Journal of Computing and Electronic Information Management

ISSN: 2413-1660

Research on Power Big Data Visualization Based on Hadoop

Pan Gao

Key Laboratory of Industrial Internet of Things & Networked Control, Ministry of Education, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

453572890@qq.com

Abstract: With the continuous development of smart grid construction, the size of modern power systems is becoming increasingly large. How to extract value information that can meet people's needs from a large number of power business data has important practical significance for the safe and stable operation of power systems. After the development of computer technology and the promotion and application of power systems, it is possible to display the relevant data of the power system with graphic image resources capable of accommodating large amounts of information, and it is possible for the auxiliary system operators to grasp the operational status of the system in a timely manner. Therefore, based on the Hadoop platform, this paper mines and analyzes the power big data and visualizes the analysis results.

Keywords: Hadoop; big data; MapReduce.

1. Introduction

With the continuous development of smart grid construction, the scale of modern power system is becoming larger and larger. The power data acquired by each acquisition system has the characteristics of large amount of data, diverse types and low degree of resource integration. How to extract value information from a large amount of power business data to meet people's needs so that system operators can respond promptly to various complicated situations is of great practical significance for the safe and stable operation of power system. According to GTM Research, a leading power industry analysis and consulting company, the world's power big data management system market will reach 3.8 billion USD by 2020, and the power big data collection, management, analysis and service industry will hit an unprecedented level Development Opportunities.

2. Hadoop Platform Introduction

2.1 MapReduce architecture

MapReduce is a software framework that can write applications. It is suitable for large-scale data processing. Large data processing often uses parallel computing. MapReduce is a programming model that simplifies parallel computing. The job usually divides the input data set into several independent data blocks. Map tasks are processed in a completely parallel manner. The framework first sorts the output of the Map and then inputs the results to the Reduce task. Usually the input and output of the job are stored in the file system, the entire framework is responsible for the scheduling and monitoring of the task, and re-execution of the task has been closed. In general, the MapReduce framework and distributed file system run on the same set of nodes. This mode of operation is conducive to the MapReduce framework to efficiently schedule tasks on the nodes, thereby increasing the network bandwidth utilization of the entire system cluster. MapReduce framework diagram as shown in Figure 1.

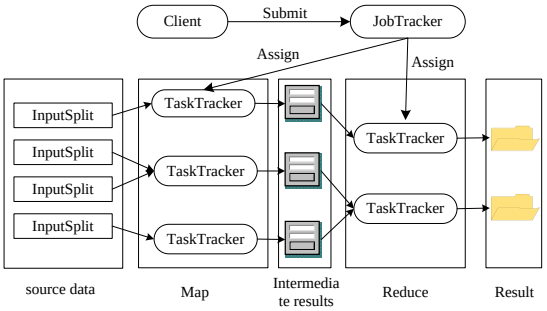

Figure 1 MapReduce architecture

2.2 HDFS

HDFS (Hadoop Distributed File System) is a distributed storage system in the Hadoop platform. As one of the core components of the system cluster, the system stores data of all storage nodes and has high fault-tolerance and high-throughput data access. Therefore, the distributed file system is suitable for large-scale data sets and provides a high-fault-tolerant and high-throughput large-data storage solution. HDFS framework diagram as shown in Figure 2.

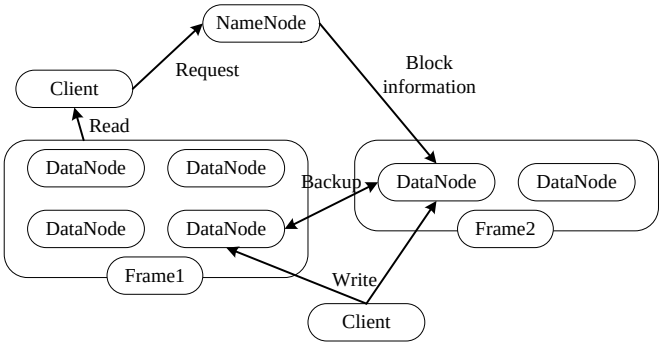

Figure 2 HDFS architecture

3. Data preprocessing

Power big data comes from many different data collection points, including text data, multimedia data, time series data and other types of data, which contains a very small amount of abnormal data, that is, power big data has a large amount of data, a wide variety of data, and value. Low density and other characteristics. Abnormal data mainly manifests as missing data attribute values, redundant data values, and inconsistent data formats. Based on this, the power big data needs to be preprocessed, and only by ensuring high quality data input can high quality decision results be produced. The preprocessing process mainly includes data cleaning, data integration, data transformation, and data protocol.

4 Power Big Data Visualization

3.1 Visual Architecture Design

Combining the characteristics of power big data data, according to the needs of power big data visualization, the overall scheme of power big data visualization based on Hadoop platform was proposed. The overall scheme architecture is shown in Figure 3.

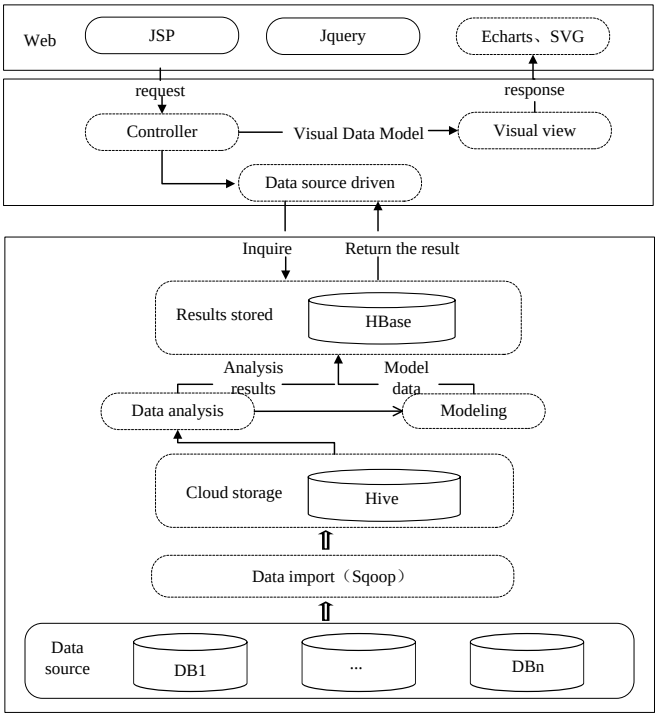

Figure 3 MapReduce architecture

3.2 Visual data model design

Data visualization has different manifestations. Different visualization requirements have different visual forms. Specific visualization needs to combine data features and visual mining types. The visual data model design is shown in Figure 4.

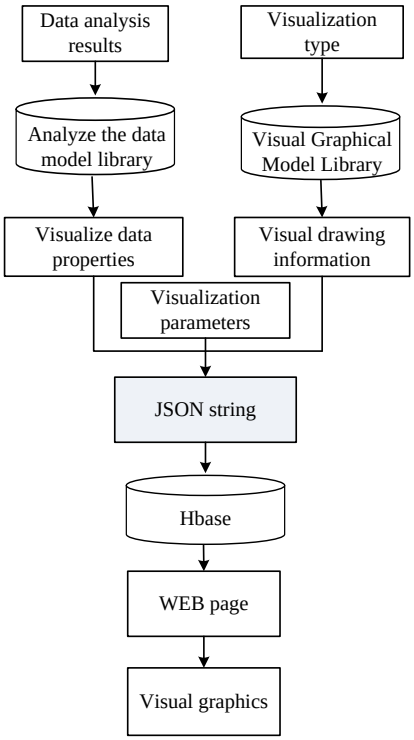

Figure 4 Visual data model design

3.3 Visual display

By interacting with the EasyUI control of the web interface, the required data rendering effect is achieved. EasyUI requests the data by sending an Ajax request, jQuery Ajax requests the visual data model from the server, and parses the model data, according to the parsing result, the attributes in the macotions object are performed. Update, load the map object parameters for the dom object, and call the map API to render the data. The data exchange diagram is shown in Figure 5.

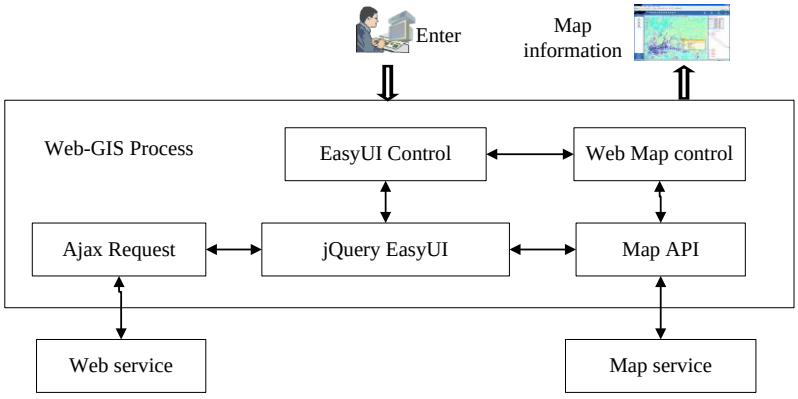

Figure 5 Data exchange diagram

References

[1] ZHANG Dongxia,MIAO Xin,LIU Liping,et al.Research on development strategy for smart grid big data[J].Proceedings of The Chinese Society for Electrical Engineering, 2015, 35(1): 2-12.

[2] WANG Xiangwei, SHI Yuliang, ZHANG Jianlin. Computation services and applications of

electricity big data based on Hadoop[J]. Power System Technology, 2015, 39(11): 3128-3133.
[3] SHE Xiangyang, ZHANG Ling. Parallel improved algorithm of FP-Growth association rules based on Hadoop [J]. Application Research of Computers, 2018, 35(1): 1-6.
[4] ZHANG Zhonglin, TIAN Miaofeng, LIU Zongcheng. Parallel hierarchical association rule mining in big data environment[J].Computer Science, 2016, 43(1): 286-289.
[5] GUO Xiaoli, QU Chaoyang, LI Xiaodong, et al. Visual data mining model in power grid based on SOM clustering[J]. Information Science, 2015, 30(2): 206-209+225.
[6] WANG Ling, LI Shulin, WU Lulu. Categorization and regression algorithm based on the quantitative association rule tree[J]. Chinse Journal of Engineering, 2016, 38(06): 886-892.
[7] LIAO JINGGUI, ZHAO YUELONG, LONG SAIQIN. MRPrePost—A parallel algorithm adapted for mining big data[C]// IEEE Workshop on Electronics, Computer and Applications. Ottawa: IEEE, 2014: 564-568.
[8] HUANG Manyun,WEI Zhinong,SUN Guoqiang,et al. A Novel situation awareness aproach based on historical data-mining model in distribution networks [J]. Power System Technology, 2017, 41(04): 1139-1145.
[9] XU Linsheng.Design and realization of electric power dispatching management system based on data mining[D]. University of Electronic Science and Technology, 2013.
[10] WU Jianjian.The research and realization of key technology of data mining visualization platform in power system. [D]. Zhengzhou University, 2016.