Research and Design of Smart Grid Monitoring System Based on Cloud Computing

Yuan Yan

College of Electrical and Electronic Engineering, North China Electric Power University (Baoding), Baoding, Hebei, 071000, China

Abstract. Real-time data of power system is an important data for power system equipment measurement and acquisition. These data are important basis for analyzing power system stability, predicting grid load and power equipment failure and aging, and are the data that must be monitored for grid operation. Aiming at the problem of massive monitoring data access and processing in smart grid monitoring system, this paper proposes a smart grid monitoring system based on cloud computing framework by comprehensively utilizing geographic information technology, network communication technology and distributed database technology. The system's workflow is more efficient and reliable for information processing than traditional database models and existing methods.

1. Introduction

Numerous power equipment and monitoring instruments in the power grid form real-time status data that requires the power information system to continuously receive and process. The amount of data in these status data is huge, and high reliability and real-time requirements are imposed on the monitoring system. However, in the face of massive state data, the traditional system shows that the storage capacity and processing capacity are seriously insufficient, which restricts the working efficiency of the power grid monitoring system to a large extent. Therefore, storing information data and processing these data has become the key to restricting smart grid monitoring. Hadoop is a decomposition/aggregation cloud computing framework for large dataset object analysis processing. Through the coordinated management of the distributed file system HDFS and the parallel programming model Map/Reduce, the system can effectively segment and reasonably schedule massive data, thus achieving efficient parallel processing for big data.

At present, China is accelerating the construction of smart grids, and adopting cloud computing technology for reliable storage and parallel processing of state monitoring data will inevitably play a huge role in promoting the establishment of smart grids in the future. Based on this, based on the research of the combination of geographic information system and smart grid, this paper develops intelligent grid monitoring system based on cloud computing to realize efficient parallel storage and processing of large dataset grid monitoring information, and it is difficult to solve real-time monitoring of grid operation. The problem of slowness and improvement of the monitoring of the operation of the power grid to improve the overall emergency support capability of the power grid.
2. Grid monitoring technology

Grid monitoring refers to the process of providing the basis for equipment fault diagnosis by detecting, transforming, analyzing, processing and displaying the characteristic signals under a certain operating state, and outputting the information collected by the diagnosis. Through the combination of advanced automatic control, communication technology, computer technology, information technology, etc., telemetry, remote signaling, remote control, remote adjustment operation, and secondary equipment and auxiliary equipment for power operation equipment facilities monitored in a physical area of the system Realize remote control, realize monitoring, measurement, control, recording and alarm of all primary equipment, realize optimal operation, optimization control and optimization management of power grid, thereby improving grid operation status, safety level and forecast level of accident disaster prediction.

The data to be faced and processed by the smart grid monitoring system mainly includes three types: basic geographic information data, grid equipment facility spatial data and other data with spatial attribute characteristics. The basic geographic information data includes national grid map data, provincial network map data, and key city grid map data of various scales. Grid equipment facility spatial data is grid data with spatial location information, including power plants, substations, overhead lines, towers, cable lines, and so on. Other data with spatial attribute characteristics are grid environment information, including meteorological information, disaster information, geological monitoring (earthquake), natural disasters (typhoons, ice disasters, etc.), and sources of danger. These will have an impact on the grid, which is the focus of the monitoring system. The smart grid monitoring system belongs to a complex adaptive system and requires comprehensive support of geographic information technology, network communication technology and database technology.

The technical advantages of GIS are its hybrid data structure and effective data integration, unique geospatial analysis capabilities, rapid spatial location search, and evolutionary simulation of spatial processes and spatial decision support functions, through the grid equipment and its geography. The combination of location attributes allows for faster identification of line problems or accident locations and grid equipment facilities, improving the efficiency of emergency response.

The smart grid puts higher requirements on the automation and operation level of the grid. To realize the smart grid, the information technology neural network must run through the whole network. Through the sensor network, intelligent terminals, intelligent control centers and information networks, the whole network can be controlled and controlled. Grid control and operational models to be compatible with new energy sources and flexible to respond to diverse user needs.

While the monitoring system displays and monitors the power parameters, in order to perform statistics, analysis and processing on the data in the future, and realize the dynamic display of the data curve, it is also required to provide a database storage function for the system. Since the power information has many kinds of data, large amount of data, inconsistent format, one-time writing, multiple readings, etc., the power monitoring equipment needs to continuously write real-time monitoring data into the database. In order to meet the requirements of reliability and real-time, the traditional relational database is not used, and the distributed data management mode based on column storage is adopted to support the efficient management of large data sets. Although the distributed database spreads the data records on each physical node, it still logically belongs to the same system. This data is shared by the method of data distribution. A global database is responsible for the management of the entire system, and some databases are undertaken by the local database of each node. In this way, the data is dispersed to maximize the local application, and the mutual interference between the computing nodes is also minimized. Tasks are shared among nodes, thus avoiding load bottlenecks and improving work efficiency.

3. System design

The smart grid monitoring geographic information system designed in this paper is divided into three levels, namely the on-site monitoring layer, the network communication layer and the management application layer.
The on-site monitoring layer is also called the sensing layer. It consists of various sensors and sensor network tubes, including temperature sensors, humidity sensors, two-dimensional code labels, RFID tags, readers, cameras and other sensing terminals. It mainly monitors the operating parameters of all links in the power grid. Real-time monitoring including transmission line monitoring, substation station monitoring, substation monitoring, etc., including geographic attribute information with geographic coordinates, network communication network operation status and grid environment information.

The network communication layer mainly realizes the transmission and control of information. The various operating parameters of the power grid obtained from the on-site monitoring layer and the grid environment information are transmitted to the management application layer through the network communication layer. On the other hand, instructions issued from the management application layer need to be forwarded to the monitor of the power running device on the field monitoring layer through the network communication layer. The network communication layer undertakes a two-way, massive grid data transmission task. The system uses a variety of communication technologies such as power private network and public wireless network to complete the data transmission of the network communication layer.

The management application layer collects, summarizes and manages the geographic information and attribute information of the grid equipment facilities in the on-site monitoring layer, and comprehensively monitors the overall operational status of the power grid to analyse, process, and analyze the acquired information to provide emergency command and decision-making. Key data sources.

According to the characteristics of smart grid equipment monitoring, this paper uses distributed redundant storage system and column storage-based data management mode to store and manage data to ensure the reliability and efficient management of smart grid massive state data. The operating condition data of the power equipment facility acquired by the sensor network, the power private network, and the like are generated and stored in the database through the Map Reduce-based grid monitoring data parallel processing platform. Other information databases with spatial attribute features include remote sensing information, road information, and weather information. For meteorological information, relevant meteorological geographic information data such as grid dispatching meteorological warning and forecasting service data are obtained from the meteorological department through a dedicated cable or the Internet. This article uses Hive and HBase database to store data. As a distributed database, HBase has high query efficiency. Install HBase on the already configured Hadoop smart grid monitoring platform, edit and modify the hive-site.xml in the conf directory. The modification is shown in Figure 1.

![Figure 1. Edit and modify the file in the conf directory](image)

Through the system read and write test, no matter which end of the data is written, the data can be read at the other end, indicating that the Hive database can be used for large data storage and processing, and HBase can quickly display data, indicating that the two technologies are obtained. Effective integration.

4. Parallel processing based on Map Reduce

When processing the monitoring data set, Map Reduce first divides it into hundreds or thousands of small data sets, and then each node in the cluster processes one or several divided small data sets and
produces intermediate results, and finally passes merging a large number of nodes translates these intermediate results into final results. The whole work process is divided into two stages: Map and Reduce.

![Parallel processing operation](image)

**Figure 2.** Parallel processing operation

Usually, there are many duplicate keys in the intermediate results processed by the Map operation. In order to alleviate the burden of Reduce operation and network transmission, the system is optimized, that is, a custom Combiner method is used to locally integrate and stipulate the intermediate results. The Combiner operation runs on each node that performs the Map operation, usually using the same process as the Reduce operation. The only difference is that the result of the Reduce operation is written to the final output file, and the result of the Combiner operation is sent to the Reduce operation as an intermediate file. The specific steps of the improved Map Reduce execution process are as follows:

1. First, the monitoring data set of the cloud computing platform that is imported into the running user program is preprocessed, and data noise reduction and the like are processed.

2. The Map Reduce function library in the user program divides the imported monitoring data file into 16 to 64 megabytes of M blocks (which can be adjusted by parameters), and then executes a copy of the program on different machines in the cluster. It should be pointed out that the split does not need to understand the internal logical structure of the file. The specific split mode can be specified by itself or by several simple partitions that Hadoop has defined.

3. In all running processes, the master program master is responsible for the allocation of the remaining execution tasks. The master program in the execution program assigns Map and Reduce tasks according to the idle status of the worker.

4. The working node assigned the Map task reads and processes the input data block, and the Map function finally outputs the intermediate result pair. In addition, in order to further shorten the processing time, when performing the Combine operation, the merged key pairs are first indexed, then merged into several large new data pairs, and after being transferred to the Reduce process, the decomposition is performed, and the index is utilized.

5. The system packs the key/value pair data and sends the index information to the Master, and then transmits it to Reduce through the Master.

6. The Reducer Worker calls the user-defined Reduce function to analyze and sort the set of intermediate results.

7. After executing all the Map Reduce tasks, the Master is responsible for controlling the corresponding monitoring data to be stored in the library and generating a corresponding copy file for transmission.
5. Conclusion
Based on the research of the combination of geographic information system and smart grid, this paper designs a smart grid monitoring system based on cloud computing for the problem of large data set storage in the traditional mode of power grid monitoring system. The method stores and processes massive amounts of monitoring data. Through the development of cloud computing-based smart grid monitoring system, the efficient parallel storage and processing of large dataset grid monitoring information is realized, the problem of difficult and slow real-time monitoring of power grid operation is solved, and the monitoring of grid operation is improved to the overall emergency of the grid.

References
[1] REN Jianfeng, DING Yawei, FU Lei, et al. An improved strategy for out-of-step separation based on phase angle principle for 1 000 kV ultra-high voltage AC power grids [J]. Automation of Electric Power Systems, 2011, 35 (10): 104 - 107.
[2] Li Chenghua, Zhang Xinyi, Jin Hai, Xiang Wen. MapReduce: A New Distributed Parallel Computing Programming Model [J]. Computer Engineering and Science, 2011, 33 (3): 129 - 135.
[3] Luo Junzhou, Jin Jiahui, Song Aibo, et al. Cloud computing: architecture and key technologies [J]. Journal of Communications, 2011, 32 (7): 3 - 21.
[4] Lu Wei, Zhang Tianbing, Li Guozhi. Design and key technology of power grid emergency command information management system [C]. Proceedings of China Communications Society 2011 Optical Cable and Cable Academic Annual Conference, 2011: 45 - 50.
[5] Shu Qiang. Information Network Support Platform for Smart Grid [J]. Equipment Manufacturing Technology, 2011 (3): 118 - 121.
[6] Li Xiangzhen, Liu Jianming. IoT technology for smart grid and its application [J]. Telecom Network Technology, 2010, 8: 41 - 45.
[7] Wang Dewen, Song Yaqi, Zhu Yongli. Smart grid information platform based on cloud computing [J]. Power System Automation, 2010, 34 (22): 7 - 12.
[8] Meteorological industry standards of the People's Republic of China. Power grid dispatching meteorological warning and forecasting service products (draft for comments) [S]. Beijing: China Meteorological Administration, 2011.