Research on Heterogeneous Data Fusion of Multiple Energy

Jie Song[1, *], Jiancheng Yu[2], Na Tian[2], Jianxu Liu[2], Xianxu Huo[2,] Jinming,Wang[1]

[1]NARI Technology Development Co., Ltd., Nanjing 211106, China
[2]State Grid Tianjin Electric Power Company, Tianjin 300010, China

Abstract. With the development of information and automation system related to smart city and new energy, the municipal system and power system produce a large number of heterogeneous data every day, including structured data, semi-structured data and unstructured data. These data are often stored in different forms in various business systems, fragmented among each other. However, in the process of large data of multi energy applications, these data need to be integrated before the analysis of them. In this paper, we study the typical characteristics and storage technology of heterogeneous data of multi energy in power system, and propose the architecture of heterogeneous data fusion system.

1 Introduction

In the process of the development of China's smart city and the construction of new energy, "Information Island" has become an important fetter for the development of smart city and new energy. The city's various data are relatively isolated and closed, such as municipal buildings, streets, transportation, lighting and other data, often not open to the outside world; Similarly, the relevant information of new energy generation, due to the reasons for the management mechanism and the basic conditions of communication, it is often difficult to achieve information integration and data sharing with the related business system of power system. Therefore, to achieve the integration and sharing of municipal pipe network data, multi energy data, power system data, new energy generation data and so on, build the integrated energy data service platform of smart city, realize the information interconnection of smart city, is the foundation for the development and construction of smart city, regional energy optimization operation.

Based on the above background, this paper focuses on the typical characteristics and storage technology of heterogeneous data, and puts forward the architecture of multi energy data fusion. The application of data fusion technology will provide basic support to establish a multi energy large data service platform, to promote energy development, configuration, consumption mode reform, to achieve large-scale clean energy and distributed power paralleling in grid friendly, to realize optimization configurations of various energy transparent and open, to enhance new experience of intelligent power and interactive services.

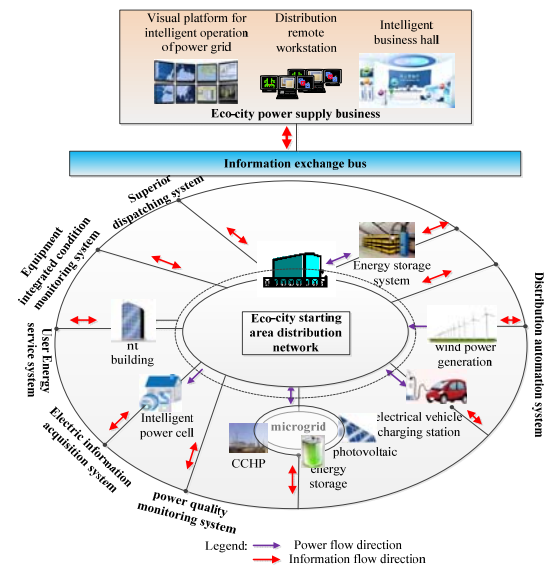## 2.Sources and characteristics of heterogeneous data of multiple energy

Fig. 1 Eco-city smart grid application system architecture diagram

November 18, 2007, the Prime Ministers of China and Singapore signed a framework agreement with the government of the people's Republic of China and the government of the Republic of Singapore on the construction of an ecological city in the people's Republic of China, to determine the cooperation between China and Singapore in the construction of Sino-Singapore Tianjin Eco-city (hereinafter referred to as eco-city). As a part of the eco-city, integrated demonstration project of smart grid of Sino-Singapore Tianjin eco-city started to construct in 2010, is a comprehensive project which is rich in connotation, involving a wide range and reflecting research results of State Grid Corporation of China in the field of smart grid. Many renewable energy generation technologies applied in the demonstration area. In accordance with the overall planning requirements of the eco-city, the expected rate of renewable energy use will reach 20%, of which, the solar power will reach 10MW, biomass power generation will reach 125MW, wind power generate 1.489MW, animation Park No. two power station gas for power generate 1.489MW.

Distributed access is the main energy access method of all kinds of energy.

In order to optimize and dispatch these multivariate energy generation systems, the eco-city from a unified perspective of the power flow, information flow, business flow, designs the overall architecture of the 6 application systems and information exchange bus, as shown in Fig. 1.

These six major applications include: 1) Distribution automation system; 2) Intelligent power network equipment integrated condition monitoring system; 3) Power quality monitoring system; 4) Electricity information collection system; 5) User energy use service system; 6) Power grid intelligent operation visualization system.

Above six applications get a lot of heterogeneous data through various sensors, control systems and other automation systems, as well as manual, semi-manual data sources and they are stored separately. From the data type, the data generated by the system include general structured data, semi-structured data and unstructured data. In this paper, these data types are summarized, as shown in Table 1.

Table 1 Heterogeneous data types of multiple energy sources in eco-city

Data type	sub types	typical data
Structured data	relational database data	load data
		AC analog data
		Power quality data
Semi structured data	Spatial vector data	power GIS data
	Three dimensional model data	underground pipe network data
	File storage data	Excel table
Unstructured data	Picture data	Monitor image
	Video data	Monitor video stream

Due to the operation of the main body and management mode, the above data is often in the local control and decentralized management of various departments. In the application of electric power large data analysis technology, according to the established business goals, data mining technology is needed to explore and analyze a large amount of diverse energy data and municipal, transportation, weather and economy and other industries, to reveal the hidden, the unknown, or verify known regularity, and further its model. The first step to achieve the above goals is to integrate the distributed heterogeneous data through the global control and high-level collaborative management of the entire system, and build a distributed information database which has the independent processing of each department and suitable for the global scope, and provide a single general view of all data stored in the local system, construct a global model, and make the data distribution and the heterogeneity transparent to users. This makes users and data analysts in the access to heterogeneous data as simple as access to similar mode data.

3.The storage of heterogeneous data based on NoSQL database

Table 1 shows the multi energy heterogeneous data have the following characteristics: 1) General structured data: to describe entities by the table, to store the attribute data in a field and to describe entity relationship by foreign key associations, to operate by a relational algebra, logic model according with relational database storage; Relational database is suitable for store it; 2) Image data: the amount of data source is enormous, but the small size relatively, so the database can be chosen as storage mode; 3) Video data: the number is not much, but the file size is large, so file storage mode is the most commonly used; 4) Space vector data: in order to information storage of GIS and other systems, the data volume of space vector data is large, and the structure is more complex, which requires higher information retrieval and access technology, so the relational database management system can be adopted; 5) 3D model data: the hierarchical structure of data is more complex, but the number of the model is not much, so specialized file storage mode can be used for it.

Considering the diversity of multiple heterogeneous data storage structure, the NoSQL database based on document storage can store and query different types of data by relatively uniform structure. NoSQL is a new database technology, which has a significant difference with relational database in the data model, reliability, consistency, and many other core mechanism of database. NOSQL data storage has been out of the traditional relational database based on the rigid data model of ranks, using loose data structure, which has the performance advantages that the traditional database cannot match in the need for high concurrent processing and large data access. The NoSQL database which is stored in the document type uses generally the data structure of JSON format, the document is used to store the data, and establish the mapping from the document to the data records to achieve the rapid query for data. Fig. 2 shows an example that NoSQL database based on the document type for storage holds multiple heterogeneous data. The data record holds the offset value and the length in user file in the corresponding each user's data, and by which the NoSQL database management system can quickly retrieve the file record that mapped by the user's query record, so that the heterogeneous data is stored in the same database system.
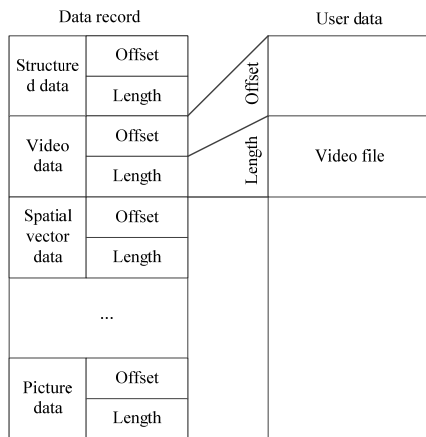
Fig. 2 The NoSQL database holds the heterogeneous data
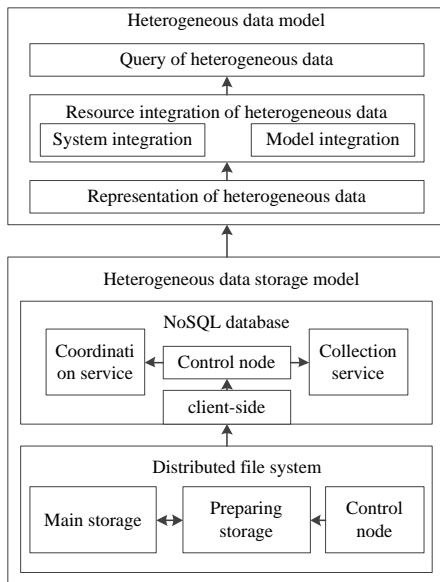
4.The global integration model of heterogeneous data

Fig. 3 The overall integration model of heterogeneous data

The overall integration model of heterogeneous data is shown in Fig. 3. The integration of heterogeneous data is divided logically into two parts, which are the storage model of distributed heterogeneous data and data model of heterogeneous data.

When multiple energy heterogeneous data are analyzed integrally, the data from each business system is required to store in the form of files in the distributed file system HDFS of the heterogeneous data storage model. When the unstructured data are added to the HDFS, the metadata information and the file storage size of the user is obtained by the application server as the offset value of the newly added file. File metadata information, offset value, and file size that is added to the NoSQL as a record is written into the data record of NoSQL database based on the data stored in the document type, and then apply the server to write a new file to the corresponding file of HDFS in an additional way. When the program reads the file, first a request is sent to the NoSQL

database system, and the relevant metadata information of the file corresponding to records is acquired. The input stream of the user file in the HDFS is acquired according to the offset value and file size so as to read out the file.

Because file structure stored in HDFS can vary, so the representation of heterogeneous in heterogeneous data model can be chosen flexibly according to the demand of application, such as virtual reality model suitable for 3D data model, point, line and plane model suitable for spatial vector data, the index description model suitable for the pictures and video data and so on. In the process of heterogeneous data resource integration, in the face of the integration problem of the system caused by the difference in the application service system or the operating system and the pattern fusion problem due to the inconsistent data patterns, the special transaction interface can be used to shield the access complexity on account of heterogeneous data. In the end user can develop a unified interface that access data for users on the basis of the NoSQL query language, which is convenient for the user to query the multi heterogeneous data in the same way as the centralized database query.

5.Conclusion

Based on the data of Tianjin eco-city, this paper introduces the typical characteristics of multi energy data, analyzes the data storage, including file storage and database storage, and proposes the use of NoSQL database technology and distributed file system to save heterogeneous data in the unified data record. Finally, it gives the overall integration model of heterogeneous data, which can be used as the basis for the design of multi energy heterogeneous data system.

Acknowledgements

Reference

[1] Xiaofan Shang: Integration of Information Resources and Information Island of Electronic Government. Modern Information No. 6 (2008), 18-20.

[2] Yinbiao Shu, Jianbin Fan: Progress in Smart City Standardization. Power System Technology Vol. 38, No. 10 (2014), 2617-2623.

[3] Xuehao Hu: Smart Grid—A Development Trend of Future Power Grid. Power System Technology Vol. 33, No. 14 (2009), 1-5.

[4] Qian Yi: *Research on the Operation Mode of Tianjin Eco-city*. Tianjin University of Technology, Tianjin, China, 2009.

[5] Kai Xie, Mingzhi Liu, Jiancheng Yu: Summary on smart grid integrated demonstration project of Sino-Singapore Tianjin Eco-city. Journal of electric power science and technology Vol. 26, No. 1 (2011), 43-47.

[6] Xu Shan: *The Research on the Method of Heterogeneous Big Data storage*. Beijing Jiaotong University, Beijing, China, 2014.

[7] Jinbo Du, Junjun Yang: The Unstructured Storage Service Based on NoSQL and Cloud Storage. J. Wuhan Univ. (Nat. Sci. Ed.) Vol. 58, No. S1 (2012), 99-101.

[8] Qi Yu, Jie Ling: Research of cloud storage security technology based on HDFS. Computer engineering and design Vol.34, No. 8 (2013), 2700-2705.