Research on Hadoop Massive Power Quality Data Based on Computer Cloud Platform

Manyu Liu¹*, Huan Liu², Lin Li³

¹State Grid Beijing Electric Power Research Institute
²Information and Communication Company, State Grid Baoding Electric Power Supply Company
³Department of Mechanical Engineering, North China Electric Power University, Baoding

*Corresponding author e-mail: liumanyu@bj.sgcc.com.cn

Abstract. With the continuous expansion of the demand for power in various industries, the scope of power data collection is also gradually expanding, which makes the explosive growth of power quality data, which brings new pressure and challenges to the analysis and processing means and tools of power quality data. Based on this, this paper first analyses the characteristics and architecture of Hadoop big data tech, then studies the power quality data analysis of Hadoop based on computer cloud platform, and finally gives the power quality data analysis of Hadoop based on computer cloud platform.

Keywords: Power Quality, Cloud Platform, Hadoop

1. Introduction
With the iterative progress and growth of social economy, the demand for electric energy of various industries in the society is constantly expanding, and the demands for the quality of electric energy are also gradually improving. The scale of the current power system continues to expand, making the analysis of power quality more complex [1]. The traditional analysis and calculation of power quality mainly rely on the centralized computing platform, which can better match the small-scale power grid. However, for large-scale power grid system, this calculation method has some practical problems, such as poor scalability, and the computing capacity is difficult to meet the actual needs. With the development of power quality system, the ability of power quality data storage and analysis also needs to be adjusted. And the expansion of power data collection will also bring the explosive growth of power quality data, which brings new pressure and challenges to the analysis and processing means and tools of power quality data.

With the iterative maturity of computer info tech, big data analysis means and platforms are constantly emerging. For example, Hadoop distributed tech, as a platform for distributed processing of massive data, has several typical advantages as shown in Figure 1, so it has gradually become a hot spot and focus of current application and research. Based on Hadoop platform, the design and establishment of power quality monitoring and data computing platform can effectively integrate the
power data computing platform and power quality monitoring system, so as to ameliorate the processing ability of the platform for power quality data [2].

Figure 1. Typical characteristics of Hadoop distributed tech

In addition, as an important support and guarantee that directly affects the smooth operation of social economy, electric energy is an important prerequisite for the stable and healthy development of various social industries. Monitoring and analyzing power quality data is helpful to provide support and auxiliary data for power quality optimization [3]. The impact of power quality on the operation of electrical equipment is mainly reflected in the cable capacity, safe operation of electrical equipment, power failure and other aspects. Too low power quality will lead to the failure of electrical equipment and load, seriously affecting the normal operation of related equipment. The big data processing based on Hadoop platform mainly realizes the storage and analysis of a large number of complex data through distributed storage database, which not only reduces the cost of hardware, but also has high scalability.

In a word, using Hadoop cloud platform to carry out the monitoring, storage and analysis of massive power quality data can not only meet the timeliness demands of power grid for massive data processing, but also have high cost performance [4]. In particular, the application of computer cloud platform significantly reduces the investment of power grid enterprises in related analysis equipment, and can achieve higher output at a lower cost. In addition, with the scalability advantages of cloud platform in storage and computing, it can effectively guarantee the growing processing requests of power quality data. Therefore, the research of Hadoop massive power quality data processing based on computer cloud platform has important practical value.

2. Characteristics and architecture of Hadoop big data tech

2.1. Characteristics of Hadoop big data tech
Hadoop big data tech based on computer cloud platform is a distributed computing that runs massive power quality data in a cluster composed of a large number of computers. It allows applications to support massive nodes and data. Hadoop is mainly composed of distributed storage and distributed computing. Hadoop big data tech has high scalability, economy, reliability and efficiency. Among them, Hadoop big data tech is the basic of its design, whether it is the scalability of storage or computing. Secondly, Hadoop framework can run on any ordinary computer device [5]. In addition, the backup and recovery mechanism of distributed file system and distributed computing task monitoring ensure the reliability of distributed processing. The efficient data interaction implementation of distributed file system and the local data processing mode of distributed computing geometry establish the technical premise for efficient processing of massive power info.

2.2. The architecture of Hadoop tech
Hadoop big data tech system based on computer cloud platform is composed of several aspects as shown in Figure 2 below. Pigs is a large-scale data analysis platform based on Hadoop, and provides...
recommended operation and programming interface for complex massive data parallel computing [6]. Chuakwa is a tool based on Hadoop, which can provide complete SQL query function and convert SQL statements into MapReduce tasks to run. In addition, Zookeeper is an efficient and scalable coordination system that stores and coordinates key shared states. HBase is an open source distributed database based on storage model [7]. HDFS is a distributed file system with high fault tolerance. It can be deployed on common hardware and is suitable for applications with large data sets.

![Diagram of Hadoop architecture]

**Figure 2.** The architecture of Hadoop tech

### 2.3. Hadoop core design based on computer cloud

The core design of Hadoop based on computer cloud mainly includes MapReduce and HDFS. Among them, MapReduce mainly includes the decomposition of characters and the summary of results. HDFS includes file management, file storage and file acquisition [8]. As a highly fault-tolerant distributed file system, HDFS can provide high-throughput data access, which is more suitable for large-scale data sets. The NameNode in HDFS can be regarded as the manager of the distributed file system. DataNode is the basic unit of file storage. It stores the file blocks. In the local file system, it stores the Meta data of the file blocks, and periodically sends all the existing file blocks to the NameNode.

### 3. Power quality data analysis of Hadoop based on computer cloud platform

#### 3.1. The necessity of power quality data analysis

With the increasing complexity of the power grid system, the number of power grid terminal equipment is also increasing. In this context, the degree of diversification and complexity of the power grid equipment industry is also growing; making the system connected to the power grid more complex [9]. With the wide application of electric energy, the info created and shared, especially the non-institutional data generated by the business of power grid industry is growing rapidly. In addition, with the continuous integration of data in the power industry, the island info originally hidden offline is continuously input online.

#### 3.2. Characteristics of electric energy data in electric power industry

The data characteristics of the power industry mainly include large amount of data, scattered data, complex storage path format, various types and so on.
Secondly, the power industry also faces several typical problems, such as the lack of effective analysis methods, numerous data formats, low efficiency of data transmission, data duplication and so on, as shown in Figure 3 above. In addition, the info contained in power quality data has its own characteristics. Only by synthesizing the info contained in various data can the real power quality be reflected. The demand of power enterprises for big data products and solutions is mainly reflected in scalable storage, high bandwidth, analysis solutions that can handle different formats of data, etc.

4. Design of Hadoop power quality data cloud platform

4.1. Functions of Hadoop power quality data cloud platform
Hadoop power quality data cloud platform first has the function of online design collection and storage. Through the online data acquisition subsystem, the data is collected regularly and stored in the data platform. Power quality data acquisition module is composed of data acquisition hardware and monitoring data management software. Secondly, the platform should also have power quality analysis function [10]. With the help of the analysis module, the collected power quality monitoring data and indicators are analyzed and calculated. In addition, the platform can also generate abnormal info reports of power quality data monitoring, and realize cluster environment based on existing equipment to make rational use of resources. Finally, the platform can be transplanted to the existing system, and can be expanded according to the needs.

4.2. Design of power quality monitoring system
Firstly, the collected power quality data are stored, statistically analyzed, calculated and displayed according to the monitoring point number. Secondly, the process of power quality data processing is composed of monitoring data, pre analysis, storage data, statistical analysis, power quality indicators and so on. Through the statistical analysis of the monitoring data of each monitoring point, each power quality index is calculated. In addition, the big data processing flow of power quality mainly includes data acquisition, data integration, data analysis and processing, and data delivery. According to the distribution of data tables in HBase, Hadoop, a MapReduce cloud platform, allocates multiple computing tasks to process the corresponding data of multiple monitoring points in a distributed parallel manner.

4.3. Hadoop architecture of power quality data cloud
Hadoop is used to develop and run applications dealing with massive data. With the advantages of Hadoop, such as high reliability, high scalability, high efficiency and high fault tolerance, the cloud architecture of power quality cloud model is built. Hadoop platform is built on the cluster, distributed files and databases are deployed, and the parallel power quality analysis algorithm is deployed to the distributed platform. Firstly, Hadoop platform is built on the cluster, and Hadoop is configured as fully distributed mode. Secondly, the distributed database is deployed, and the operation of the database is
tested by starting and connecting the hive and metadata managed Derby database. In addition, parallel power quality analysis algorithm is carried out, and MapReduce programming model is adopted to reduce the workload of parallel programming. Finally, the application system is released, and different database operation modules are called by jsp foreground or Xfire encapsulated web service.

5. Conclusion

In summary, using Hadoop cloud platform to carry out the monitoring, storage and analysis of massive power quality data can meet the timeliness demands of power grid for massive data processing, and has high cost performance. The application of computer cloud platform reduces the investment of power grid enterprises in related analysis equipment, achieves higher output at lower cost, and effectively guarantees the growing processing demands of power quality data. Based on the analysis of the characteristics and architecture of Hadoop big data tech, this paper studies the core design of Hadoop based on computer cloud. Through the research on power quality data of Hadoop based on computer cloud platform, the necessity of power quality data analysis is analyzed. Through the design and analysis of Hadoop power quality data cloud platform, the architecture and design process of power quality monitoring system are given.

References

[1] Cui Yan. Research on smart grid info platform based on cloud computing [J]. Digital tech and application, 2015 (5): 103-103.
[2] Ghazi M R, Gangodkar D. Hadoop, MapReduce and HDFS: A Developers Perspective [J]. Procedia Computer Science, 2015, 48: 45-50.
[3] Huang Bin, Xu Shuren, Pu Wei. Design and implementation of data mining platform based on MapReduce [J]. Computer engineering and design, 2013, 34 (2): 495-501.
[4] Maitrey S, Jha C K. MapReduce: Simplified Data Analysis of Big Data [J]. Procedia Computer Science, 2015, 57(3): 563-571.
[5] Nachiappan R, Javadi B, Calherios R, et al. Cloud Storage Reliability for Big Data Applications: A State of the Art Survey[J]. Journal of Network &Computer Applications, 2017, 97.
[6] Qu Guanglong, Yang Honggeng, Li Lanfang. Design and implementation of real time power quality monitoring system for active distribution network [J]. Power system automation, 2015 (10): 117-123.
[7] Song Jie, Sun Zongzhe, Mao Keming, et al. Research progress of MapReduce big data processing platform and algorithm [J]. Acta Sinica Sinica, 2017, 28 (3): 514-543.
[8] Tan Jiamao, Huang Shaoxian. Study on comprehensive evaluation method of power quality based on fuzzy theory [J]. Relay, 2016, 34 (3): 55-59.
[9] Wang Jianxiong, Wang Yulan, sun Haoxue, et al. Design of power quality computer monitoring system based on Ethernet [J]. Power tech, 2013,37 (7): 1237-1238.
[10] Zhang Yi. Research on online monitoring and evaluation system of coal mine power quality [J]. Zhongzhou coal, 2013 (12): 17-18.