A Hierarchical Visualization Analysis Model of Power Big Data

Yongjie Li *, Zheng Wang and Yang Hao
Information and Communication Engineering Center, Information & Telecommunication Company of state grid Henan electric power corporation, Zhengzhou 450052, China

*Corresponding author e-mail: 154094079@qq.com

Abstract. Based on the conception of integrating VR scene and power big data analysis, a hierarchical visualization analysis model of power big data is proposed, in which levels are designed, targeting at different abstract modules like transaction, engine, computation, control and store. The regularly departed modules of power data storing, data mining and analysis, data visualization are integrated into one platform by this model. It provides a visual analysis solution for the power big data.

1. Introduction
Power big data is a new interdisciplinary emerging with the development of smart grid and the technological development of big data and the specific application of big data technology in the modern smart grid [1-4]. To date, the development of modern smart grid has a lot of power system data which are sourced from each production and operation process of power system. On the one hand, it poses a great challenge to the storage of existing power system data and the computation and analysis of data, and on the other hand, it offers valuable data resources for the value mining of power big data.

Visualization method has been used to display and analyze data [5-7]. Studies have shown that human get the information from the outside world more than 80% come from the visual system. When the data display in the form of intuitive graphical, the user can immediately know the information hide in graphics data, and quickly turn it into knowledge. In the era of big data, data is surge and data dimension increases. The current application of data mining in power grid enterprise research and development is rapid, but how to display the analysis results of data mining in a way that is easy to understand and how to use the result to guide the grid enterprise work are challenging problems.

2. Visual analysis task processing flow
For conventional power system visualization tasks, the whole process of data generation and visualization can be divided into the following 3 steps:

1) Collect data and serialize it into the file system (or further persistence to the database);
2) After preprocessing the original data, the intelligent algorithm is used to analyze the data;
3) The output data is imported into the visualization module and analyzed according to the visualization results.

It can be seen that, after each step, the off-line manual intervention can be carried out for the next step, not only the degree of automation is low, but each process is independent of each other. Under the
framework of big data, steps 1) and 2) can be processed continuously, but it still needs to be imported into the visualization system, and there is a lack of reasonable visualization means. The 3D scene is regarded as the core of the visualization system, and the visualization system is integrated into the big data frame, and then the data visualization is carried out on the basis of the visualization of the 3D scene. The power of big data has spatial characteristics, and it can set up the 3D spatial information in the data field. The data through specific interface and 3D map directly, thus the data and spatial information associated in the visualization level.

![Figure 1. Conventional visualization analysis process](image1.png)

![Figure 2. Visual analysis process of large power data system](image2.png)

From the above figure, combined with three-dimensional scene, and integrated into the framework of large data, visual analysis of the task of the following processes are as follows:
1) Collect data, serialized into the big data file system HDFS, and at the same time to the database HBase;
2) decide what kind of scheme to mine and analyze the data, and use the corresponding intelligent algorithm;
3) The output result set is mapped to the visualization module, and the visualization engine integrates the result set with the scene

Due to the large data environment, data, business, and the scene is more complex, the need for the entire task processing process to achieve reasonable organization to meet the needs of practical applications. The hierarchical model for power system design process large data visualization tasks.
Considering the Hadoop ecosystem has become the industry standard in the field of big data, the proposed model design based on the Hadoop ecosystem. Hadoop system is widely used in the field of large data processing and easy to use. It has been widely concerned and studied by the academia since it was launched, Hadoop is the most successful and widely accepted mainstream technology and system platform for large data processing, which provides a full range of functional modules for distributed cluster computing. Since the Hadoop platform has evolved into a complete ecological cluster system. Hadoop platform running on a server or by ordinary commercial machines on the cheap, cheap, convenient, big data expansion solutions.

3. Hierarchical visualization framework for big data analysis

In the visualization of large data analysis of actual application, the user needs an integrated platform. Because of the business with many function modules of the platform, the application platform for the abstract model, discuss the organizational structure and hierarchy model.

First of all, from the perspective of functional modules, the modules should be discussed:

1) Business module: provide high-level abstract interface to achieve the business needs of the user layer
2) Visualization engine: as the core subsystem of the model, it should realize the integration of data and scene, and realize the rapid rendering of large scale 3D scene to meet the needs of practical application;
3) Calculation module: used to implement a variety of intelligent algorithms to complete the work of data mining;
4) The control module is used to complete the task scheduling of large data, to achieve a reasonable load balance control;
5) Storage module: used to store data, the need to achieve large data file systems and database systems;

The possible business is combined with the above modules to model the layers, named as the interface layer, the engine layer, the computing layer, the control layer and the persistence layer, as shown in the following figure

![Diagram of Visualization Analysis Platform](image)

**Figure 3.** Visualization analysis platform function module of Power big data

After the division level, for each layer from the bottom of the specific design

1) Persistence layer
Hadoop file system HDFS and database system HBase is used to store all types of data, including scene data, numerical data and log data generated in actual operation.

2) Control layer
Hadoop task scheduling module ZooKeeper and chip level parallel technology MPI are used to control different types of computing tasks. ZooKeeper is used to control the data intensive computing mode STORM, MPI is used to control the computation intensive computing model CUDA. These control modules perform task scheduling, low level load balancing and simple fault tolerance.

3) Computing layer
The classification of computing tasks is not divided into graphic computing and data calculation, but is divided into data intensive and compute intensive.

The purpose of this method is to improve the computational efficiency as much as possible. For the two types of computing tasks, the parallel computing model is used. For the data to be calculated, all the data intensive are assigned to the STORM module execution; all the computing intensive computing are assigned to the CUDA module

4) Engine layer
The engine is the core subsystem in the model, the module implements a fast rendering engine, a large-scale 3D scene for large data environment design several optimization algorithms and strategies, to ensure the efficiency of real-time rendering the scene can meet the needs of practical application. The problem of real-time rendering of large-scale scene has been the focus of research, this paper propose two approaches: one is used to accelerate the rendering method based on octree culling, and rendering method for multi resolution weight function based on LOD. The two methods will be introduced in detail in the following sections of the algorithm and the implementation process, and through experiments to verify its efficiency

5) Interface layer
As a high-level abstraction, the interface layer needs to define a series of interfaces for the user's direct operation, these interfaces should include the operation of the scene and the operation of the data, as well as other operations, such as scene import, data import, data analysis, and log export and so on. These interfaces should also be reserved for fields that indicate the type of computation the requested task is data intensive or computationally intensive. For tasks that do not require parallel computing, the field should be set to NULL to avoid unnecessary task scheduling and inter node communication time loss.

In summary, the system architecture diagram of the model can be obtained:

| Interface Layer | Functional Modules and Interfaces |
|----------------|-----------------------------------|
| Engine Layer   | Visibility Analysis Engine        |
| Computing Layer| STORM                             |
| Control Layer  | ZOOKEPER                          |
| Persistence Layer | HDFS             |

**Figure 4.** Large scale data visualization analysis model
4. Summary
A hierarchical visualization analysis model of power big data is proposed, in which levels are designed, targeting at different abstract modules like transaction, engine, computation, control and store. The regularly departed modules of power data storing, data mining and analysis, data visualization are integrated into one platform by this model. It provides a visual analysis solution for the power big data.

References

[1] Zhan J, Huang J, Niu L, et al. Study of the key technologies of electric power big data and its application prospects in smart grid Power and Energy Engineering Conference (APPEEC), 2014 IEEE PES Asia-Pacific. IEEE, 2014:1-4.
[2] Peters D P C, Havstad K M, and Cushing J, et al. harnessing the power of big data: infusing the scientific method with machine learning to transform ecology. Ecosphere, 2016, 5(6):1-15.
[3] Yang Z, Zhang C, Hu M, et al. An Electric Power Big Data Deployment Solution for Distributed Memory Computing. 2015:155-161.
[4] Guo L, Vargo C. The Power of Message Networks: A Big-Data Analysis of the Network Agenda Setting Model and Issue Ownership. Mass Communication and Society, 2015, 18(5):557-576.
[5] Munzner T. Visualization Analysis and Design. Wiley Interdisciplinary Reviews Computational Statistics, 2015, 2(4):387–403.
[6] Munzner T. Visualization Analysis and Design: Principles, Techniques, and Practice// ASME 2011 Summer Bioengineering Conference. 2014:1119-1120.
[7] Xie P. Study of international anticancer research trends via co-word and document co-citation visualization analysis. Scientometrics, 2015, 105(1):611-622.