Research on Big Data Analysis Platform of Power Grid Enterprise Accounting Based on Cloud Computing

Jia Tian¹, *, Dan Xu², * and Meng Cui², b

¹School of business administration, Jilin Engineering Normal University, Changchun 130052, China
²School of Economics and Trade, Jilin Engineering Normal University, Changchun 130052, China

*Corresponding author email: 2364146255@qq.com, 48485658@qq.com, 1554314667@qq.com

Abstract. With the development of smart grids, big data is generated in all aspects of the power system, and the traditional financial system is no longer able to accurately analyze the large amount of financial data. In this paper, based on the problems arising from the traditional financial system decision-making process, the decision-making of grid financial system based on big data and cloud computing is proposed. Through the analysis of the big data characteristics of the power system, the application of big data and cloud computing in the smart grid is clarified, and the external application oriented to the customer group and the internal application oriented to the system flow are explored. Taking the operation and maintenance cost of an actual substation in the power grid as an example, the paper uses the idea of big data to conduct scientific research on its operation and maintenance cost, which intuitively reflects the application of big data, and makes the readers clearer and vividly see big data. How to apply it to accelerate the application development of big data in China.

1. Introduction

Since entering the 21st century, human society has officially entered the era of network informationization. Under the influence of the background of informationization, enterprise informationization and management systematization has become the mainstream trend. In power companies, due to their own data. The complexity and hugeness of processing not only realizes information management, but also comprehensively applies cloud computing technology to realize big data processing. In recent years, the operating environment of power grid companies has undergone tremendous changes, and the current financial management of power grid enterprises has shown great limitations. However, with the development of information technology and the advent of the era of big data, it also provides grid companies with more space for technical solutions [1-2]. This paper explores the establishment of a cloud-based grid company financial shared service center, effectively separates the transactional accounting business from the value creative management business, and builds a new grid enterprise financial management model with information support, resource use efficiency and efficiency improvement. Achieve resource integration, deepen data analysis and use, and further
improve the decision support ability and risk management and control capabilities of financial enterprise financial management.

2. Analysis of Big Data Characteristics of Power Enterprises Accounting in Cloud Computing Environment

With the development of information technology, especially Internet of Things technology and mobile computing technology, power enterprise accounting must collect and process large amounts of big data with different structures. In addition to the four "V" features of general big data, these power enterprise accounting big data are also characterized by intangibility and stickiness:

2.1. The number of data is large
The amount of data generated under the Internet of Things is large. It is not described by large-scale data, huge data, and massive data in the past, but should be summarized by big data. The data size is not measured in GB, TB but in PB.

2.2. Data heterogeneous data
The data generated under the Internet of Things includes not only structured data such as numbers, but also unstructured data such as sounds and images. Because of the relevance of these data to business events, structured data and unstructured data are more complex and difficult to handle.

2.3. Data generation and processing real-time
Traditional data does not require high time processing. However, the behavior of the Internet of Things and the behavior of people generally require completion in the present, so the generation and processing of data is real-time.

2.4. Low value density
Power enterprise accounting big data will be generated continuously, but valuable data is only a piece or a part of the continuously generated data. Take video as an example. In the continuous monitoring process, the data that may be useful is only one or two seconds. Therefore, the value density of big data is low.

2.5. Intangibility and Stickiness of Accounting Data in Power Enterprises
The digital information collected, transmitted and processed by enterprises on the Internet of Things is mainly non-valued quantity information. These data can be directly perceived by the sensor and thus easily propagated; while the power enterprise accounting data is intangible data, which cannot be perceived by the sensor. At the same time, the accounting data of power companies is directly bonded to the business data and cannot exist away from the business data. Therefore, power enterprise accounting data is intangible and sticky.
3. Traditional grid financial decision-making process and existing problems

Decision-making is an important function of enterprise financial management. Throughout all aspects and functional systems of corporate financial management, scientific decision-making is the core of financial management, and the key to decision-making is the decision-making process and procedures [3]. The traditional decision-making process is divided into four steps: first, discover the problems that arise in the daily financial process; secondly, perform a detailed logical analysis of the problems found; again, find out the causal relationship of the problems, and identify the causes and key factors Finally, develop a solution.

With the rapid development of the domestic economy and the strategy of electric power first, the development of the power grid across regions, industries and diversification has brought tremendous vitality to the development of the power grid. At the same time, however, the complex financial environment has also increased the difficulty of financial management of power grid companies. For example, the power grid management level is continuously extended, the business diversified and expanded, and the quality of personnel is uneven. As a result, financial management faces business data dispersion, slow information flow, multiple levels of accounting reports, difficult data quality assurance, and financial information sharing. Difficulties [4].

4. Establishment of grid accounting big data platform based on cloud computing

4.1. Demand Analysis

Cloud computing-based big data analytics technology has a complete reference architecture and software implementation, and has been applied in some industries. However, most of these systems are from the Internet industry, and their design, implementation, and functional features fully reflect the needs and characteristics of Internet big data and its services. When applied to industries other than the Internet, it is common practice to develop and optimize the limitations of existing systems based on the analysis of the big data characteristics and business needs of the industry. There are three main differences between power accounting big data and Internet big data.

(1) In the Internet scenario, a typical big data application needs to scan the entire data set sequentially. Therefore, the distributed parallel big data analysis system Hive or Impala does not provide good support for the index; in the power accounting big data analysis, Multi-dimensional area queries are extremely
common, and because there is no index, it will result in access to a large amount of unneeded data and significantly reduce the performance of the query. It is necessary to design a suitable index structure and corresponding data retrieval mechanism for the characteristics of multi-dimensional area query.

(2) The typical feature of Internet big data is "write multiple times at a time." For this data feature, the Distributed File System (HDFS) and Hive do not provide a data rewriting (update or delete) mechanism, and can only indirectly achieve the purpose of rewriting data by covering all existing data. In the power accounting big data business scenario, there are a large number of data rewriting statements, and executing these queries in a manner of covering existing data will lead to inefficient execution. Therefore, it is urgent to provide a data rewriting mechanism in the existing system.

(3) Big data query language designed by Internet companies according to their own business needs, such as HQL is only a subset of SQL, and power data analysis systems are mostly written in standard SQL language, which requires a lot of manpower and time to bring existing The tens of thousands of SQL statements are translated into equivalent HQL statements. Therefore, it is necessary to design a tool to automatically translate the SQL language into the HQL language, thereby improving the migration speed of the legacy application and realizing the seamless and smooth migration of the power data analysis service.

Based on the above analysis, in order to meet the needs of smart grid for in-depth analysis and mining of power accounting big data, based on the typical characteristics of power accounting big data and its business logic, combined with the latest developments of cloud computing technology and actual industry deployment experience, based on the cloud Calculated power accounting big data analysis system.

4.2. System Structure Design

The system is based on distributed parallel computing framework (Hadoop), using Hive as data analysis software, and developed multi-dimensional index based on grid file, SQL-to-HQL automatic translation tool and support based on query rewriting for power accounting big data features. Key technologies such as data-updated hybrid storage models have enhanced Hive's performance and ease of use.

From the collection to the final analysis and calculation of power accounting big data, it takes several steps, and its flow is shown in Figure 1. A large number of sensors and smart meters collect data periodically at a fixed frequency and send it to the data center via a communication network. For errors or omissions in the data collection process, supplementary acquisition of power consumption information (hereinafter referred to as supplementary mining) is performed irregularly. To alleviate the access pressure of the cloud storage system, the collected data first enters the front-end buffer pool for decoding and pre-processing. In addition to the data collected by the data source on a regular basis, relatively static information such as equipment and personnel constitutes an archive database, which is stored in a relational database and directly copied to the cloud storage system. When the archive database is updated, the update should be synchronized to the cloud storage system to ensure the accuracy of the calculation results. The parallel computing environment accesses the cloud storage system, performs complex analysis and calculation on the data according to the business logic, and writes the calculation result to the cloud storage system. The online query system obtains data that satisfies the user request from the cloud storage system and returns the user. Some queries for archive data need to calculate result data. Therefore, the cloud storage system also synchronizes the calculation results into a relational database so that users can query data through a relational database.
The power big data analysis system architecture is shown in Figure 3. The main modules include:

1. Distributed File System (HDFS) [5] module is one of the core modules of the entire big data analysis system, used for persistent storage of various types of power big data. HDFS consists of a metadata server and multiple data servers. The file is divided into 64MB data blocks and distributed to different data servers. Each data block has 3 copies distributed on different nodes. When a copy is inaccessible, the system automatically creates a new copy to maintain load balancing.

2. Hadoop [6] is an open source implementation of the Map-Reduce parallel programming framework proposed by Google. The MapReduce program consists of a Map function and a Reduce function. Each time the Map function converts an input (key, value) pair into a set of intermediate (key, value) pairs; the Reduce function processes the same set of values of the key to produce the final result is written to the distributed file system HDFS.

3. Hive [7] is a data warehouse system based on Hadoop platform for analysing and calculating big data. Hive provides the SQL-like query language HiveQL (HQL), which provides a familiar interface for data analysts. The HQL query is compiled by the Hive parser into a set of MapReduce programs. Similar to relational databases, Hive organizes data into tables and stores them in HDFS.

4. Monitoring tools and running scheduling tools. The monitoring tool is used to monitor the running status of the system and the execution status of the data analysis job. The running scheduling tool schedules the data analysis job according to the scheduling policy and resource status specified by the administrator, and analyses the association or dependency between the operations.

5. The development toolset includes SQL translation, parallel ETL tools, index management, task management and other tools, providing system administrators with a Web-based graphical interface to simplify system configuration management.
4.3. Case Analysis

In the grid financial system, equipment operation and maintenance, maintenance and retiring are often encountered. The cost is always changing throughout the process. For example, the transformer that has just been put into operation has a low cost of maintenance and operation. With the continuous use of the transformer, the number of years of service will increase, and the cost of maintenance and operation will increase. Therefore, the analysis of the costs incurred in different stages after the substation is put into operation is conducive to timely accounting of the required costs and improving the management and control capabilities of financial operations in terms of substation operation and maintenance. This paper takes a transformer as an example, based on the accumulated operation and maintenance cost data, analyzes it based on big data and cloud computing, and fits the cost of the substation, as shown in Table 1.

| Year | Original sequence | Fitting sequence |
|------|-------------------|------------------|
| 2009 | 161553.4          | 163552.4         |
| 2010 | 177038.8          | 178049.8         |
| 2011 | 193482.3          | 191165.3         |
| 2012 | 204815.8          | 202271.8         |
| 2013 | 221810.3          | 217623.3         |
| 2014 | 232476.8          | 231338.8         |
| 2015 | 248965.3          | 248544.3         |
| 2016 | 259556.8          | 258890.8         |
| 2017 | 273258.2          | 274881.2         |

According to the above data, it can be judged. In the initial stage, the equipment has just been put into operation, and the operation and maintenance cost are small. After a period of time, some of the equipment slowly began to appear some faults, which greatly increased the investment in repair or technical improvement, so the operation and maintenance cost increased. After the technical renovation, the equipment status was improved, and the operation and maintenance cost were restored to a relatively normal state. This is a basic rule. In the process after the equipment is put into operation, there may be a year that may increase every time. Through the analysis of big data and cloud computing, it can be
predicted that these years of cost spurt are in the first few years after commissioning, thereby pre-
arranged to prevent sudden occurrence of the situation, which can significantly improve financial
planning and enhance financial Management capabilities provide a solid analytical foundation.

5. Conclusion
This paper summarizes and analyzes the characteristics of smart grid data, combines the big data concept
with financial data for these characteristics, and implants the concept of cloud computing into the
analysis process, and finally combines the specific substation operation and maintenance cost examples.
It provides a role for the combination of big data and power grid, and lays a foundation for the
subsequent application of big data cloud computing in the grid financial system.

References
[1] Feng Guorui, Wang Liang. Research on Decision Making of Grid Financial System Based on
Big Data and Cloud Computing. Telecommunication Power Technology, Vol. 1 (2017) No.28,
p. 103 - 105.
[2] Huang Lei, Tao Rui. Power Big Data Analysis Technology and Application Based on Cloud
Computing. Digital Technology & Applications, Vol. 2 (2017) No.15, p. 175 - 183.
[3] Dong Huabiao. Power Big Data Analysis Technology and Application Based on Cloud
Computing. Nanfang Agricultural Machinery, Vol. 5 (2017) No.23, p. 105+107.
[4] Xie Xinwen. On the Application of Big Data and Cloud Computing Technology in Enterprise
Financial Management. Financial Economy, Vol. 3 (2015) No.14, p. 204 - 205.
[5] Yang Xiaoyan, Liao Qingyuan. Design and Implementation of Intelligent Financial
Information Platform Based on Cloud Computing in Big Data Era. Information and Computer
(Theoretical Edition), Vol. 7 (2016) No.18, p. 26 - 28.
[6] Wang Xianghong, Chen Yuyi. Research on the Status Quo and Development Trend of Cloud
Computing and Power Big Data Technology. Shaanxi Electric Power, Vol. 6 (2017) No.13, p.
52 - 55.
[7] Lou Fengdan, Yan Xubin, Wang Zhiqiang, et al. Research on Power Search Engine
Technology Based on Cloud Computing and Big Data Technology. Power System and Clean
Energy, Vol. 12 (2016) No.32, p. 86 - 92.