Research on failure rate of self-service payment terminal of power supply company based on cloud computing

Ming Feng Xu¹, Fei Song¹, Yan Jun Sun¹, Shan Lang¹, Wen Long Ma²
¹. State Grid Dalian Electric Power Supply Company, Dalian 116001
². Liaoning Hanhua Information Engineering Co. Ltd, Shenyang 110004
xmf_dl@ln.sgcc.com.cn

Abstract. With the continuous development of power companies, the load on the user side continues to diversify, the traditional payment methods in the power system are constantly updated, and the scale of self-service payment terminals has been fully expanded. However, as the scale of self-service payment terminals becomes larger and terminal equipment functions become more abundant, the failure rate of terminal equipment is also on the rise. Cloud computing can provide a large number of low-cost computing and storage resources, and has the characteristics of easy deployment and strong scalability, which can meet the software and hardware needs of big data analysis. Therefore, the effective integration of cloud computing and big data technology and the use of their respective advantages have important practical significance in the big data analysis of equipment failure. Completing the prediction of the failure rate of the terminal equipment according to the designed big data analysis framework will help to arrange the maintenance plan of the terminal equipment reasonably and improve the reliability of the power self-service payment terminal.

1. Introduction
With the deepening of the electricity market reform, power supply services have become one of the important tasks of power companies, especially the implementation of the "one household, one meter" project and the advancement of urbanization, and the number of electricity users will continue to increase. With the continuous development of economy, science and technology, the pace of people's life has accelerated, and the standard of living has entered the information age. The traditional electricity bill payment method of electric power has not kept up with the development of the times and the actual needs of marketing work. In order to further expand the channels of electricity payment, reduce the payment pressure of banks, business halls and other methods, and meet the actual needs of the vast number of electricity customers, through the research on the technology of electricity mobile tolling terminals, the customer service capabilities are further improved and marketing methods are enriched.

The power mobile charging terminal performs data interaction with the terminal platform system through the GPRS/CDMA communication network, and the terminal platform system and the power supply marketing system perform data interaction through the data interface, so as to realize the data interaction between the terminal and the power supply marketing system, and realize the remote electricity bill payment and card meter Electricity business such as electricity purchase, electricity bill inquiry, payment correction, and credit card payment.

The networking mode adopts APN dedicated line, and all electric mobile charging terminals adopt fixed IP in the internal network. This specification requires access to the mobile company's GPRS
network through a 2M/10M APN dedicated line, a private fixed IP address is used between the interconnection routers of the two parties for wide area connection, and a GRE tunnel is used between the GGSN and the mobile company's interconnection router. The SIM card used for the GPRS private network only activates the dedicated APN and restricts the use of other APNs. After obtaining the APN, allocate mobile internal fixed IP to all power mobile charging terminals. End-to-end 3DES encryption is used between the terminal and the terminal management platform to avoid possible information leakage during the entire transmission process. Both parties use firewalls for isolation, and perform IP address and port filtering on the firewall. This kind of specification has greatly improved safety and stability regardless of timeliness, and is suitable for application environments with high security requirements, more data points, and high timeliness requirements.

2. Design principles
Security and high reliability: The power mobile charging terminal involves a large amount of customer information, electricity data information, etc., so the security and high reliability of the system must be ensured in the selection of terminal equipment and system design.

Compatibility: Although the power mobile toll terminal has a certain degree of independence, it is inseparable from the support of the basic business system. Therefore, the compatibility with other basic systems must be considered when the system is designed.

Manageability and efficiency: In order to ensure the manageability of the data of the power mobile charging terminal, the standardized design should reduce the complexity of management and ensure the high reliability of the system. The terminal has the ability of high-speed and sustainable access to data.

Scalability: As an innovative power payment method, in the system design and implementation, the scalability of the system should be fully considered, and advanced technology and craftsmanship should be adopted to facilitate the smooth upgrade of the entire system.

3. Design projects
The power supply company controls the actual operation of the self-service payment terminal equipment and monitors it, and can provide real data generated by the self-service payment terminal faulty equipment. Based on real data, it can be further divided into offline data and real-time streaming data. Offline data represents historically generated data. We need to save it for subsequent data analysis. Real-time streaming data refers to the data generated in real time during operation of the equipment. Through real-time analysis of streaming data, it can provide early warning of failures in time and reduce losses.

Self-service payment terminal equipment status monitoring and fault diagnosis The "cloud computing" platform architecture design follows the methodology of system science, essentially abstracting and encapsulating the functions of the entire system. There are clear levels, functions and structures inside the system for external display It is a whole and provides specific services. In the design of the architecture, a top-down approach is generally used to analysis the problems to be solved, and a bottom-up approach is adopted for specific requirements. In computer science, regarding system design, it is common to simplify, encapsulate and abstract the overall requirements in a layered manner, and implement them as module services, that is, the upper layer services call the interfaces of the lower layer services to obtain the required data, and the lower layer Services provide data support for upper-level services. This article studies the bottom-up model to complete the overall architecture design. The entire computing platform is divided into five parts, including data source, data access layer, data storage layer, resource layer, and data analysis layer. The specific structure is shown in Figure 1.
3.1. Data source

Data structure is mainly divided into structured, semi-structured and unstructured data. Different data structures have different preprocessing and storage methods for one. For structured data, relational databases can be used for storage, such as Mysql, Oracle and other databases. For unstructured data, non-relational databases such as MongoDB and HBase can be used.

If the scale of data to be analyzed is not large, then the conventional technical route can meet the requirements. However, for self-service payment terminal equipment, with a variety of data collection methods, its data structure is relatively complicated, which in turn generates large quantities of data. These data can be divided into two types according to actual analysis scenarios, namely offline processing and real-time stream processing. For the analysis of non-real-time data, we need to manually write related task programs, and use batch processing to analyze the data in batches. During the analysis, we also need to monitor the execution tasks, and schedule and allocate different resources according to computing requirements.
3.2. Data access layer
Flume is a distributed system for log management provided by Apache, with high reliability and high availability. In terms of log management, it can efficiently complete log collection, aggregation, and transmission from different data sources. In the actual operation process, multiple self-service payment terminal devices constitute a variety of data sources. Through the Flume log collection module, a variety of data can be transmitted on the basis of ensuring load balance, timeliness, fault tolerance and timeliness. Analyze on Hadoop. The principle of Flume is simple and consists of four major components, namely data source, data channel, data outlet and interceptor.

Specify the data source when using it, including TCP, Avro, Kafka, etc. At the same time, declare the data storage location through Channel, and finally specify the final transmission path through Sink, including HDFS, Hbase, File, etc. In the demand analysis, it is mentioned that in the application scenario, there will be multiple self-service payment terminal devices to generate log data, and the corresponding Flume can be configured to use multiple Flume components. Each Flume declares a different Source to represent different data from Sources obtain data to meet actual needs. After satisfying the data collection of multiple data sources, you will face a problem, that is, when the amount of data generated per unit time is too large, if only one Flume component is used for the final data transmission, disk IO will be frequently performed, which will be serious Affects the efficiency of data transmission, so it is necessary to perform load balancing operations on the data transmission of Flume. For data obtained from multiple data sources, add multiple Flume components as the intermediate transmission process of transmission, and set relevant algorithms internally to support round robin, random Algorithm to select different data output, store the data in different places, to achieve load balancing.

3.3. Data storage layer
The data storage layer mainly provides a place to receive data from the data access layer, and then provide the data to the big data analysis module for use. The analysis methods are divided into real-time analysis and offline analysis. The problem to be considered is that the raw data obtained by Flume should not be consumed in real time, and it is not necessary to obtain all the data. Instead, some key data can be selectively obtained and then analyzed.

For the storage of faulty data, HDFS is used to complete the persistence of data, write it to the hard disk, and read it when needed. It can be used for a long time and is suitable for offline data storage. For the data generated in real time, due to the relationship of multiple data sources, a large number of logs will be generated. Real-time analysis is required to further judge the current working status of the equipment. According to the analysis results, it can be certain that the running equipment may malfunction It is an early warning function, and if these data are written into HDFS and then read from the hard disk, there will be a certain delay, and there will be a relatively large lag when the analysis results are obtained. Therefore, a reasonable channel is required for storage of the real-time data to be analyzed., Can read and analyze in real time.

Based on the data requirement analysis of the above real-time streaming data analysis, the use of Kafka can well meet the above requirements. First of all, it is a distributed system, based on the Hadoop-Spark cluster we designed, which can be compatible and extended well. For multiple data in the data access layer, Kafka also supports multiple subscription points, multiple consumers and multiple subscribers. It can distribute data well for different data sources, and can provide high throughput during publishing and subscription consumption.

3.4. Source layer
In a big data cluster, a large number of tasks will be submitted in actual operation, and each task has different requirements for computing resources. How to allocate the existing computing resources in the cluster to improve resource utilization can be effective The efficiency of submitting tasks. In the research of this article, the Yarn resource manager is used for resource management, which can
effectively manage the resources on the Hadoop-Spark cluster. In the industry, the Yarn resource management mode is widely used.

Yarn is a new feature of Hadoop 2.0. In the previous Hadoop 1.x version, although a distributed computing framework was implemented, there were still many problems in resource management. For example, the data sharing mechanism of the cluster was not perfect, and the computing nodes in the cluster were not perfectly utilized, and more memory resources were required during operation. Compared with the centralized resource management mode in Hadoop 1.x, Yarn resource management is obviously different in architecture design. The more prominent one is that the JobTracker is divided into two parts, namely the resource manager RM and the application master node AM. The resource manager RM is mainly used for resource management, and the application master node is mainly used for managing and scheduling applications submitted by users. In general, the Yarn source management system is mainly composed of resource managers, application master nodes and containers.

Different from offline analysis of data, real-time streaming of data puts forward higher requirements for the timeliness and stability of the system. For clusters where real-time data needs to be transmitted synchronously, data analysis needs to be completed quickly to obtain feedback results. This method is particularly important for equipment failure analysis. For equipment that is in operation, through real-time analysis of various equipment operating parameters, it can alert in time when the data is abnormal, which is beneficial to equipment maintenance and failure detection.

3.5. Data analysis layer
The Spark computing framework provides the Spark Streaming module to complete the analysis of real-time data streams. The data is decomposed in chronological order to form RDDs. The subsequent processing flow is the same as offline analysis. The key to the design of this module is real-time access to data. The framework studied in this paper uses Flume+Kafka+SparkStreaming to complete the real-time stream computing module design.

Because the Hadoop distributed computing framework is implemented based on the Java language, and Spark needs to use Scala as a common language, it is necessary to configure the Java and Scala languages on the server to provide coding support. For the choice of Java version, the Java version 1.8 was chosen because it has been stable, is widely used in engineering practice, and is widely used in the development of projects by Internet companies in the industry, while the Scala version selected 2.12.7, version selection The reason is also based on stability considerations.

Figure 2 shows the relationship between the usage time of the payment terminal and the failure rate. From the figure, it is easy to see that the use of the big data analysis method designed in this paper to reasonably arrange the maintenance plan can greatly reduce the equipment failure rate.

![Figure 2 Time-varying characteristic fitting diagram of equipment failure rate](image)

4. Conclusion
Electricity fee recovery is the top priority of the power marketing work of power supply companies. The system has realized the network, community, and ubiquity of power mobile tolling outlets, which
facilitates the payment of electricity fees by the majority of power customers, improves the recovery rate of electricity fees for power supply enterprises, and makes effective use of social rationality.

Resources. The multi-dimensional assessment method of self-service payment terminal equipment health status proposed in this paper and the use of "cloud platform" information system for big data analysis and reasonable arrangement of terminal equipment maintenance plans can effectively reduce the failure rate of self-service payment terminal equipment and greatly increase Operational reliability of terminal equipment.

Acknowledgements
This paper is supported by “State grid science and technology project (2020YF-49).”

References
[1] Ghemawat S, Gobioff H, Leung S T. The Google file system[J]. Acm Sigops Operating Systems Review, 2003, 37(5):29-43.
[2] Wu M Y, Shu W, Zhang H. Segmented min-min: a static mapping algorithm for meta-tasks on heterogeneous computing systems[C]// Heterogeneous Computing Workshop, 2000. (HCW 2000) Proceedings. 9th. DBLP, 2000.
[3] Lei Y, Jia F, Lin J, et al. An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data[J]. IEEE Transactions on Industrial Electronics, 2016, 63(5):3137-3147.
[4] Lv S, Zhang M, Li Y, et al. Design on the Fault Diagnostic System Based on Virtual Instrument Technique[C]// International Workshop on Knowledge Discovery & Data Mining. IEEE, 2009.
[5] Prieto M D, Cirrincione G, Espinosa A G, et al. Bearing Fault Detection by a Novel Condition-Monitoring Scheme Based on Statistical-Time Features and Neural Networks[J]. IEEE Transactions on Industrial Electronics, 2013, 60(8):3398-3407.
[6] Daengduang S, Vatekul P. Applying One-Versus-One SVMs to classify multi-label data with large labels using spark[J]. 2017:72-77.
[7] Azab A, Domanska D. Software Provisioning Inside a Secure Environment as Docker Containers Using Stroll File-System[C]// 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing. ACM, 2016.