Access Data Analysis Technology and Implementation of Electric Power Big Data Achievement Sharing Platform through Artificial Intelligence

Wang Qian1, *, Mi Na1, Yi Zenan1, Sun Ming yue1, Liu Qing2

1 Big Data Center of State Grid Corporation of China; Beijing 100052, China
2 State Grid Info-Telecom Great Power Science and Technology Co., Ltd.; Fuzhou Fujian 350003, China

*Corresponding author: wang_qian@sgcc.com.cn

Abstract. With the rapid development of the information age, the data generated by all walks of life is showing an increasing trend of "blowout". According to statistics, the total amount of data generated by mankind in the past 40,000 years is less than the total amount of data generated from 2010 to 2013 alone. The global big data reserves reached 8.61ZB in 2015 alone, and the growth of data in the future will reach an unpredictable value, entering the data age in an all-round way. Traditional power data computing technology and intelligent analysis technology are undergoing profound changes, and emerging big data intelligent analysis platforms are gradually emerging. With the in-depth development of power informatization and the concept of smart grid, the power industry data has grown exponentially, and the business demand for intelligent analysis of large amounts of power data is increasing day by day. Therefore, the access data analysis technology and realization of the power big data achievement sharing platform based on AI is of great significance. The AI-based power big data achievement sharing platform constructed in this article is a research on previous data analysis, and it aims to use the current cutting-edge artificial intelligence technology to build a scalable and highly available power big data analysis and processing platform to provide fast and reliable smart data services for the power industry, smart grid and other grid businesses. Research has shown that the overall availability of the ultrasonic partial discharge recognition system of the artificial intelligence-based electric power big data achievement sharing platform is 99.9967%, which meets the high availability index requirements, and verified that the artificial intelligence-based power big data achievement sharing platform provides highly available computing, storage and other services for its applications.

Keywords: Artificial Intelligence, Electric Power Big Data, Achievement Sharing, Platform Access.

1. Introduction

With the development of big data and AI, traditional data computing technology and intelligent analysis technology are undergoing profound changes, and emerging big data intelligent analysis
platforms are gradually emerging [1-2]. With the in-depth development of power informatization and the concept of smart grid, the power industry data has grown exponentially, and the business demand for intelligent analysis of large amounts of power data is increasing day by day. Therefore, the access data analysis technology and realization of the power big data achievement sharing platform based on AI is of great significance [3-4].

In recent years, the problem of global energy shortage has become increasingly prominent, which is a severe test for the sustainable development of mankind. Countries around the world have successively carried out research work on smart grids. Faced with the four general characteristics of big data: volume (sea quantification), variety (diversity), velocity (rapid), value (value) [5]. Electric power big data also has these four characteristics. Massive quantification means that the amount of data is extremely large. Nowadays, PB, EB, and ZB-level data are the norm. How to store massive amounts of data will be the cornerstone of big data. Diversification means that the types of data become increasingly complex with the development of business diversification. How to structure massive data is the bridge of big data [6]. Rapidization refers to the rapid real-time iterative update of data in the form of streams, from minutes to seconds to milliseconds. How to process massive data in real time is an accelerator of big data [7]. Valuation refers to the common problem that data has sparse value density. How to turn massive data into treasure is the ultimate goal of big data. From a technical perspective, the continuous development of modern smart grid-related technologies and the advancement of big data technology have contributed to the proposal of big data in electric power [8]. Among them, the development of the main related technologies of modern smart grids includes related technologies of smart substations, such as IEC61850, a new generation of protection devices, a new generation of telecontrol systems, and so on. The development of big data technology mainly refers to the proposal of new derivative products and a new generation of computing framework after the Hadoop ecosystem is proposed, such as Spark, HBASE, etc. [9-10].

In this paper, big data technology is introduced into the field of power data analysis. Aiming at the problems of low efficiency, poor reliability, and high cost of processing massive data in existing systems, an AI-based power big data achievement sharing platform is proposed. It is planned to adopt cheap server clusters, distributed management software, and with the help of mature data mining algorithms to realize the effective analysis and processing of massive electric power big data [11-12].

2. Access Data Analysis Technology and Implementation of Electric Power Big Data Achievement Sharing Platform Based on AI

2.1. System Requirements Analysis

(1) Analysis of overall system requirements

The main requirement of an AI-based power big data achievement sharing platform is to start from the bottom, from the unified storage and management of traditional power system data to the highest-level power system customized calculation and analysis services. The overall system demand analysis is shown in Figure 1.
In general, the system requirements are divided into 5 levels. The first level is to store data securely, unified access, and securely access and manage, so as to realize distributed storage of power big data and safe and fast access. The second level is to provide basic support for data visualization and data reporting. This demand is an implicit demand and cannot work independently, but provides service support for the upper-level needs, so that the upper-level needs can be met. The third level is the core power big data calculation and analysis function support module, which realizes the determination of data rules for power big data, basic statistical analysis and basic machine learning algorithm analysis, and meets the basic needs of power big data calculation and analysis. The fourth layer is the algorithm programming entrance and task management services for big data professionals, so as to meet the needs of basic statistical analysis and big data calculation method analysis to carry out further programming calculations and analysis of electric power big data. The last layer is the customized calculation and analysis service of the power system for senior management personnel of the power system. The main function is to solidify some more mature theoretical analysis results into customized analysis and calculation solutions, so that the big data of the power system can be quickly processed.

Figure 1. Analysis of overall system requirements

(2) Hierarchical analysis of system requirements

The basis of the AI-based power big data achievement sharing platform is to store the power big data for access and calculation. On the basis of being able to store, provide service support for secure access, fast access, and unified management.

The underlying requirements of the system are first of all basic data storage requirements, which include distributed file system services and distributed non-relational database building services. Traditional power system data uses relational database storage to meet typical data business logic requirements. However, after entering the era of big data, traditional relational databases can no longer meet the needs of power big data calculation and analysis, and require larger distributions. Therefore, the basic requirement of power big data is to introduce a distributed file system, so that traditional power big data can be stored and managed in a distributed manner, so as to share computing tasks, improve the computing performance of power big data, and meet true power big data calculation and analysis.

After completing the requirements for the storage and access management of power system data, then naturally the visualization and report support for power system big data are further requirements.

The system's visualization support requirements are friendly UI support and visualization mapping services for the complex system backed on the basis of power big data storage and access management. It mainly includes the core UI engine requirements of the system. The first layer of the system needs to be able to store and access management of the power system data. After the first layer needs are met,
the backed of the system is still obscure and difficult to humanize operations. Only professionals can manage and operate them. It is difficult to meet the original intention of the system. At this time, the system's humanized interface requirements came into being, and the basis of the humanized interface is the system's core UI engine requirements.

Finally, there is the demand for drawing the results of the system calculations. After satisfying the system's basic UI engine requirements and other visualization requirements, the last independent thing is the system's calculation result mapping requirements, that is, the visualization requirements in the real big data field. This visualization requirement is based on the basic data visualization, allowing the system to call various drawing engines to draw various complex data graphics, so as to realize the visualization support of the real data and calculation and analysis results. The quality of the drawing structure can directly affect the overall big data calculation results of the system and the user experience, so it is a very important independent requirement. The functional modules that meet this requirement are often composed of several different independent drawing engines, so as to meet different drawing requirements such as pie chart, line chart, cube chart, scatter chart, etc., and combine drawing functions to draw complex data result graphics, so as to meet the needs of all kinds of complex calculation results presentation, and give full play to the potential drawing ability of drawing construction.

2.2. Design of Overall System Function Module

The overall functional module design of the system can be divided into platform front-end design and platform back-end design. The back-end of the platform mainly realizes data storage management and calculation of electric power big data, and the front-end of the platform focuses on the realization of system operation business logic functions.

The overall function module of the power big data achievement sharing platform based on AI can be divided into two parts, the distributed data storage management and computing module and the system business logic function realization module. The module can also be physically divided into the platform front-end module and the platform back-end module. The back-end modules include business computing core modules, business support modules, and data storage and access management modules. The front-end modules mainly include platform business support basic modules, computing core business modules, and power big data custom analysis and calculation support modules.

2.3. Related Algorithms

Time domain characteristics

(1) Variance

In signal processing, the variance calculates the degree of deviation between the signal strength and the overall mean value. The calculation formula is as follows:

\[ VAR = \frac{1}{N-1} \sum_{i=1}^{N} (x(i) - \bar{x})^2 \]

(2) Absolute Integral Average

Absolute Integral Average can represent the average intensity of the ultrasonic audio signal. The calculation formula is as follows:

\[ A VA = \frac{1}{N} \sum_{i=1}^{N} |x(i)| \]

(3) Kurtosis

Kurtosis is used to characterize the degree of aggregation of signal strength, and the calculation formula is as follows:
3. AI-Based Power Big Data Achievement Sharing Platform Experiment

3.1. Hardware Deployment
The hardware environment uses 3 physical servers to build a cluster, one master node as the control node of the cluster; two slave nodes as the working node of the cluster. The parameters of the three servers are: 8-core processor: Intel(R)Xeon (R)CPUE5606@2.13GHz; memory size: 16GB.

3.2. Software Deployment
The platform software environment deployment includes: Java environment deployment, Scala environment deployment, Zookeeper cluster deployment, Kafka cluster deployment, HBase cluster deployment, Hadoop cluster deployment, Spark cluster deployment, Tomcat deployment.

After calling each software script program to start, you can view the process status of the master node and the slave node through Linux instructions to determine whether the cluster software is successfully deployed and started.

3.3. Platform Operation Test and Analysis
The AI-based power big data achievement sharing platform designed based on the Lambda architecture idea splits the whole into a batch processing layer, a stream computing layer and a service layer, and the platform operation process is therefore divided into three independent operation processes. The batch processing layer and the stream computing layer are deployed in the form of two separate Spark applications and run in a Spark cluster; the service layer is deployed in the form of a web service and run separately in Tomcat.

4. Experimental Analysis of Power Big Data Achievement Sharing Platform Based on AI

4.1. Verify the Efficiency of the AI-Based Power Big Data Achievement Sharing Platform
In order to verify the efficiency of the AI-based power big data achievement sharing platform, this section verifies the platform's computing performance and real-time processing performance.

First, verify the computing performance of the power big data sharing platform based on AI, and use a larger data set for full data training of the K-Means algorithm model. The data set has a total of 146,155 records. Compare the SSHA platform and the local PC (Personal Computer). Calculation time. The local PC uses the matlab program to train the K-Means algorithm model on the data set. The software environment of the SSHA platform is different, but the hardware environment is the same.

The local PC configuration parameter is a 4-core processor: Intel(R) Core (TM) i5-3210M CPU@2.50GHz; Memory size: 8GB. The experimental results are shown in Table 1:

| Computing platform | This platform | PC          |
|--------------------|---------------|-------------|
| Time used (s)      | 354.251       | 1345.254    |

Through experiments, the AI-based power big data achievement sharing platform model training takes about 355 seconds, and the PC takes about 1385 seconds. The computing efficiency of the SSHA platform is about 4 times that of a single machine. Due to the experimental setup that the number of K-Means model training iterations is small, only 50, and the K-Means model training time complexity is linear, the computational cost is not high, so it fails to fully show the advantages of the SSHA platform. If a model with higher time complexity for iterative calculations is selected, the computing advantages

\[
BK = \frac{\sum_{i=1}^{N} (x(i) - \bar{x})^4}{(N-1)sd^4}
\]
of the AI-based power big data achievement sharing platform will be more obvious. The basic computing performance test verifies the efficiency of the AI-based power big data achievement sharing platform.

Secondly, to verify the real-time processing capabilities of the platform, the experimental idea is: send a piece of data to the AI-based power big data achievement sharing platform every 1 second, and calculate the average time of the platform flow calculation. The test time is about 28 minutes. The experimental results are shown in Table 2.

| Send volume | Actual reception | Data loss rate | Average scheduling delay time | Average processing delay time | Average total time delay |
|-------------|------------------|----------------|-------------------------------|-------------------------------|-------------------------|
| 1652        | 1652             | 0%             | 134 ms/item                   | 254 ms/item                   | 388 ms/item             |

Experimental results show that the real-time processing delay of the AI-based power big data achievement sharing platform is 388 milliseconds, which can meet the second-level response requirements of general big data real-time computing scenarios; the data loss rate is 0%, ensuring zero data loss and meeting the requirements of platform high availability and data disaster tolerance also further verify the feasibility and availability of the middleware "zero data loss" mechanism and the high availability mechanism of the data storage layer.

4.2. Analysis of Partial Discharge Recognition Results
A total of about 800 ultrasonic audio samples are divided into training set and test set at a ratio of 7:3, and they are respectively sent to the system data preprocessing module for preprocessing and feature extraction. After preprocessing, the training set is fully read by the batch processing layer of the AI-based electric power big data achievement sharing platform into the model training for the SVM recognizer, and the test set is continuously sent to the AI-based electric power big data achievement sharing platform. The stream computing layer is used in the test experiment of the recognizer, and the experimental results are shown in Figure 2.

Figure 2. Partial discharge SVM recognizer test results
The new features in the figure refer to the 4-dimensional time-domain features and the 3-dimensional frequency-domain features; the existing features refer to the 15-dimensional feature parameters such as the Mel cepstrum coefficient MFCC in the results of previous studies. From the results of recognition accuracy, the new features increase the recognition rate from 80.49% to 85.37%, and the recognition effect is optimized by nearly 5 percentage points.

Based on the experimental results, the overall availability of the ultrasonic partial discharge recognition system of the AI-based electric power big data achievement sharing platform is 99.9967%,
which meets the high availability index requirements, and verifies that the AI-based electric power big data achievement sharing platform provides high performance for its applications, available computing, storage and other services.

5. Conclusions
In today's era of rapid development of smart grids, its accompanying problems have also emerged. Many system operating data cannot be stored because of the high collection frequency and the large amount of data units. Even if it is solidified storage, the storage time scale is also small. The overall data capacity is still too small. In the provincial control and operation center of the entire power system, the server load is close to saturation, and effective storage and access to data has been difficult. It is even more fable to perform complex analysis and calculation on the data. Therefore, the power system continues to be a new generation of data centers, and can undertake powerful data storage and computing tasks to explore the potential value of the massive data in the power system. The AI-based power big data achievement sharing platform proposed in this paper is proposed in response to this background, and the goal is to deploy a software system in the new generation of power system data. In the next 10 years, or even 5 years, the darlings of technology in the new era, such as AI, may replace big data technology and become everyone's new focus. Another convenience is that there is still much room for improvement in the data storage and management of the power system, waiting for the people of the power system to improve. At the same time, the Internet of Energy has risen rapidly again, grabbing reform resources.

References
[1] Lee J Y, Yoon J S, Kim B H. A big data analytics platform for smart factories in small and medium-sized manufacturing enterprises: An empirical case study of a die casting factory [J]. International Journal of Precision Engineering & Manufacturing, 2017, 18(10):1353-1361.
[2] Marek, Grabowski, Wladek, et al. Sharing Big Data[J]. IUCrJ, 2017, 4(1):3-4.
[3] Zhang Y, Mi Z. Environmental benefits of bike sharing: A big data-based analysis[J]. Applied Energy, 2018, 220(JUN.15):296-301.
[4] Frazier J A, Hodge S, Poline J B, et al. 2.7 Everyone Likes Candi: Data Sharing and Neuroinformatics to Address Big Data Questions[J]. Retour au numéro, 2017, 56(10): S138.
[5] Kang C, Wang Y, Xue Y, et al. Big Data Analytics in China's Electric Power Industry[J]. IEEE Power and Energy Magazine, 2018, 16(3):54-65.
[6] Wu W T, Lin W W, Hsu C H, et al. Energy-efficient hadoop for big data analytics and computing: A systematic review and research insights[J]. Future Generation Computer Systems, 2017, 86(SEP.):1351-1367.
[7] Sanderson C, Pang T M. The challenge of level platform access: A best practice approach [J]. Permanent Way Institution Journal & Report of Proceedings, 2017, 135(pt.4):28-31.
[8] Scassa T. Sharing Data in the Platform Economy: A Public Interest Argument for Access to Platform Data[J]. University of British Columbia law review. University of British Columbia, 2017, 54(4):1017-1071.
[9] Ferhatoglu S Y, Kudsioglu T. Evaluation of the reliability, utility, and quality of the information in cardiopulmonary resuscitation videos shared on Open access video sharing platform YouTube[J]. Australasian Emergency Care, 2020, 23(3):211-216.
[10] Almagrabi A O. A Pervasive Controlled Access with Privacy Delegation Design for Smart Internet of Things Applications [J]. Measurement, 2020, 172(2):108875.
[11] Cui L, X Wang, Su S C. Techniques Used on E-Clinic Portal Access Control[J]. International journal of computational intelligence research, 2018, 14(1):79-88.
[12] Zhu Q, Fisher S A, Dueck H, et al. PIVOT: platform for interactive analysis and visualization of transcriptomics data [J]. Bmc Bioinformatics, 2018, 19(1):6.