Monitoring and Analysis of Distributed New Energy Resources Based on the Internet of Things

Hengzhi Cui\textsuperscript{1,*}, Chong Wang\textsuperscript{2} and Haixuan Liu\textsuperscript{3}

\textsuperscript{1}State Grid Jiangsu Electric Power Co., Ltd Jiangsu Nanjing 210029
\textsuperscript{2}Information and communication branch of State Grid Jiangsu Electric Power Co., Ltd Jiangsu Nanjing 210024
\textsuperscript{3}China Electric Power Research Institute Co., Ltd Beijing 100192

*Corresponding author: zjg_chz@js.sgcc.com.cn

Abstract. In order to accurately monitor and analyze new energy resources in real time, the analysis and design of a distributed new energy resource monitoring device based on the Internet of Things cloud platform is carried out, and a Hadoop distributed new energy system based on the Internet of Things cloud computing is designed. It analyzes the computing architecture of parallel collection and distributed storage of new energy resource monitoring data, analyzes the infrastructure system of the distributed system, and builds the big data processing architecture and computing service architecture for the electricity information collection system based on the Hadoop cluster. Tasks are processed in parallel to improve the computing efficiency of electricity consumption big data; at the same time, in order to solve the problem of insufficient storage space caused by the sharp increase in the amount of data in the electricity consumption information collection system and the difficulty of data interaction at all levels, a distributed database is developed. The modular design of the parallel data collection system of HBase and the data warehouse tool Hive, and the distributed data calculation design based on the MapReduce method have improved the efficiency of data collection and the unified storage of heterogeneous data. The research provides reference for realizing an efficient and safe new energy resource monitoring data processing system.

Keywords: Internet of things, distributed, new energy, resource monitoring, data analysis.

1. Introduction
In order to solve the problem of the accumulation of large amounts of multi-source heterogeneous data, domestic and foreign scholars have conducted in-depth research on the monitoring of new energy resources based on big data processing [1-3]. The main research directions are: ① storage of big data; ② calculation and analysis of big data; ③ the value of big data.

Scholars have conducted research on distributed power supply cloud services and big data analysis platforms for the global energy Internet, and designed distributed power supply cloud services and big data analysis platform business architecture, application architecture, data architecture and physical
architecture [4-5]. Model, realize the integrated application of computing, communication and power system, support distributed power cloud service and big data analysis platform for user behavior decision-making and intelligent optimization management; on this basis, the Hadoop cloud platform is used to study the distributed monitoring platform [6], A complete monitoring platform is realized, and the cloud monitoring system is responsible for resource guarantee, service application status, site operation status and availability, and log monitoring.

2. Distributed data storage

The difference between traditional centralized data storage technology and distributed data storage technology is that, distributed data storage technology is to store a large amount of data on each machine, while centralized data storage technology is to store data in multiple machines or on a machine. At present, the distributed data storage of the Hadoop platform is mainly embodied by HBase and HDFS [7-8]. Among them, HBase is based on HDFS and belongs to the columnar database, while the underlying support for distributed storage is mainly provided by HDFS. The framework composition of HDFS and the HBase system composition are shown in Figure 1.

![HDFS framework and HBase framework](image)

**Figure 1.** HDFS framework and HBase framework.

3. Design of distributed monitoring data processing system

3.1. System design

Since the monitoring, collection and analysis of new energy data require units of different levels to cooperate and operate together, the monitoring data of different regions is collected and stored by the main unit, and the main unit completes the corresponding data calculation and analysis. There are some problems in the current system [9-10]. One of them is the data of units at all levels are independent of each other, lack of mutual interaction and unified management. What’s more, the new
energy models are independent of each other. If advanced computing applications are applied to the entire network, data splicing is required. Increased workload, and real-time performance cannot be effectively ensured.

In order to achieve distributed computing or storage of massive amounts of new energy monitoring data, the Hadoop distributed system can not only provide basic support for distributed computing, but also provide massive data distributed storage functions, and realize new energy the unified management of automatic implementation, new energy monitoring data calculation/storage system architecture is shown in Figure 2.

![New energy monitoring data calculation/storage system architecture](image1)

**Figure 2.** New energy monitoring data calculation/storage system architecture.

Municipal and provincial-level units need to establish dedicated data high-speed channels for effective real-time data transmission. The logical structure of the cloud platform composed of new energy sources at all levels [11-13] is shown in Figure 3, mainly composed of main servers, It consists of a backup server, several clients, and several storage/computing servers.

![Logic of the cloud platform at all levels of new energy](image2)

**Figure 3.** Logic of the cloud platform at all levels of new energy.

3.2. *Cloud platform system composition*

In order to meet the needs of new energy units for the calculation and storage of monitoring data, the use of Pass (service form) to provide deployment platforms and applications to other units is to integrate the cloud computing platform (application development, application design, application hosting, application testing) Provided to users, the main advantages of the system are shown in Table 1.
Table 1. Main advantages of the system.

| Number | Project                               | Measure                                         | Advantage                                      |
|--------|---------------------------------------|-------------------------------------------------|-----------------------------------------------|
| 1      | Storage localization                   | establish separate private clouds               | Ensure the efficiency of network transmission and data security |
| 2      | Parallel data collection               | Use multiple parallel front-end units           | Improve collection efficiency                 |
| 3      | Data sharing                          | connect to other private clouds to obtain monitoring data | Used for whole network calculation |
| 4      | Data distributed processing            | The data is analyzed and calculated by the nearest computing server | Ensure the efficiency of data processing |
| 5      | System robustness                     | Main and standby server mechanism, multiple front-end machine mechanism | Ensure the stability of the system |

The data processing system using Hadoop can provide a fast, advanced and convenient operating platform for new energy data monitoring, and develop and build distributed automation system software (DMS, TMR, EMS) based on the cloud platform. The system structure is shown in Figure 4. The system is based on new energy monitoring data. The main components of the monitoring data are shown in Table 2.

![Figure 4. System structure.](image)

Table 2. Main components of monitoring data.

| Number | Composition        |
|--------|--------------------|
| 1      | SCADA Data         |
| 2      | PMU Data           |
| 3      | Fault recording data |
| 4      | New energy data    |

4. Design of distributed data storage scheme
The monitoring data of new energy mainly includes structured data (represented by SCADA data) and unstructured data (represented by relay protection data and recorded data). In order to fully consider the long-term development of new energy informatization, distributed storage systems require the realized functions are shown in Table 3.
Table 3. Distributed storage system needs to achieve the function.

| Number | Function                                      | Content                                                                 |
|--------|-----------------------------------------------|------------------------------------------------------------------------|
| 1      | Solving the heterogeneous problem of data    | design a new data storage structure to integrate and store data in different databases stored in a cluster of cheap machines, and the size of the storage space can be allocated or added as needed |
| 2      | Strong storage capacity                       | operate and access data resources in the network environment transparently |
| 3      | Efficient and transparent operation          | Faced with different application requirements, set the corresponding storage structure to use |
| 4      | Scalability and adaptability                 | Faced with different application requirements, set the corresponding storage structure to use |

4.1. System design
In order to meet the needs of the new energy monitoring system, the parallel data processing feature of the HBase database is used to design the data acquisition scheme. The parallel data acquisition system is shown in Figure 5.

![Figure 5. Parallel data acquisition system.](image)

As shown in Figure 5, the main advantages of the system are: ① Parallel collection; ② Division management; ③ Easy to expand; ④ Strong fault tolerance.

4.2. Structured data storage model design
The SCAD system is mainly composed of monitoring remote signaling data, remote measurement data, SOE data, electrical measurement data, etc. Based on the characteristics of HBase, taking remote measurement and remote signaling data as an example, the designed model is shown in Table 4.

Table 4. Structured data storage model design.

| Row Key            | YC value | YX state | Time Stamp |
|--------------------|----------|----------|------------|
| 201609025534_454   | 231.224  | 0        | T1         |
| 201609025534_454   | 229.473  | 1        | T2         |
| …                  | …        | …        | …          |
| 201609025534_454   | 231.654  | 0        | T1         |
| 201609025534_454   | 229.013  | 0        | T2         |
| …                  | …        | …        | …          |
In Table 4, with the manufacturer of the equipment, the unique identification number of the equipment, and the monitoring date as the primary keys, there are 2 column clusters (YX and YC). YX represents the remote signaling information of the storage device, and YX represents the remote measurement information of the storage device. Through practice Stamp Time Stamp to save the data.

5. Distributed data processing scheme

5.1. Distributed data processing architecture design
The monitoring data processing program based on the cloud platform is mainly represented by the processing of the fault record data in the project, and the Hadoop system is used to realize the efficient access and processing of the new energy monitoring data.

In order to realize the rapid diagnosis of the fault record information, the module is mainly based on the time selected by the user, according to the call time as the condition, calls the corresponding record source data, and then analyzes the data accordingly, so as to quickly and accurately find the fault point. The system structure of the distributed data processing scheme is shown in Figure 6 for the type of failure and the difference of the failure.

![Distributed data processing scheme system architecture](image)

**Figure 6.** Distributed data processing scheme system architecture.

5.2. Distributed data processing scheme design
In order to realize the rapid processing of monitoring data, the Map Reduce method is used to process the fault recorder data, which improves the data processing capability and carries out detailed design of the fault recorder data. The Reduce and Map operations are shown in Figure 7.

![Calculation models of Reduce and Map](image)

**Figure 7.** Calculation models of Reduce and Map.

Figure 8 shows the overall execution flow of distributed data processing tasks based on MapReduce. It can be seen from Figure 8 that the monitoring data is stored in the HBase database,
which is based on HDFS (Data Node). The data in HBase is obtained through the Map function, and then in a specific conversion format, a specific key-value pair is generated, where the key in the key-value pair (key, value) adopts a special combination form, and value is the corresponding monitoring data value.

Figure 8. Distributed data processing tasks to perform the overall process based on Map Reduce.

6. Conclusion
Through the method design of distributed new energy monitoring data based on the cloud platform, it solves the current needs of new energy system big data for storage and analysis and calculation, and solves the shortage of storage space caused by the rapid increase in data volume and unit data at all levels. The difficulty of interaction realizes the parallel writing of grid detection data, improves the efficiency of data collection and unified storage of heterogeneous data.

Acknowledgements
This research was funded by Science and Technology project of SGCC Research and Application of Service Design and Governance Technology Based on Data Middle Platform.

References
[1] W. Wang, M. Zhang, L. Zhang and Q. Bai, "Imbalanced Data Classification for Multi-Source Heterogenous Sensor Networks", in IEEE Access, vol. 8, pp. 27406-27413, 2020.
[2] I. Kiaei and S. Lotfifard, "Fault Section Identification in Smart Distribution Systems Using Multi-Source Data Based on Fuzzy Petri Nets", in IEEE Transactions on Smart Grid, vol. 11, no. 1, pp. 74-83, Jan. 2020.
[3] L. Shuai et al., "Multi-Source Feature Fusion and Entropy Feature Lightweight Neural Network for Constrained Multi-State Heterogeneous Iris Recognition", in IEEE Access, vol. 8, pp. 53321-53345, 2020.
[4] H. Chen, R. Kazman and S. Haziyev, "Agile Big Data Analytics for Web-Based Systems: An Architecture-Centric Approach", in IEEE Transactions on Big Data, vol. 2, no. 3, pp. 234-248, 1 Sept. 2016.
[5] M. Goudarzi, "Heterogeneous Architectures for Big Data Batch Processing in MapReduce Paradigm", in IEEE Transactions on Big Data, vol. 5, no. 1, pp. 18-33, 1 March 2019.
[6] H. Alshammari, J. Lee and H. Bajwa, "H2Hadoop: Improving Hadoop Performance Using the Metadata of Related Jobs", in IEEE Transactions on Cloud Computing, vol. 6, no. 4, pp.
1031-1040, 1 Oct.-Dec. 2018.

[7] A. Gadkari, V. B. Nikam and B. B. Meshram, "Implementing Joins over HBase on Cloud Platform", 2014 IEEE International Conference on Computer and Information Technology, Xi'an, 2014, pp. 547-554.

[8] C. Lin, C. Li, W. Huang, W. Liao and W. Chen, "A Sensor Data Processing and Access Platform Based on Hadoop for Smart Environments", 2014 17th International Conference on Network-Based Information Systems, Salerno, 2014, pp. 455-460.

[9] M. M. U. Rathore, A. Paul, A. Ahmad, B. Chen, B. Huang and W. Ji, "Real-Time Big Data Analytical Architecture for Remote Sensing Application", in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 10, pp. 4610-4621, Oct. 2015.

[10] K. Lavanya and G. Murali, "Efficient analytical architecture for sensor networks using Hadoop", 2016 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, 2016, pp. 1-6.

[11] X. Sun and N. Ansari, "Green Cloudlet Network: A Sustainable Platform for Mobile Cloud Computing", in IEEE Transactions on Cloud Computing, vol. 8, no. 1, pp. 180-192, 1 Jan.-March 2020.

[12] M. Pau et al., "Design and Accuracy Analysis of Multilevel State Estimation Based on Smart Metering Infrastructure", in IEEE Transactions on Instrumentation and Measurement, vol. 68, no. 11, pp. 4300-4312, Nov. 2019.

[13] A. Raha and V. Raghunathan, "Approximating Beyond the Processor: Exploring Full-System Energy-Accuracy Tradeoffs in a Smart Camera System", in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 26, no. 12, pp. 2884-2897, Dec. 2018.