2.1 Information Identification Service

Information identification service is also a high abstraction of industrial information model and the core of industrial information model.By identifying the industrial information model, an industrial information identification tree is formed. Each identification tree is a logical abstraction of an industrial business domain,by expanding the branches and leaves of the tree, the object in the field of industrial data can be identified comprehensively and uniquely, and the "ID number" can be provided for the identification, effective utilization and processing of the data.

The instantiation of industrial information identification is the instantiation of industrial information model, which can bring the following advantages.

1) The aggregation of markers can form a marking system.

2) The instantiation of identification can form industrial information model.

3) The flexibility of identification, that is, the flexibility of data services;

In order to satisfy the identification management of mass industrial information, the system provides the identification service of industrial information, including the management of identification rules, the inspection of identification structure, the construction of identification and the query of identification.

2.2.2 Data Access Service

The open RESTful API interface is used to provide various data access services through the data asset management system service access gateway.

The standard Elastic Search retrieval specification is used to provide access interface for data asset management and to achieve efficient retrieval of data assets.

2.2.3 Data Synchronization Service

The platform includes data synchronization function, which can easily realize real-time data synchronization between two isomorphic data centers and ensure data synchronization storage between different levels of data centers. Data centers need to support customized source and target to achieve data synchronization between two isomorphic data centers. The definition of synchronous data has the following functions:

(1) Custom synchronization grouping: the synchronization requirements (synchronization frequency, etc.) of user-defined grouping in the target data center and the encoding contained in the grouping are set. The data center synchronizes the encoding data contained in the grouping according to the user's requirements.

(2) Data synchronization at coding level granularity: Data synchronization between different groups and in the same group does not affect each other;

(3) Incremental data synchronization: The data center itself controls how to obtain incremental data without requiring any changes from existing systems.

(4) Efficient Synchronization: Supporting Multi-instance Synchronization

(5) Breakpoint Continuous Transmission: When transmission fails due to abnormal network conditions, it can be automatically retransmitted.

(6) Synchronization logs: The amount and time of each synchronization can be displayed in the logs.

(7) Supporting cross-isolation synchronizatio

(8) Data Access Authority Management Service

(9) Using RESTful API, data rights management services are provided through data asset management system service access gateway, including data query, storage, editing, authorization, authentication, field control and other rights management.

2.3 Data Governance

Data governance runs through all stages of large data processing process. By collecting, querying and analyzing metadata, it establishes data audit strategy, completes data traceability, data map, data consanguinity analysis and other functions, and realizes the whole process control of data life cycle.

Data quality management refers to the quality control of industrial data according to the definition of data standards and quality rules, such as data quality verification, data retransmit, data cleaning and so on, and the generation of data quality reports according to the rules.

Effectively identify all kinds of data quality problems, establish data supervision, form data quality management system, monitor and reveal data quality problems, provide detailed queries and quality improvement suggestions, comprehensively improve data integrity, accuracy, timeliness, consistency and legitimacy, and reduce data management costs,reduce decision bias and loss caused by unreliable data.

2.3.1 Data Quality Cleaning and Discrimination Contents

1) Data Quality Verification

Real-time data quality verification is used to ensure the integrity and continuity of the data collected by the equipment on site, and to monitor the status of the acquisition interface and the usable status of the interface components by monitoring the status of the acquisition system.

2) Data Retransmit

Through data quality checking, the quality of data can be determined according to business rules, and data can be recalled according to the configuration of rules for lost data.

3) Data cleaning

According to the definition of industrial data quality execution standards and rules, the quality inspection and data cleaning of all kinds of collected industrial data are carried out. Data cleaning methods include: time axis unification, physical rule elimination, data interpolation, data filtering and classification, data anomaly elimination, duplicate record elimination and so on.

4) Data Quality Report

According to the quality rules, the quality of statistical data is regularly analyzed, and corresponding data quality reports are generated to urge data managers to improve data quality.

5) Data cleaning and discrimination

Due to sensor failures or network transmission errors, some abnormal data quality will be included in the original production data. This kind of abnormal data will be identified by the data quality discriminant module according to the quality discriminant rules, and the data that does not conform to the quality rules can be distinguished.

2.3.2 Data Quality Cleaning and Discrimination Method

1) Abnormal Value Processing

Through anomaly data detection and anomaly data correction processing, referring to the data before and after the data anomaly points, the anomaly data points can be processed.

2) Abnormal Data Detection

Abnormal data detection is realized based on monitoring statistical analysis detection, probability distribution abnormal data detection, box diagram analysis detection and so on.

3) Abnormal Value Correction Method

Abnormal data are corrected based on digital filtering methods such as mean filtering, median filtering and mode filtering.

4) Missing Value Processing

The missing values are filled by interpolation methods, such as mean interpolation, median interpolation and mode interpolation.

5) Statistical aggregation

The data errors such as noise and disturbance produced by measurement are reduced and eliminated by filtering and other algorithms.

6) Common Data Cleaning Function

Such as: comparison, merge, association, label, duplication, etc

2.4 Data Security

Data security also runs through all stages of large data processing. Data security in this chapter mainly considers all kinds of security in large data processing environment, focusing on platform security and data security itself. Construction teams with strong technical force, complete organizational structure and more skilled workers should be selected to ensure the smooth implementation of construction safety, quality and progress. In the process of construction, we should strengthen communication, coordination and cooperation with relevant departments in design, construction and local areas, timely adjust the construction progress, timely command and allocation of construction site personnel and machinery and equipment. In the process of construction, construction safety must be put in place. From the construction unit to the construction unit, special personnel should be set up to manage safety problems. Safety personnel should be given the right to stop all construction with potential safety hazards.

Work to reduce accidents caused by safety problems. After unit operation, the operation management unit should formulate corresponding management system, flood season safety flood

control plan, strengthen the inspection and maintenance management of engineering and equipment, strengthen the training of operation and maintenance personnel, and ensure the normal and safe operation of power station.

Data security refers to the comprehensive protection of confidentiality, integrity, availability, persistence, authentication, authorization and non-repudiation of user data information assets. Protecting users'data and information assets is the core of cloud platform security strategy.Cloud platform should follow the advanced standards of data security life cycle management, and adopt excellent technologies, practices and processes in identity authentication, authority management, access control, data isolation, transmission security, storage security, data deletion and physical destruction to ensure tenants'privacy and ownership of their data and the right of control is inviolable, providing users with the most effective data protection.

2.4.1 Access Isolation

1) Identity authentication and access control

Cloud platforms need to have a unified identity authentication service, centralized management of users'management rights and access rights to cloud resources. Administrators can ensure the security of user accounts and reduce the information security risk of tenants by assigning users minimum privileges, setting up high-security login authentication, password policy, access control list and other operations.

2) Data isolation

Cloud platform isolates the network of different tenants through virtual private cloud (VPC), which ensures that data between different VPCs will not be obtained beyond the authority. Through VPC, tenants can fully control their own virtual network and achieve complete isolation between different tenants in two or three layers of network. On the one hand, VPC and traditional data center of tenant's intranet are interconnected by VPN or cloud dedicated line, so as to realize smooth migration of tenant's application and data from tenant's intranet to cloud,On the other hand,the security group function of VPC is used to configure security and access rules on demand to meet the needs of more fine-grained network isolation for tenants.

2.4.2 Transmission Protection

In the process of data transmission, the special longitudinal encryption device is used to protect the data, so as to protect the security of key data in the power industry in real time.

2.4.3 Data Destruction

If the user terminates the service and the user proposes data deletion, the data should be deleted immediately unless there is a special agreement. Cloud platforms need the following means to ensure that data is not leaked during and after deletion.

Memory deletion:

Before the operating system reallocates the memory to the user, it needs to clean up the allocated memory, so as to ensure that malicious memory detection software can not detect useful information in the newly launched virtual machine and prevent data leakage caused by deleting data through physical memory recovery.

Data security (soft) deletion:

Cloud platform needs to provide logical deletion of discarded data. Tenants can achieve flexible one-click deletion of data in storage services such as RDS through the management console according to their needs.

Disk data deletion:

Cloud platform needs to adopt zero-clearing measures for virtual volumes of retailers to ensure that data can not be recovered, effectively prevent malicious tenants from using data recovery software to read disk data, and eliminate the risk of information leakage.

Physical disk scrap:

When the physical disk is scrapped, the data is removed by degaussing, bending or crushing the storage medium, and the data removal operation is kept a complete record to meet industry

standards to ensure user privacy and data are not unauthorized access.

2.5 Platform Management

Platform management mainly refers to the overall management of large data platform, including the unified scheduling of user rights, management processes, service deployment, platform security, data provision and other services. Data storage computing layer is the core of large data processing environment. It stores all kinds of data collected by data acquisition layer, and provides data computing and analysis capabilities for upper application. In the construction of large data in this scheme, massive structured data and unstructured data need to be processed, among which structured data is the main part. In order to meet the needs of current and future data storage and processing capabilities, we use advanced "Hadoop + MPP" hybrid architecture to build the entire storage computing layer. MPP platform is mainly used to process massive structured data with high value density. Hadoop platform is mainly used to process other unstructured data and structured data with low value density.

3. Core Technology and Difficulties of Big Data Scheme

After the big data hardware platform is built, a lot of manpower, energy and time are needed to create and calculate the data model, constantly modify the relevant characteristic parameters, and strive to make each model adapt to the current situation of the equipment as far as possible. It is recommended that the industry modeling engine developed by the major cloud platforms be used to improve work efficiency. Rate, general modeling engine is a data analysis application system with data warehouse construction, data retrieval, data analysis, data governance and data display as the core. Operators use this system, from data access, data warehouse establishment, data association, analysis, mining, to visual presentation. Now complete data analysis process.

Because there are many types of equipment in hydropower station, the data types and quantities are all in mass. It will be a long-term and arduous task to find the association rules of different data in so many data and apply them to production, operation and daily management. The law between data is endless, and new types of data can always be found. It can be seen that only by constantly discovering laws, summing up experience and transforming them into applicable results can the intellectualization of hydropower stations become more and more advanced and mature.

Core technologies: data management, data processing, data services, data security

4. Conclusion

Hydropower station is a kind of comprehensive engineering facility, which mainly converts water energy into electric energy for use in many fields such as agriculture, industry and so on. China has abundant water resources, and the demand for electricity for social development is gradually increasing, the number of hydropower stations is increasing. Small hydropower stations have significant advantages over large hydropower stations in terms of construction cycle, investment and functional development. Therefore, in the process of operation and management of mechanical and electrical equipment in small hydropower stations, we must do a good job in safety management. Only in this way can we continuously promote the long-term sustainable development of hydropower stations. At present, the major domestic hydropower companies are promoting the work of intelligent power stations, but they are basically in the stage of scheme design or hardware platform construction. There are not many better data models developed and used to guide production and operation. Therefore, we should vigorously promote the data modeling of hydropower stations and the association rules between different data. Summarizing and refining is a breakthrough point in the construction of intelligent hydropower stations, an important part of the application of information technology in the power industry, and also an inevitable trend. It can effectively improve the level of system intelligence and self-diagnosis and self-analysis function, enhance the efficiency and reliability of equipment operation, and build a safe, stable, intelligent and efficient intelligent system of hydropower plants. Unity lays the foundation.

References

[1] Liu Juan. Discussion on Equipment Management of Small Hydropower Stations [J]. Henan Water Conservancy and South-to-North Water Transfer, 2009, (9): 57.

[2] Sun Hui. Problems and Countermeasures in equipment management of small and medium-sized hydropower stations [J]. New technologies and products in China, 2010, (14): 70.

[3] Cui Shuzhi. Exploration of Mechanical and Electrical Material Equipment Management in Water Conservancy and Hydropower Construction Project [J]. Industry and Science Forum, 2011, 10 (20): 242.