

Research on Wide-area Distributed Power Quality Data Fusion Technology of Power Grid

Li Jin, Li Haosong, Xu Zhongping, Wang Ting, Wang Shuai, Wei Yutong, Hu Dongliang, Kang Chunting
 State Grid
 Beijing Sgitg Accenture Information Technology Center
 Co., Ltd.
 Beijing, China
 e-mail: hudongliang@sgitg.sgcc.com.cn

Wu Jia, Su Dan
 State Grid
 State Grid Jibei Electric Power Co.,Ltd.
 Beijing, China
 e-mail: alice.0729@163.com

Abstract—With the advancement of the “big operation” system construction, the online monitoring system for power quality has been integrated, and various power quality data have been incorporated into relevant organizations for unified management. Power quality management has a larger range of data, more types, and higher frequency. It needs to realize the unified storage management and efficient access of massive heterogeneous power quality data for the characteristics of data applications and the collection and aggregation of these effective data. This paper proposes a new type of grid wide-area distributed power quality data integration architecture, which is designed for multi-source, heterogeneous, distributed data integration technology and wide-area distributed data storage technology to solve the big data source problem and realize the sharing of power quality data information of the whole network.

Keywords—Power Quality; Wide Area Distribution; Data Integration

I. INTRODUCTION

Data fusion refers to the integration of interconnected distributed heterogeneous data sources and maintains overall data consistency to improve the efficiency of information sharing which enable users to access these data sources in a transparent manner. In the field of enterprise data integration, there are already many mature frameworks that can be utilized.

A. Data Integration Technology

The traditional methods of data integration include pattern integration method and data replication method. The pattern integration method is divided into federated database and middleware technology. These technologies solve data sharing and provide decision support for enterprises in different priorities and applications. With the further development of data integration technology, new solutions emerged, including distributed object technology (DOC) and XML data integration.

Schema Intergration is the first data integration method which is also the basis for other data integration methods. The basic idea is to integrate the data views shared by each data source into a global mode when the integrated system is

built so that the user can transparently access the data of each data source according to the global mode. The method does not need to store a large amount of data repeatedly which can make sure that the latest data is available and is suitable for integration of heterogeneous data sources with large integrated data and fast update changes. The data replication method is to access the source database indirectly by accessing the central database by copying the data of each data source into the central database. Data replication can be a copy of the entire data source or one of some data which have been changed.

DCOM is a relatively typical distributed object technology (DOC) whose largest user is Microsoft. DCOM is a further extension of the Object Model (COM) which defines the way components interact with each other to make the components communicate with the clients without any intermediary components and customers access the methods in the components directly through the interface provided by the component object. XML stands for Extensible Markup Language which is a system that defines other languages. It enables that the same XML source documents is written once and can be expressed in different ways by the means of separating structure, content, and performance. The working mechanism of xml data integration is shown in Figure 1:

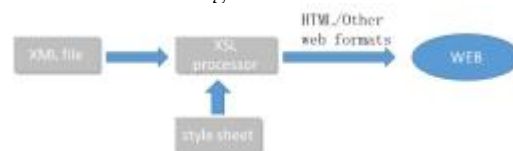


Figure 1. xml data integration working mechanism

B. Data Management

In order to solve the problems of redundant power quality data indicators (including voltage deviation, frequency deviation, harmonics, three-phase unbalance, voltage fluctuation and flicker, transient data, etc.), large business calculation amount and big data volume in the the power quality data fusion application scenario, a wide-area distributed data storage, computing and management technique should be studied. It includes three aspects: distributed physical storage structure design pattern, data management method and computing mode. The data storage

mainly uses the HDFS which is a underlying storage architecture and the data calculation uses Spark which is a distributed computing technology.

C. Heterogeneous Data Sharing

In order to solve the problem of different communication modes, data storage methods and time tags in date sharing through multi-type heterogeneous systems, mainstream heterogeneous data sharing technologies mainly include two types which are Webservice and XML-based technologies. Webservice is a brand-new distributed computing model that enables different systems to communicate and share data easily. It has good encapsulation, loose fit and high integration ability which uses standard protocol specifications. It is a popular new platform for interoperable distributed applications. XML (ExtensibleMarkupLanguage) is a simple and flexible text format extensible markup language which originated from SGML (Standard Generalized MarkupLanguage) and is a sub-collection and simplified version of SGML. It is very suitable for the Web or other data sources to exchange data.

II. TECHNICAL SCHEME FRAMEWORK

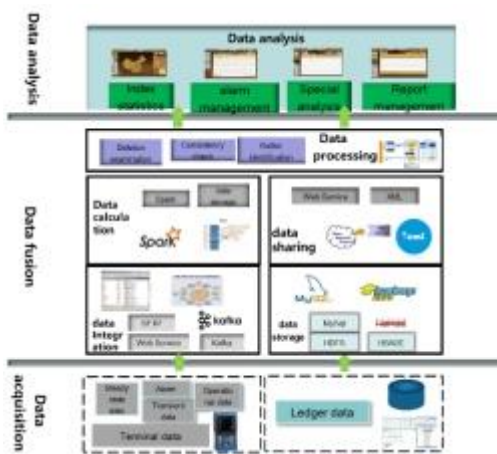


Figure 2. Technical architecture of the overall power quality monitoring platform

The wide-area distributed power quality data integration scheme proposed in this paper is based on a unified power quality data integration project which consists of cross-domain and heterogeneous data. It can improve the sharing of power quality data and reduce the isolation of data to promote the application level of the advanced computational analysis function in power quality system and supports the power quality data analysis by providing data sources.

The technical architecture includes data acquisition, data fusion and data analysis. Data collection is used to collect basic data of power quality. Data fusion technology is used to integrate, store, calculate, share and process basic data. Advanced power quality data analysis is based on fusion data. The architecture diagram is shown in Figure 2.

The data collection is mainly from different data sources according to the power quality monitoring services needs to meet the basic requirements. It includes power quality monitoring terminal data and PMS accounting information.

Data fusion mainly performs a series of integration processing on the basic data according to business needs and the requirements and characteristics of power quality data which includes the data integration technologies (Webservice, Sftp, Kafka, etc.), the data storage technologies (HDFS, HBASE, MySql, etc.), the data calculation technologies (Spark, stored procedures technology, etc.), the data sharing technologies (Webservice, XML, etc.) and the data preprocessing technologies (missing verification, consistency check, outlier identification, etc.).

The data analysis application mainly counts the fusion data and analyzes the power quality problems existing in the power grid in combination with the power quality standard specification. At the same time, the analysis results are visually displayed to the user through charts and evaluation reports to provide governance decisions which includes metrics, alarm management, special analysis, and report management.

III. TECHNICAL SCHEME DESIGN

This program focuses on the design and implementation of data fusion technology solutions which includes data integration, data storage, data computing, data sharing and data processing.

Data integration

Based on the above analysis of characteristics of wide-area distribution, multi-source system, heterogeneous, massive of power quality data, the data integration architecture we designed is shown in Figure 3.

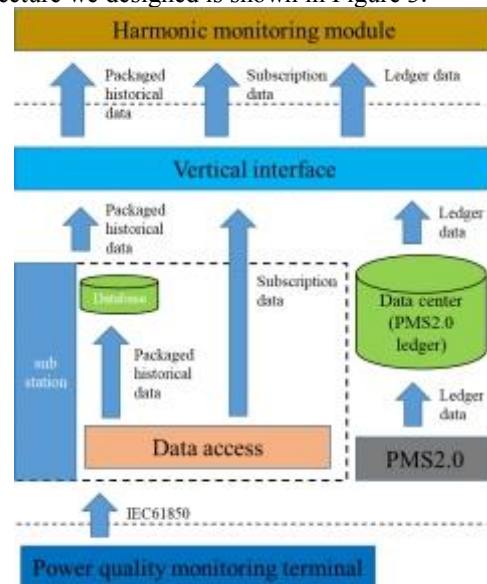


Figure 3. Data integration architecture

After the provincial power quality monitoring terminal data is connected to the communication front-end machine according to the IEC61850 protocol, the historical data is processed, compressed packed and stored in their database server by their sub-station, and then periodically collected into the headquarters platform through the Webservice interface and sftp. As for the subscription (real-time) data, it is accessed and collected into the headquarters platform by the

communication front-end machine through the Webservice interface and the Kafka message service bus.

The parameter data is periodically pushed to the provincial data center by the PMS2.0 system, and then the PMS2.0 related parameter data of the provincial data center is collected into the headquarters platform through the Webservice interface.

This study uses Web Service, Sftp, Kafka to build interfaces to achieve data integration. The interfaces are divided into two categories: data interface and query/control interface.

There are three data interfaces, namely the parameter data interface, the historical data interface and the real-time data interface. The parameter data interface and the historical data interface are deployed on the main station side of the provincial company, and the real-time data interface is deployed on the main station side of the headquarter.

For each type of data interface, there is a corresponding query/control interface which is deployed on the provincial company main station side. The data integration interface is shown in Figure 4.



Figure 4. data integration interface

1) Data Storage

The basic data of power quality comes from various collection point monitoring terminals. The existing traditional relational database cannot adapt to the storage of massive data. To solve the current storage problem, HBase is selected as the storage method of massive power quality basic data.

The power quality calculation result data is much smaller than the basic data. However, HBase can not reflect its superiority in the use of mall data set storage and there will be higher query delay of data access than the general relational database. Considering all the different performance of databases, we choose MySql relational database as the storage container of power quality result data. The storage capacity can be 524PB and is enough to store all the result data of this scheme. We can access all kinds of required relation data by using simple SQL statements query. MySql also supports Web pages and the analysis results can be published to the Internet or displayed on the webpage to achieve higher degree of visualization.

2) Data calculation

The gradual intelligence of power grid equipment enables systems such as power data acquisition and monitoring to continuously generate large-scale time-series data streams with real-time and infinite features. The data streams value decreases with time, so large-scale streaming data processing technology is in uegent need. In this paper, we use

distributed parallel processing based on Spark and its framework diagram is shown in Figure 5.

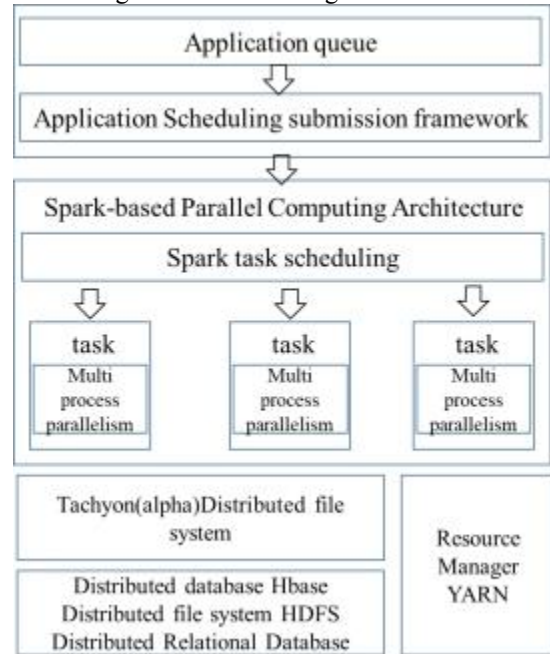


Figure 5. Power grid data distributed parallel processing framework diagram based on Spark

From top to bottom of the framework diagram, there is the grid data application management module, the Spark-based parallel computing framework, the distributed memory file system Tachyon, the data storage system and the resource management system YARN in turns. After the grid data application is submitted, the application management service component checks and parses the parameters input by the user at the same time and decomposes the grid data application into several grid data processing jobs, and in the end, submits them to the job queue. The data processing jobs are executed directly by single-machine processing procedures and parallel processing jobs are submitted to the calculation framework, schelduled to excute when there are idle computing resources. For large-scale streaming computing operations for grid condition monitoring applications, Spark will break it down into a series of short batch jobs, which divide the input data into discrete data processing at a certain time interval or unit batch throughput. Every part of data is assigned to eache node for computing in Spark. The calculation results of each node are combined and stored in the memory and are superimposed or saved to the disk according to the calculation result of the latter part of the data. The performance achieved by Spark multi-machine multi-process data parallel processing can meet the requirements of real-time monitoring of grid status.

3) Data Sharing

In order to solve the difference between the terminal data model of each manufacturer and the terminal communication modes of multi-vendor, and in the mean time, realize data sharing for other systems, the power quality terminal data

sharing is realized through application layer network protocol, consistent measurement data labels and time labels designing.

Based on IEC61850 protocol, the unified power quality online monitoring device data model includes the modeling of logical device nodes, data objects and data attributes, logical device modeling, communication service modeling, etc. It realizes power quality monitoring device data sharing between different equipment manufacturers.

In order to eliminate the differences of terminal data models between different manufacturers, by defining a unified quaternary format specification and converting all kinds of power quality terminal monitoring data into quaternary data to unify measurement data labels and time labels, we eliminate format differences between each power quality terminals and realizes the sharing of power quality data. We can send every provinces' unified data to the headquarter and share the calculated quaternary data to other organizations according to their system requirements.

The quaternary data consist of four fields: monitoring points, measurement indicators, measurement timing and measurement values. These four fields data is processed of all power grid data according to the requirements of power grid monitoring business needs.

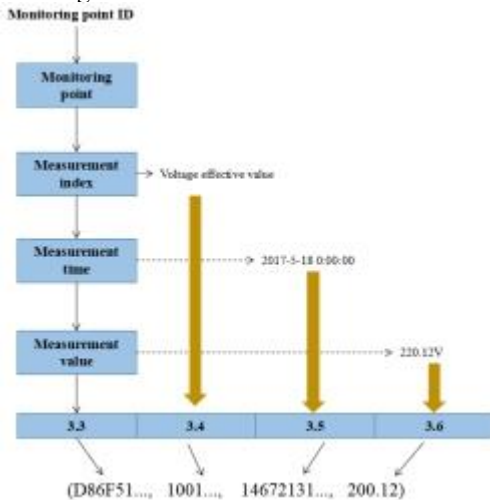


Figure 6. Quaternary example diagram

IV. CONCLUSION

This paper focuses on the characteristic analysis of wide-area distributed power quality data, feature analysis of power quality data, power quality data fusion technology and power quality data integration mode and proposes a wide-area

distributed power quality data fusion architecture for power grid. In the future, we will further deepen the theory and application of project research results, expand the breadth of project research and include other business system data indicators into the research scope. We will strive to develop scale, speed, security, quality and efficiency of the power grid in order to analyse the development of its data and discover main problem in the key lines under the support of accurate and rich fusion architecture. We will also do further research in how to enhance the comprehensive sharing of power grid data based on rich and accurate data fusion architecture.

REFERENCES

- [1] Han Qiang, Hang Gang, Feng Tao, Liu Di. Smart Grid Multi-service Data Integration Technology[J]. Electric Power IT. 2012, 10(12):92-96.
- [2] Zhang Liuliu, Ji Kun, Liu Yang, Deng Dawei. Research and implementation of data integration technology for integrated power grid operation system[J]. Electrical application. 2015, S2:428-432. 2012, 10(12):92-96.
- [3] Li Shuguang, Qi Linhai, Ma Suxia. Data processing technology in power quality analysis system[J]. ELECTRIC POWER IT. 2008, 6(12):109-112.
- [4] Wang Shupeng. Research on power quality information system based on IEC 61970[D]. Beijing: North China Electric Power University, 2013.
- [5] Zhang Yi, Yang Honggeng, Ye Maoqing. A Data Management Scheme for Massive Power Quality Monitoring Data Based on Distributed File System[J]. Automation of Electric Power System. 2014, 38(2):102-108.
- [6] Zhang Huaying, Yao Senjing, Geng Jiewen, Wang Xin. Research of Data Integration Technique for Wide-Area Distributed Power Quality Monitoring System[J]. Electrical and energy efficiency management technology. 2016, 17:62-66.
- [7] Song Yi, Xue Zhenyu, Teng Lin, Jing Hai. Data Integration Technology Research and Application in Integrated Distribution Network Planning Platform[J]. Power System Technology. 2016, 7(40):2119-2205.
- [8] Matei Zaharia, Mosharaf Chowdhury, Tathagata Das, et al. Resilient distributed datasets: a fault-tolerant abstraction for in-memory cluster computing[EB/OL]. 2011[2015]. <http://www.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-82.html>
- [9] Gregorio Convertino, Umer Farooq, Mary Beth Rosson, et al. Supporting intergenerational groups in computer-supported cooperative work[J]. Behaviour & Information Technology. 2007, 26(4):275-285.
- [10] Yang Li, Tang Chaobo, Li Beibei, Chen Ya. Research on Holographic Time-Scalar Measurement Data Integration and Applied Technology for Power operation[J]. Computer Applications and Software. 2015, 32(6):120-126.