



# Application of Big Data Processing Technology in Power Consumption Information Acquisition

Jin Wang<sup>1</sup>(✉), Yukun Xu<sup>1</sup>, Chao Jiang<sup>1</sup>, Jingrui Yan<sup>2</sup>, Bo Ding<sup>2</sup>, and Qiusheng Lin<sup>3</sup>

<sup>1</sup> State Grid Shanghai Municipal Electric Power Company, Shanghai 200120, China  
j f 1 2 5 1 2 5 @ 1 6 3 . c o m

<sup>2</sup> LongShine Technology Group Co., Ltd., Shanghai 201600, China

<sup>3</sup> Guangzhou Huali College, Guangzhou 511325, China

**Abstract.** In the big data environment of power enterprise operation, power consumption information plays an important role in user behavior monitoring and power generation task planning. In order to provide effective information reference for power work, big data processing technology is applied to optimize the design of power information acquisition method. According to the composition structure and working principle of the power network, the power network model is constructed. The power information collector and processor are installed, and the A/D conversion circuit is used to complete the A/D conversion of the power information acquisition signal. Synchronously control the power consumption information acquisition program, and use big data processing technology to count the power consumption information parameters, such as electricity consumption and electricity charges. Through noise filtering, missing value compensation and other steps, the pretreatment of the information collected at sea is completed, so as to realize the collection of electricity information. Compared with the traditional acquisition method, it is found that the acquisition error of the optimal design method in the two aspects of electricity consumption and electricity cost is reduced by 0.345 kWh, 0.095 yuan, reducing the redundancy of the acquisition information, and improving the integrity of the acquisition information.

**Keywords:** Big Data Processing Technology · Power Consumption Information · Information Collection

## 1 Introduction

User power consumption information acquisition is to realize power consumption monitoring [1], implement step pricing, load management, and line loss analysis through the acquisition and analysis of power consumption data of distribution transformers and end users, and finally achieve the purposes of automatic meter reading, off peak power consumption, power consumption inspection, load forecasting, and power consumption cost saving. To establish a comprehensive user electricity information acquisition system, it is necessary to build the system master station, transmission channel, acquisition equipment and electronic watt hour meter. The coverage rate of power information acquisition

of State Grid Corporation of China is 100%. With the application of smart meters in China becoming more and more popular, to a certain extent, it has increased the daily collection of electricity data in China, and also increased the difficulty of data acquisition and data analysis of the electricity consumption information acquisition system. Fortunately, the development and application of big data technology have effectively solved the problem of large real-time data volume of the electricity consumption information acquisition system. The practice shows that the big data power consumption information acquisition system that conforms to the IEC international communication standard can support tens of millions of users at the same time, and can complete real-time acquisition of all kinds of power data, remote parameter configuration, two-way time division measurement, power quality detection, power failure detection, on-off control, break-point detection, etc. Under this technical support, the power consumption information acquisition system can complete various services for power customers, such as power price and rate adjustment Power demand, analysis and statistics of power consumption data [2]. By the end of 2021, China's power supply companies are carrying the power supply demand of 130 million residents, covering more than 90% of the land area, forming a 220 kV power supply scale with three layers and two networks. Data sources are extensive, processing is difficult, and users are distributed dispersedly, increasing the difficulty of traditional manual meter reading, causing errors and losses, affecting work efficiency and decision-making speed, and data privacy is easy to leak. As a result, the security of users' basic information cannot be guaranteed, so it is necessary to design power consumption information acquisition.

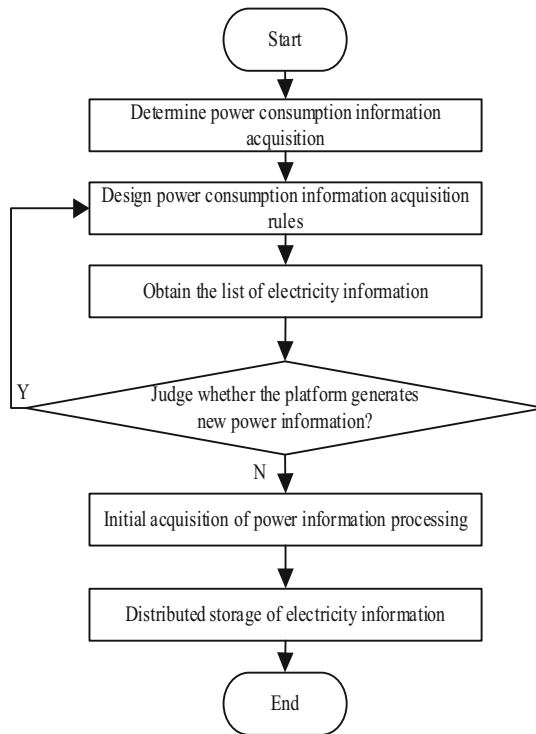
At this stage, the more mature power information acquisition methods include: cloud storage based power information acquisition method, DLPTP clock synchronization based power information acquisition method and redundancy analysis based power information acquisition method. The cloud storage based power acquisition method can be used to record the voltage, current, effective power, reactive power, frequency Weighted harmonic distortion, unweighted harmonic distortion, harmonic and other parameters. The principle of information acquisition and measurement recording is: measure the effective value of voltage in one cycle or half a cycle. When the effective value of voltage exceeds the set upper or lower threshold, record the parameters of each channel at that time; When the effective value of voltage is within the set upper limit and lower limit threshold, it will not be recorded. High quality flash memory is used internally to compress and store the measured data. The amount of data compressed for a single channel can be reduced to 5% of the original amount, and the compressed data can be transmitted through LAN/WAN. In addition, the power consumption information acquisition method based on DLPTP clock synchronization mechanism pays more attention to the synchronization of information acquisition, and avoids repeated acquisition and acquisition deviation of power consumption signal as far as possible. The power consumption information acquisition method based on redundancy analysis analyzes the redundant part of the acquisition results to minimize the redundancy of the power consumption information acquisition results. However, the above power consumption information acquisition methods still have the problem of poor acquisition quality in the actual use process, which is mainly reflected in the low acquisition accuracy and low integrity of

the acquisition information. Therefore, big data processing technology is applied to the acquisition of power consumption information.

Big data analysis technology specifically refers to big data application technology, covering various big data platforms [3], big data index systems and other big data application technologies. The big data analysis technology can be divided into distributed computing, data storage, data parallel computing and other parts. The big data analysis technology is applied to the power consumption information acquisition work, and the acquisition method is optimized to improve the power consumption information acquisition performance.

## 2 Design of Power Consumption Information Acquisition Method

With the support of big data technology, the power consumption information acquisition process is shown in Fig. 1.



**Fig. 1.** Power consumption information acquisition flow chart

Figure 1 Power consumption information collection process can be summarized as the following steps:

- Step 1: Determine the power consumption information collection;
- Step 2: Design power consumption information acquisition rules;

Step 3: Obtain the power information list;

Step 4: Judge whether the platform generates new power information? If yes, retrieve the power information list according to step 2; If not, go to step 5;

Step 5: Initial acquisition of power information processing;

Step 6: Distributed storage of power information.

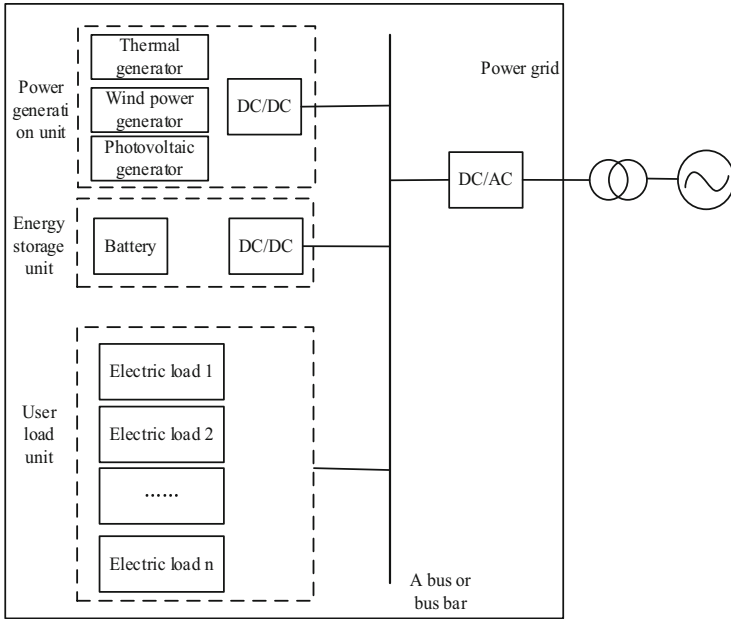
The power consumption information acquisition includes three aspects, namely, real-time acquisition, timing task acquisition and terminal active reporting. Real time acquisition refers to the data required for real-time measurement, including terminal parameters, real-time data, historical data, terminal events, etc. The timed task collection is the most important aspect of data collection. The system terminal data is collected by timed automatic tasks and then stored in the database. The main content of the timing task includes the collection of historical data terminals. After the first data collection, the data shall be analyzed and measured immediately to calculate the success rate, and then the invalid data obtained from the collection and analysis shall be supplemented. Scheduled automatic task collection needs to configure the task type, task start time, number of tasks to be executed each day, task priority, task collection date, task collection data set and other related information. The timing task can complete the collection of more than 98% of the data in four hours. In this way, timed task collection has two advantages. First, it can take the data of the day in time and make statistics and analysis on it. Second, it does not affect the operation of other businesses and does not occupy working time. By configuring tasks and saving them, you can obtain relevant task information, including the priority, start time, cycle, interval between task execution, and running status of all tasks. Configure the scheme to realize the active reporting of terminal data, and transmit the reported task scheme to the designated terminal [4]. The terminal will confirm the task, and then actively report the task data regularly according to the content of the task plan. After getting the actively reported data, the collection server and front-end server will process the data and store it in the database after processing. After actively reporting the task scheme, you can view the terminal and system's active reporting forecast, and use this as a reference to adjust the time and data volume of the active reporting task, so as to ensure the balance rationality of the system's all day reporting data. The task scheduling module is formed by integrating real-time data acquisition, timed task acquisition and terminal active reporting data processing.

## 2.1 Construction of Electricity Network Model

The power grid is generally connected with the power generation network, distribution network and transmission network. In consideration of power generation, distribution, transmission, storage and other factors, the power network model is built as shown in Fig. 2.

The real-time output characteristics of the power generation terminal in the network model environment can be expressed as Formula 1:

$$I = I_g - I_b \left\{ \exp \left[ \frac{q(U_e + I_e R_{\text{series}})}{\kappa_{\text{diode}} \gamma} \right] - 1 \right\} - \frac{U_e + I_e R_{\text{series}}}{R_{\text{parallel}}} \quad (1)$$



**Fig. 2.** Structure Diagram of Power Network Model

In formula (1),  $I_g$  and  $I_b$  respectively represent the real-time generated current and reverse saturation current,  $q$  represents the amount of charge generated by power generation,  $I_e$  and  $U_e$  represent the rated current and rated voltage,  $R_{series}$  and  $R_{parallel}$  represent the series resistance and parallel resistance in the grid environment,  $\kappa_{diode}$  represents the characteristic parameters of the diode element, and  $\gamma$  represents Boltzmann constant [5]. The available power generated by the grid in real time can be expressed as Formula 2:

$$Q = I^2 R_{all} \tag{2}$$

Substitute the calculation result of Formula 1 into Formula 2 to get the calculation result of real-time power consumption in the power network model. Use Formula 3 to calculate the maximum power consumption provided by the power network.

$$Q_{all} = Q + Q_{storage} \tag{3}$$

In Formula (2),  $Q_{storage}$  represents the total electricity storage in the grid model. The real-time work process of the power network and the initial network power data are substituted into the power network structure to complete the construction of the power network model.

### 2.2 Installation of Power Information Collector and Processor

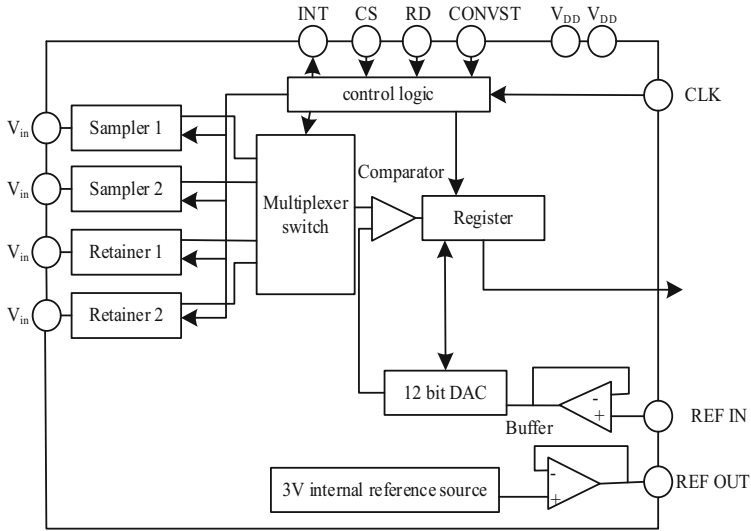
In the constructed power network model, install the power consumption information collector as the hardware support for the operation of the acquisition program, and install a

processor device to connect with the information collector to process the real-time collected power consumption information and provide hardware support for the operation of big data processing technology. In view of the particularity of power consumption information acquisition, the working principle of power consumption information acquisition is to collect power grid signals in real time, including analog signals such as grid voltage signals and grid current signals, with a total of 18 differential signals [6, 7]. After the EMI filter effectively suppresses the influence of external environmental noise on the system, the analog signal passes through the differential operational amplifier to obtain the signal to be converted by A/D sampling. The normal operation of each component in the power consumption information collector is guaranteed by the acquisition board FPGA. The function of the acquisition board FPGA is to receive the synchronous sampling signal sent by the core board FPGA through the I/O port, and control the synchronous sampling and holding of three A/D chips; Then, the grid data after A/D conversion is sequentially stored in the RAM of the acquisition board FPGA, and is sequentially stored from 1 to 18 channels; Then, under the control of the core board FPGA, these digital signals are transmitted to the core board FPGA through the I/O port. The function of the core board FPGA is to receive the GPS signal, double the frequency of the GPS pulse signal to 1 MHz through a phase-locked loop, and send it to any frequency division module. The frequency division module divides the 1 MHz signal as required to generate a synchronous sampling clock that controls A/D synchronous sampling conversion, and records the GPS time scale during synchronous sampling. The grid data after A/D sampling conversion is obtained from the acquisition board FPGA and stored in the dual port RAM in the form of ping-pong buffer. When either of the two storage spaces of the dual port RAM ping-pong cache is full, it sends an identification signal to the DSP. After that, the DSP can process the digital signals stored in this part of RAM, mainly including lossless compression and lossy compression of data, and transmit and store the processed data [8]. The core board FPGA generates control signals to control the dual port RAM, including enable signals and read/write signals. At the same time, the core board FPGA also generates a write clock to control the dual port RAM writing process.

The processor used in the user information collection method takes the dual core processor chip as the core component, and the dual core processor chip selects ARM-Cortex - A9 as the processor core. A9 adopts the ARM-v7 architecture, which has rich functions and runs fast. At the same time, A9 has its own NEON coprocessor. The media signal processing structure in this coprocessor enriches audio, video and image processing instructions. These advanced multi data stream instructions can be used for ARM and Thumb status at the same time. Cortex.A9 microarchitecture can be used for both traditional single core processors and scalable multi-core processors. In this design, the speed index of dual core processor chip is 1.3 GHz.

### 2.3 Analog Digital Conversion of Electric Power Information Acquisition Signal

The A/D circuit is used to realize the A/D conversion processing of the real-time electricity information acquisition signal. The designed A/D conversion circuit is shown in Fig. 3.



**Fig. 3.** A/D conversion circuit diagram

The four channel 12 bit high-speed sampling/holding device AD7874 is used in the design of analog digital conversion circuit to sample the instantaneous value of voltage and current. The AD7874, a synchronous data collector with high accuracy and small relative delay, is made by the analog device company using the LC MOS process. The AD7874 can synchronously sample the input of four channels and save the size information and relative phase between the four channels, which makes the AD7874 particularly suitable for three-phase voltage and current detection, AC motor control and other occasions where the size information and relative phase of each channel are required [9]. The A/D conversion circuit shown in Fig. 3 can synchronously sample four channels, each channel has its own sample/hold amplifier, and the sample/hold time is 2 s; There is a high-speed 12 bit A/D converter inside, which converts the input sampling values of each channel in turn. The conversion time of each channel is 8 s, the conversion accuracy is  $\pm 1$  LSB, and the sampling rate of all four channels can reach 29 kHz. With fast data transmission interface, it can be directly connected with 6-bit microprocessor and other interfaces. Moreover, it is equipped with a 3 V precise reference source and a clock generator, with an input voltage range of  $\pm 10$  V, a typical input impedance of 30 k, and a data bus output form of binary complement.

#### 2.4 Synchronous Control Electricity Information Acquisition Procedure

The synchronization control of the power consumption information acquisition program is divided into two stages: clock alignment and correction. First, the slave clock is consistent with the master clock, and then the deviation between the slave clock and the master clock is reduced within the clock alignment interval. An appropriate error correction algorithm can prolong the clock interval. It is important for resource constrained clocks to reasonably determine the clock synchronization interval and reduce unnecessary clock

synchronization operations. Clock synchronization is the process of adjusting its local time from the clock to the standard time of the master clock. The specific operation process can be quantified as Formula 4:

$$\begin{cases} t_{\text{Slave},1} = t_{\text{Master},1} + \varepsilon + T_{\text{MS}} \\ t_{\text{Master},1} = t_{\text{Slave},2} - \varepsilon + T_{\text{SM}} \\ \delta(T_{\text{MS}}, T_{\text{SM}}) = 0 \end{cases} \quad (4)$$

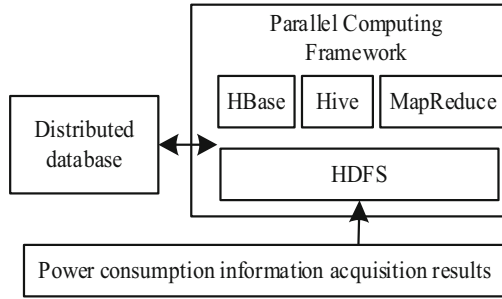
In Formula (4),  $t_{\text{Slave}}$  and  $t_{\text{Master}}$  respectively represent the slave clock and master clock of the power consumption information acquisition program,  $\varepsilon$  represents the time offset of the slave clock relative to the master clock,  $T_{\text{MS}}$  and  $T_{\text{SM}}$  represent the propagation time of the message between the master clock and the slave clock, and  $\delta(T_{\text{MS}}, T_{\text{SM}})$  represents the round-trip time difference. The local time of slave clock is consistent with the standard time of master clock after clock alignment [10]. However, due to the drift of the clock, the local time will deviate from the standard time until the next clock calibration. The maximum allowable deviation is assumed to be  $\varepsilon_{\text{max}}$ . The slave clock can run a time error correction algorithm to reduce this deviation. In the same way, we can get the correction algorithm of the non-linear error of the electric power information acquisition program, so as to greatly increase the clock time interval and realize the synchronous control of the operation process of the acquisition program.

### 2.5 Applying Big Data Processing Technology to Statistics of Electricity Information Parameters

The distributed computing technology in big data processing technology is applied to statistics of electricity consumption parameters, electricity charge parameters and line loss parameters in electricity consumption information. There are two kinds of distributed computing technologies: offline computing and real-time computing. Offline computing is dominated by Hadoop parallel computing framework, while real-time computing is dominated by Storm stream computing framework. The traditional way of computer analysis and data collation is to collect data first, then store it in the database program and search the data after receiving the request [11, 12]. This is an efficient processing method, but it is a tight structure, and usually causes a waste of time. In stream computing, stream data can be analyzed when it is received, and real-time computing can be achieved. The operation framework of distributed computing technology in big data processing technology is shown in Fig. 4.

Using the distributed computing technology shown in Fig. 4, the collection result of electricity consumption information is shown in Formula 5:

$$y_{\text{Power}}(t) = x_{\text{Power}}(t) + \sum_{t=1}^{n_t} \xi_{\text{influence}} \lambda_t^{(T)} \quad (5)$$



**Fig. 4.** Operation architecture of distributed computing technology

In Formula (4),  $x_{\text{Power}}(t)$  is the power consumption data acquisition sequence,  $t$  and  $n_t$  are the information acquisition time and total duration,  $\xi_{\text{influence}}$  is the impact factor of information acquisition, and  $\lambda_t^{(T)}$  is the pulse function. In addition, the collection result of electricity charge information is Formula 6:

$$y_{\text{fees}}(t) = y_{\text{Power}}(t) \times c_{\text{Unit}} \quad (6)$$

In Formula (6),  $c_{\text{Unit}}$  is the cost of unit electricity. Similarly, the statistical results of power consumption information parameters in any period can be obtained.

## 2.6 Pretreatment of Initial Acquisition Power Information

In order to ensure the acquisition quality of power consumption information, it is necessary to preprocess the noise information, repeated information and missing information in the initial acquisition information. The electric noise information is processed by median filtering, and the filtering window is constructed to move point by point in the initial acquisition of electric information. After the calculation and processing of Formula 7, the filtering results of information noise are obtained.

$$y'_{\text{filter}} = \begin{cases} y_{\frac{n}{2}}, n \in U_{\text{even}} \\ \frac{1}{2}(y_{\frac{n}{2}} + y_{\frac{n+1}{2}}), n \in U_{\text{odd}} \end{cases} \quad (7)$$

In Formula (7),  $U_{\text{even}}$  and  $U_{\text{odd}}$  respectively indicate that the initial collection set of electricity information belongs to an even number or an odd number [13]. The purpose of repeated information preprocessing is to reduce the redundancy of power consumption information acquisition results. First, use Formula 8 to judge the similarity of any two power consumption information in the initial acquisition results.

$$s(y_i, y_j) = \frac{y_i \cdot y_j}{\|y_i\|_2 \cdot \|y_j\|_2} \quad (8)$$

In Formula (8),  $y_i$  and  $y_j$  respectively represent any two of the initially acquired electricity information [14]. If the calculation result of formula 8 is higher than the

threshold value  $s_0$ , it means that  $y_i$  and  $y_j$  are highly similar and need to be filtered. Otherwise,  $y_i$  and  $y_j$  are considered as non repeated information and no filtering operation is required. After repeated data filtering, the statistical result of electricity consumption information is as shown in Formula 9:

$$N_y = N_0 - \frac{N_{\text{repeat}}}{2} \quad (9)$$

In Formula (9),  $N_0$  and  $N_{\text{repeat}}$  respectively represent the initial acquisition of electricity information and the detection of repeated electricity information. Finally, the missing values in the electricity information are processed, and the processing results are shown in Formula 10:

$$y'_{\text{defect}} = \frac{[\omega_1(y_{t-i} + y_{t+i}) + \omega_2(y_{t-1} + y_{t+1})]}{2} \quad (10)$$

In Formula (10),  $\omega_1$  and  $\omega_2$  represent the weighting coefficients of the power consumption information of the acquisition time before and after the missing value. After the above process, the preprocessing of the initial acquisition of electricity information is completed.

## 2.7 Realize Power Consumption Information Acquisition

### 2.7.1 Classification of Electricity Information

In order to improve the utilization rate of collected electricity information, with the support of big data processing technology, determine the type of current information, and provide reference for the classified storage of electricity information. According to the acquisition object of power consumption information, different power consumption information is selected as the standard information type. Formula 11 is used to calculate the matching degree between the real-time acquisition power consumption information and the standard information.

$$\psi = \sqrt{y_i^2 - y_{\text{set}}^2} \quad (11)$$

In Formula (11),  $y_{\text{set}}$  represents the set type standard feature. Compare the relationship between the matching degree and the threshold value. If the calculation result of the matching degree is higher than the threshold value, it is considered that the current power consumption information is consistent with the  $y_{\text{set}}$  type, otherwise it is considered that the two types are inconsistent, and the next standard type data needs to be matched. Finally, the electricity information divided into the same type will be fused, and the processing result is shown in Formula 12:

$$\psi = \sum_{i=1}^{n_{\text{type}}} \varpi_i \cdot y_i \quad (12)$$

In Formula (12),  $\varpi_i$  represents the fusion coefficient of electricity information, and  $n_{\text{type}}$  represents the amount of electricity information divided into the same type. This completes the classification of electricity information.

### 2.7.2 Distributed Storage of Collected Electricity Information

The big data technology is applied to the information collection system, and the distributed storage mode is used to solve the problem of data loss to a certain extent. At the same time, the distributed storage mode can also support cluster data for stream computing, effectively ensuring the real-time performance of computing data. Distributed data acquisition has the advantages of high availability and high reliability. It can effectively ensure the real-time performance of the collected data and the timeliness of HDFS data in the process of power information acquisition. Stom distribution and stream computing have the timeliness of data acquisition and data storage. The distributed data acquisition method can set multiple agents according to the number and location of data acquisition terminals to complete multi-threaded data acquisition, At the same time, Hadopyam is used to improve the appending function of data files [15]. Figure 5 is a schematic diagram of distributed data storage of power consumption information acquisition results.

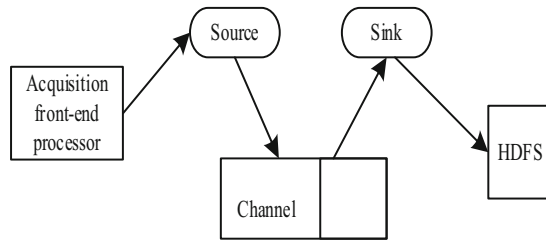


Fig. 5. Distributed storage of information collection results

Multi thread and storage buffer pool technology are adopted to dynamically apply for database connection according to the number and priority of stored data. Multi thread technology is adopted to realize parallel storage. Concurrent control access strategy is adopted to reduce resource competition for data storage space access and improve the storage efficiency of power consumption information collection results. The above process completes the real-time collection of user information.

## 3 Application Experiment Analysis

For the purpose of testing the application performance of big data processing technology in power consumption information acquisition, the application performance test experiment is designed. The power consumption information acquisition method before the application of big data processing technology and the power consumption information acquisition method using cloud storage technology are selected as the comparison method of the experiment. Through the test and analysis of power consumption information acquisition error, acquisition redundancy and acquisition integrity, It shows the advantages of the optimization design method in information collection performance.

### **3.1 Setting Up the Operation Environment of Power Consumption Information Acquisition Method**

Restricted by the hardware conditions, the optimized design of electricity information acquisition method deployed a big data processing technology platform on the laboratory server through virtual machines. The configured server type is P8000 workstation, the processor is IntelXeonCPUE5-2620v32.40 GHz \* 2, the memory capacity is set to 256 GB, the network card is installed with Gigabit Ethernet Card, and the operating system is WindowsServer2012R2DataCenter. Install VMware virtual software on the server, use VMware to virtualize the server, virtualize several virtual machines, install Ubuntu Server14.04.5LTS (64bit) operating system on all virtual machines, configure root user password, set static IP/DNS, and close the firewall. Since all node hosts trust each other through hosts, you need to configure the mapping between IP and hostname in the node host/etc./host file. Finally, you need to configure password free SSH login for all virtual machines, install NTP services, and unify the time source. Through the configuration of the operating environment of the acquisition method, it is ensured that the optimized acquisition method can operate normally in the experimental environment, and the acquisition results will not be affected by the external environment.

### **3.2 Selection of Power Consumption Information Acquisition Object**

In the experiment, the power supply area of a power supply company is selected as the collection object of power consumption information. The area of the power supply area is 502.6 km<sup>2</sup>, and the number of power users in the area is about 1980, including 1240 ordinary users, and the rest are commercial or industrial users.

### **3.3 Preparing Power Consumption Information Collection Samples**

By controlling the working parameters, working hours and working quantities of the electrical equipment within the scope of the power consumption area, the preparation of the power consumption information acquisition samples is realized. The preparation of some power consumption information samples is shown in Table 1.

The electricity charging standard within the selected electricity consumption area is 0.54 yuan/kWh, so we can get the setting result of electricity charge information. The data in Table 1 represents the power consumption data within one hour at the beginning of the experiment. According to the above method, the design results of power consumption information samples at any time during the experiment can be obtained, which is used as the comparison standard to judge the collection results of the design method.

**Table 1.** Sample Setting of Power Consumption Information Collection

User No	Power consumption/kWh	Electric charge/yuan	Power consumption information/MB
1	4.5	2.43	503
2	2.7	1.46	525
3	1.3	0.70	637
4	1.5	0.81	414
5	2.0	1.08	603
6	3.4	1.84	578
7	3.6	1.94	562
8	4.2	2.27	485
9	1.8	0.97	377

### 3.4 Describe the Application Experiment Process

The hardware equipment is installed in the test environment, and the communication network is used to connect the hardware equipment with the host computer. Under the condition that the user uses electricity normally, start the big data processing technology and electricity information acquisition program at the same time, and obtain the collection results of electricity information through the operation of the optimization design method. Figure 6 shows the power consumption information acquisition results of the optimized design method.

Under the same experimental environment, the comparison method set in the experiment is developed, and the corresponding output results are obtained through the operation of the comparison acquisition method, as shown in Fig. 7.

Similarly, the output results of the power consumption information acquisition method using DLPTP clock synchronization mechanism can be obtained.

### 3.5 Setting Quantitative Test Indicators for Application Experiments

The experiment is conducted from three aspects: the error, redundancy and integrity of electricity information collection. The error of electricity information collection refers to the error between the collected information and the prepared data samples. The test result of this indicator is shown in Formula 13:

$$\varepsilon_{\text{collection}} = |y_{\text{collection}} - y_{\text{standard}}| \quad (13)$$

In Formula (13),  $y_{\text{collection}}$  and  $y_{\text{standard}}$  represent the collected value and standard value of electricity information respectively. The lower the acquisition error of electrical information, the better the acquisition effect of the corresponding method. In the actual test process, the test is conducted from the two aspects of electricity consumption and electricity cost. The smaller the acquisition error is, the better the performance of the corresponding acquisition method is.

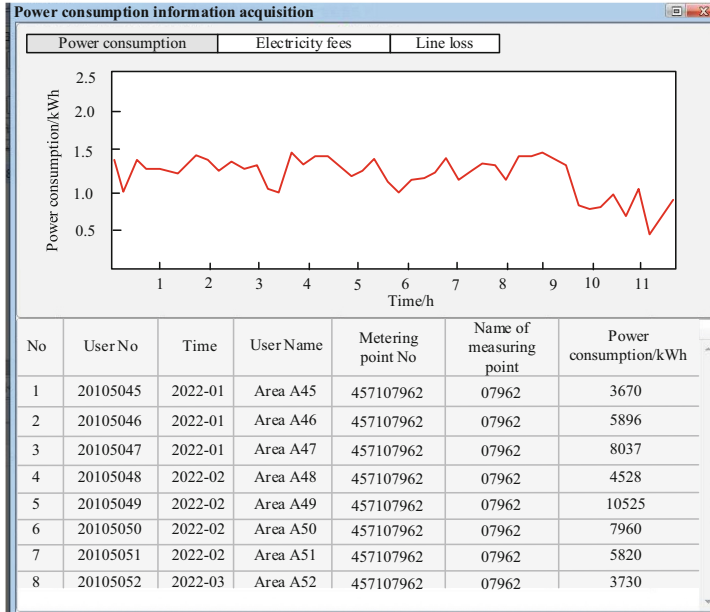


Fig. 6. Information Collection Results of Big Data Processing Technology

The numerical results of information redundancy are shown in Formula 14:

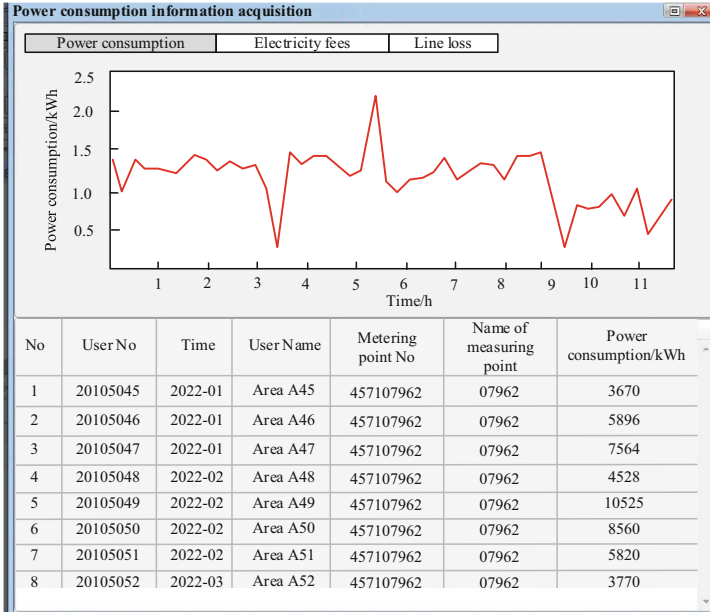
$$\eta_{\text{redun}} = \frac{\frac{N_{\text{repeat}}}{2}}{N_{\text{collection}}} \tag{14}$$

In Formula (14),  $N_{\text{repeat}}$  and  $N_{\text{collection}}$  are respectively the amount of duplicate data and the total amount of collected information in the collection results. The higher the redundancy of the collected information, the higher the rate of missed recovery in the process of power consumption information acquisition.

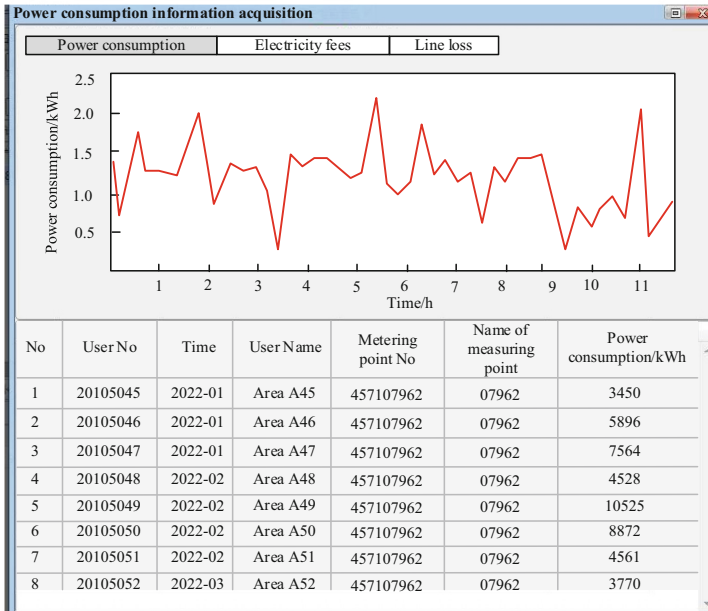
The test results of collected integrity indicators can be expressed as Formula 15:

$$\eta_{\text{complete}} = \frac{N_{\text{collection}}}{N_{\text{all}}} \times 100\% \tag{15}$$

In Formula (15),  $N_{\text{all}}$  represents the set number of power consumption information samples. The calculation shows that the lower the redundancy and the higher the integrity of the collected information, the better the collection performance of the corresponding method.



(a) Acquisition method without big data processing technology



(b) Collection method using cloud storage technology

Fig. 7. Power consumption information acquisition results of comparison method

### 3.6 Application Experiment Results and Analysis

Through the statistics of relevant data, the test results of power consumption information acquisition error are obtained, as shown in Table 2.

**Table 2.** Test Data of Electric Power Data Acquisition Error

User No	Power consumption information acquisition method without big data processing technology		Power consumption information acquisition method using cloud storage technology		Power consumption information acquisition method using big data processing technology	
	Power consumption/kWh	Electric charge/yuan	Power consumption/kWh	Electric charge/yuan	Power consumption/kWh	Electric charge/yuan
1	4.1	2.21	4.0	2.35	4.5	2.41
2	2.3	1.25	2.2	1.31	2.5	1.45
3	0.8	0.54	0.9	0.52	1.3	0.70
4	1.1	0.67	1.1	0.66	1.5	0.80
5	1.7	1.03	1.6	1.02	2.0	1.05
6	3.0	1.66	3.0	1.62	3.3	1.83
7	3.2	1.89	3.3	1.90	3.5	1.92
8	3.7	2.22	3.8	2.23	4.2	2.25
9	1.4	0.94	1.3	0.93	1.6	0.94

By substituting the data in Table 1 and Table 2 into Formula 13, it is calculated that the average power consumption acquisition error of the power consumption information acquisition method before and after the application of big data processing technology is 0.41 kWh and 0.07 kWh, and the average power charge acquisition error is 0.12 yuan and 0.02 yuan, while the average power consumption and power charge information acquisition error of the telecommunication information acquisition method using cloud storage technology are 0.42 kWh and 0.11 yuan, respectively. In addition, through the calculation of Formula 14 and Formula 15, the test comparison results of the redundancy and integrity indicators of power information acquisition are obtained, as shown in Fig. 8.

It can be seen from Fig. 8 that the redundancy of the proposed method is lower than 0.05 and the integrity is higher than 0.19. Compared with the two comparison methods, the results obtained by the power consumption information acquisition method using big data processing technology have lower redundancy and higher integrity. This proves that the big data processing technology has high application value in the power consumption information collection.

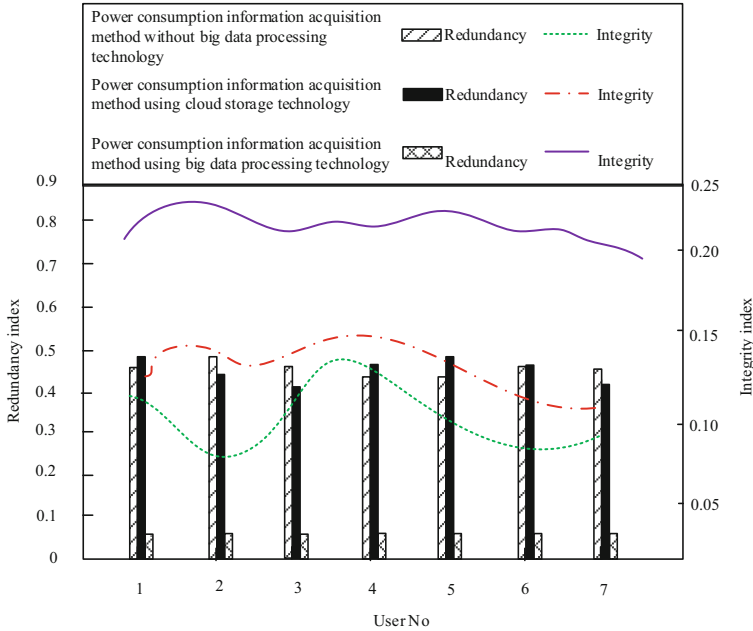


Fig. 8. Information Acquisition Redundancy and Integrity Test Results

## 4 Conclusion

With the development of society, power grid business has become more and more complex, which has caused difficulties in data collection, data calculation, data statistics and analysis, and its data storage has also increased. Therefore, this paper proposes the application research of big data processing technology in power consumption information acquisition. Through the integration and application of big data technology, the performance bottleneck of the power consumption information acquisition system has been effectively solved, so as to better assist the normal operation and maintenance of power consumption information acquisition and grid development.

## References

1. Yiying, Z., Haoyuan, P., Fei, L., Cong, W., Kun, L., et al.: Research on power consumption information acquisition system based on high speed carrier interconnection protocol. *Int. J. Syst. Assur. Eng. Manag.* **11**(4), 806–811 (2020)
2. Almshari, M., Tsaramirsis, G., Khadidos, A.O., et al.: Detection of potentially compromised computer nodes and clusters connected on a smart grid, using power consumption data. *Sensors* **20**(18), 5075 (2020)
3. Zhang, Q., Wang, P., Li, S., Jing, Y., et al.: A portable real-time EEG signal acquisition and tele-medicine system. *J. Comput. Meth. Sci. Eng.* **21**(4), 891–901 (2021)
4. Li, X., Hong, P., Zhu, R.: Simulation of vulnerability identification of mobile terminal software cache side channel. *Comput. Simul.* **39**(2), 496–500 (2022)

5. Wang, Z., Wei, Y., Qian, Y.: A simple direct heating thermal immersed boundary-lattice Boltzmann method for its application in incompressible flow. *Comput. Math. Appl.* **80**(6), 1633–1649 (2020)
6. Zhou, H., Guo, D., Honig, M.L.: Beam acquisition and training in millimeter wave networks with narrowband pilots. *IEEE J. Sel. Areas Commun.* **37**(12), 2759–2771 (2020)
7. Roy, S., Acharya, D.P., Sahoo, A.K.: Fast OMP algorithm and its FPGA implementation for compressed sensing-based sparse signal acquisition systems. *IET Circuits Devices Syst.* **15**(6), 511–521 (2021)
8. Ge, Y., Ye, H., Loparo, K.A.: Agent-based privacy preserving transactive control for managing peak power consumption. *IEEE Trans. Smart Grid* **11**(6), 4883–4890 (2020)
9. Norouzi, E., Hosseini, F.S., Vaezmosavi, M., et al. Effects of quiet mind training on alpha power suppression and fine motor skill acquisition. *J. Motor. Behav.* **53**(1/6), 399–408 (2021)
10. Zhang, N., Wang, M., Duan, Z., et al.: Verifying properties of mapreduce-based big data processing. *IEEE Trans. Reliab.* **71**(1), 321–338 (2020)
11. Yiying, Z., Haoyuan, P., Fei, L., et al.: Research on power consumption information acquisition system based on high speed carrier interconnection protocol. *Int. J. Syst. Assur. Eng. Manag.* **11**(4), 806–811 (2020)
12. Herodotou, H., Chen, Y., Lu, J.: A survey on automatic parameter tuning for big data processing systems. *ACM Comput. Surv.* **53**(2), 1–37 (2020)
13. Badr, B.M., Makosinski, A., Dechev, N., Delaney, K.R., et al.: Controlling wireless power transfer by tuning and detuning resonance of telemetric devices for rodents. *Wireless Power Transf.* **7**(1), 19–32 (2020)
14. Peng, W., Xiangning, X., Bingshan, X., et al.: Design and implementation of the data acquisition system of a large scale network based on four-in-one meters. *Indian J. Power River Valley Dev.* **70**(9/10), 162–172 (2020)
15. Othman, A., Mesbah, W., Iqbal, N., et al.: Sum-rate maximization and data delivery for wireless seismic acquisition. *Wireless Netw.* **26**(8), 6095–6110 (2020)