# Research of Smart Distribution Network Big Data Model

Guangyi LIU
China Electric Power Research Institute
liuguangyi@epri.sgcc.com.cn

Yang YU
Stanford University
yangyu1@stanford.edu

Feng GAO
Smart Grid Research Institute North America
feng.gao@ieee.org

Wendong ZHU
Smart Grid Research Institute North America
Wending.zhu@gmail.com

## *ABSTRACT*

*In this research, we tease out three rules of the big data for the electricity sector and induce a systematic criteria framework for building and evaluating data models for supporting the electricity-system big data. On the basis of the three rules, we induce a criteria system, which contains five types of criteria, to building a data model or evaluate the performance of an existing data model to facilitate the electricity-system big data. We then build a big-data model for Chinese electricity system according to the criteria system. In order to satisfy the requirements of the criteria, we choose CIM model to unify the protocol of data collection and communication. We also explore a data-classification policy to functionally structure the data and maintain their spatial and sequential relations. Finally, we suggest a framework to integrate the database into a platform by using multiple data-storage and data-communication technologies.*

## INTRODUCTION

Even though the current development of data-mining technologies provides opportunities for big-data analysis for the power system, the lack of a utility unified data model blocks the efficient data integration and module deployment [1]. This research therefore builds an experimental big-data model to explore the mechanism for framing the big data to support the real-time operation and adaptive innovation of the smart distribution network [3].

The big-data use for the smart distribution network faces to a sequence of challenges differing from its use in other system; therefore the experimental data model is designed aiming to explore the solutions for these challenges. One main challenge is that the data gathered from the smart distribution network has complicate logic structures, multiple sources, and large volumes [4]. Therefore, any single database technology is insufficient to support the big data analysis in the electricity system. In addition, because of the lack of understanding for use of the data, the data model must comprehensively and smartly keep all the information about facilities, market participants, and relations connecting them [3]. Furthermore, the data model must support the development of the intellectual and interactive visualized operation system [3].

## 2 CRITERIA SYSTEM FOR BUILDING AND EVALUATING AN ELECTRIC SYSTEM DATA MODEL

### 2.1Three rules of building a big data system for electric power system

In this section we tease out three rules for building a big data system for electric power system, which is supported by an electric system data model. Therefore, an effective electric system data model must be designed to comprehensively facilitate the functionalization of the three rules.

The three rules for building a big data system for electric power system are

1. Comprehensive: the system should appropriately store, process, and communicate all available data and information and integrate all available tools and interfaces for data analyses。
2. Adaptive and dynamic: the system should be adaptively organized, dynamic innovated, and function-based structured to support all kinds of dynamic data process and analyses.
3. Distributed and multi-layered: the system should be multi-layered; and each layer or knot need to be, to some extent, intelligent, such that the data and information can be effectively processed and analyzed in appropriate layer or knot on the system.

The three rules are different from the rules for a big-data system in other field because of the electricity systems' particular physical characteristics and operational requirements. We summarize these three rules by embedded the characteristic of a big data system into considerations of electric system operation.

In particularly, we would like to emphasize that the three rules listed above are for the building of a smart-grid system. A smart-grid system, which is a decentralized system equipped with intelligent processing and communicating facilities at multi layers in the system, is an integrated platform to support the utilization of integrated distributed renewable-energy sources, smart meters, and the technologies of demand-side response.

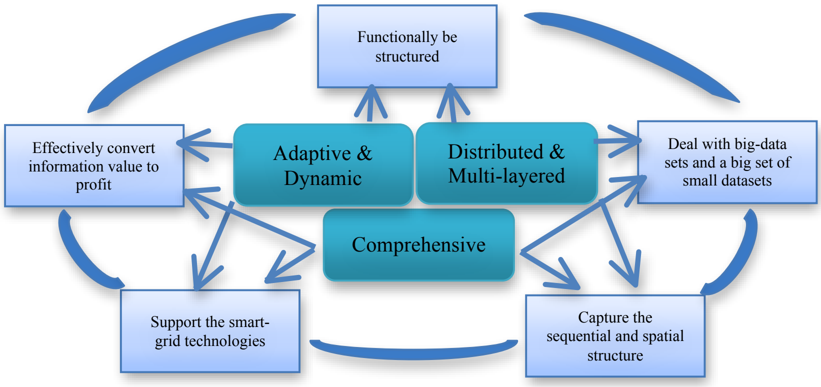


**Figure 1 Three Rules and five criteria**

## 2.2Requirements for an effective electricity-system data model

According to the three rules of building a big-data system for electric power system, we correspondingly induce a criteria system for an effective data model, which is used to collect sample, storage data, and support data analysis under a big-data environment in an electricity system. The criteria system contains five types of criteria. In figure 1, we describe the relation between the three rules and their corresponding criteria types.

According to the criteria system, we not only could evaluate the effectiveness of a data model, but also design a new framework.

### 2.1.1Capture the sequential and spatial structure

The electricity-system data has heterogeneous data forms, which are determined by the sampling frequency, and contains complicated relations, which reflect the hierarchical network structure of the electric system.

The electricity-system big data is highly time-scaled heterogeneous. The frequency for data sampling varies from one time per microsecond to one time per month or year. For example, the data captured by the Phasor Management Unit (PMU) is sampled per millisecond. In contrast, the data used for the resource-capacity planning only include the monthly or yearly information. The large variation of sampling sequence not only causes the big difference of information included in the data, but also requires different recording methods. For instance, the data from PMU is stream data. Therefore, the data-recording system must timely tease out the information from the data and determine how to deal with the data. In contrast, the data for resource-capacity planning does not need to be highly frequently analyzed but must include a broad set of information, such as demographic change and economical growth forecast.

In addition to the data-form heterogeneity caused by the time sequential, the other characteristic of the big data for electric system is that the data include information about complicated spatial relation. The data collected by smart meter is centralized into the data center at each node in the distributed networks and then submitted to the system-control center. All data not only include the situation at the position where the data is collected, but also contain the information about how demand, supply, and power flow perform with the constraints of the grid topology.

Therefore, we suggest the first requirement for an effective electric power system data model as follows.

**Type 1 criteria:** An effective big data model for electric power system should

- have an adaptive data-sampling protocol according to the data's time-sequential characteristic and
- contain the information about the spatial relation between different data sources in well-organized data storage framework.

### 2.1.2 Deal with big-data sets and a big set of small datasets

The electric power system data is included in different databases. A few of the sub systems have very huge database. Each of these databases contains big data sets, for example, the smart-meter data.

Many other sub systems have databases with limited size. However, the number of these small databases is large enough, such that the set of these databases can be considered as "big data".

Then, we have the following requirement

**Type 2 criteria**: An effective data model should have ability to

- parallel process multiple big-data datasets and
- effectively and timely connect multiple small datasets among a large number of datasets.

### 2.1.3 Support the smart-grid technologies

The big data for electric power system is the foundation of the smart-grid creation. A smart grid system is built upon the technologies of interactive devices, internet, data analytics, and distributed control. The data model must effectively connect with these technologies and support the target functions of operating a smart-grid system.

The basic functions of a smart-grid system include the energy management at the terminal devices, the adaptive response to the market signal from the agents on the distributed networks, and the distributed communication and control for the electricity dispatch and frequency regulation [3][6].

In order to support the energy management at the terminal devices and the adaptive response from the agents on the distributed network, the electricity-system data model need the ability to analyze high-frequency stream data and differentiate the information to be submitted from raw data timely treated.

In order to support the distributed process and control, the data model should be designed to associate with the optimal distributed signal-communication algorithm and distributed control system.

**Type3 criteria**: An effective data model for the big data in an electricity system should support the functions of

- converting terminal device to management platform,
- distributed processing and controlling, and
- adaptively distributed-side response.

### 2.2.4. Functionally be structured

In contrast with the current data model, the electricity-system big-data model need to be structured to support the function of system operation by integrating data and information from multi-sources.

The current data model is separated by collection systems because the marginal benefit of functionally integrating databases cannot balance its costs. Then, the same data can be collected by two systems and then stored in different databases. However, the value of functionally integrating databases in a big data system is sufficiently large. Furthermore, the collection-system-based data model significantly weakens the ability of using new information from big data to benefit the smart-grid building and operation.

The challenges of creating a function-based integrated data model include not only the effective storing data but also integrating data with different formations and types. For instance, the demand-response management needs stream data collecting from smart meter and also point-wise data from feeder monitors. In some cases, the data model needs to support the processing of effectively integrating the raw data with some processed information from data processors and computers on other layers of the system.

**Type 4 criteria**: An effective electricity big data model should be functionally structured such that

- Data is effectively collected and adaptively connected according to the possible functions.
- Data with different types and formations can be appropriately sorted for the utilization.

### 2.2.5. Effectively support the function of convert information value to profit

The big-data model for an electricity system must have ability to convert the information value to profit. The big-data technologies reveal a great amount of information, for instance individual consumer's hourly load data and high frequent PMU data. These pieces of data include valuable information for improving the whole-system efficiency. Therefore, the acquisition of these data and information provides the opportunity to the whole system to gain profit from reducing the costs caused by information incompleteness and asymmetry. However, because of the large size of the database and sparse information matrix, the data-model design highly affects the effectiveness of converting potential information value to profit.

The data model must appropriately support multi tools for profit mining. There is a sequence of tools of mining the value of the information contained in the electricity-system big data. For example, the information about consumers' behaviors collected from smarter meter must have appropriate tools to analyses. Simultaneously, a tool is needed to match the consumers with the distributed renewable-energy capacities. Therefore, the data model must support multi tools.

The data model also needs to support the new market designs and mechanisms that utilizing big data to improve the social efficiency. For instance, the big data made the trade inside a distributed system is available. However, a new market mechanism and a dispatch protocol need to be designed and implemented. The data model is a foundation of the new mechanisms and protocols built upon big data.

**Type 5 criteria**: An effective electricity big data model should

- Effectively support analyses and processing tools for converting potential value contained in the data into profit.

- Effectively support the new market designs and management protocols built upon the electricity-system big data

## 3. DATA MODEL FOR THE ELECTRICITY-SYSTEM BIG DATA

We in the last section systematically build a criteria framework for evaluating and building a data model for an electricity-system big-data model. With considering the length of the paper, we do not include the example of evaluating a current data model by our criteria.

We in the section provide an example to build a big data model for Chinese electricity system. We would like to emphasize that the data model need to particularly design for each individual electricity system because the operation protocols are different by electricity systems. However, our criteria framework can be generalized to any electricity systems.

In this section, we first select technologies used to frame the data model. Then, we build a policy to functionally classify and store data according to spatial and sequential relations of data. At last, we state how to create the integrated platform of the data model according to knowledge-management theory.

### 3.1 Technologies used for data-model building

From criteria type 1 to type 4, we argue that the core issue of the big-data model data is effective data structure and connection. Therefore, the unified data model integrates the data according to the relational model based on the Common Information Model (CIM) [4]. We choose the CIM model because

- The CIM model provides a common language for data collection from different devices located in different layers of the system.
- Furthermore, the CIM has a uniform standard and protocol for data communication and classifies and connects data.
- The CIM model is simultaneously able to functionally connect data and information and appropriately maintain the spatial and sequential relations in the real world [4].

### 3.2 Data classification and functional organization

In order to functionally structure the data, we summarize the data groups, which are in the Chinese electricity system and grouped according to the geographic collecting position, data formation, and colleting frequency. This information reflects the spatial and sequential relations of the data, as well as the characteristics inside a dataset. Then, we map the classes to their functions in the smart-grid system. According to the mapping, we reclassify the data. We in Table 1 show this classification method.

**Table 1Function-based structure for the electricity-system big data model**

<table>
<tr><th colspan="2">Sequential Relations (Dynamic Information)</th><th colspan="6">Spatial Relations（Consumer Management、GIS、Device Management）(Cross-section Information)</th></tr>
<tr><td>Sampling Frequency</td><td>Data Type</td><td>Terminal Devices</td><td>Distributed Feeders</td><td>Transformer</td><td>Local Dispatch</td><td>Transmission Dispatch</td><td>Inter-transmission management</td></tr>
<tr><td>Event Data（μs）</td><td>Wave</td><td rowspan="2">Relay protection, consumption analysis, smart inverter for distributed RE et al.</td><td rowspan="4">Feeder fault detection and self recover, feeder-level forecast for demand and distributed generation capacity, feeder-level consumption decompose and analysis</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Stream Data（ms）</td><td>PMU</td><td rowspan="4">Distributed automation, distributed device management, self recover, forecast and control, fault detection, device monitor and analysis</td><td rowspan="5">On-time processing and on-line detection, Integrated Distributed management system、demand analysis、assess management</td><td rowspan="3">Emergency response, on-line parameter detection, WANs control, energy management, pricing and energy efficiency</td><td rowspan="3">On-line parameter detection, WANs control, energy management, pricing and energy efficiency</td></tr>
<tr><td>Steam Data（s）</td><td>SCADA</td><td></td></tr>
<tr><td>Stream Data（minute）</td><td>Smart Meter</td><td>Consumption decompose, household energy management</td></tr>
<tr><td>Event and Image Data（Minute）</td><td>Device Monitor</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Time and Image Data (Hour)</td><td>Customer Service and Maintaining</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Batch Processing（Day/Month/Year）</td><td>Planning</td><td></td><td></td><td></td><td>Distributed network planning</td><td>Transmission network planning</td><td>Inter-transmission plnanning</td></tr>
</table>

Each cell of the Table 1 represents a data class. We reclassify data in this way to effectively reflect the information of data's functions, maintain the data's spatial and sequential relations, and minimize the efforts to record these pieces of information.

From the table, we show that our classification is not restricted by the sources of the data collection. Instead, the data, which is used for the same types of tasks, are grouped together. Furthermore, the data is organized in a hierarchic protocol. These hierarchic structures support the multi-layered adaptive intelligent structure of a smart grid system.

### 3. 3 Knowledge management and data storage

The unified data model is built upon a technology portfolio of state-of-the-art computing. In the experimental platform, a data-storage-and-processing system based on the Hadoop technology supports an integration platform and an exploring platform. The Hadoop system provides an environment for storing semi-structure and non-structure data, and exploring the optimal solution for the data framing [2].

The integration platform, which is built upon the Data Warehouse technology, is for the instant inquiry and the decision making for the data analysis and the module development. The exploring platform, which is based on the paralleling processing technology, is established for visualized analysis and data mining. The modules for data analysis, inquiry and decision-making are integrated into the two platforms [5].

The unified data model also provides a research environment to examine the optimal technology portfolio to achieve cost-efficiently and stably data management and analysis. A self-reinforce learning process on the basis of the RM theory makes the data mode adaptive and self-innovative.

A policy classifying the data will be established according to the uses and the structures of the data. According to their uses, the data will be classified to 'cold', 'warm', and 'hot' data. The data will simultaneously be grouped as 'structured', 'unstructured', and 'stream' data according to their logic structures. The storage protocol is determined by the classification of the data.

## 4. CONCLUSION

In this research, we tease out three rules of the big data for the electricity sector and induce a systematic criteria framework for building and evaluating data models for supporting the electricity-system big data. The three rules, which include comprehensive, adaptive and dynamic, and distributed and multilayered, are the basic principles for the smart-grid building and data-mode framing. On the basis of the three rules, we induce a criteria system, which contains five types of criteria, to building a data model or evaluate the performance of a exist data model to facilitate the electricity-system big data.

We then build a big-data model for Chinese electricity system according to the criteria system.In order to satisfy the requirements of the criteria, we choose CIM model to unify the protocol of data collection and communication. We also explore a data-classification policy to functionally structure the data and maintain their spatial and sequential relations. Finally, we suggest a framework to integrate the database into a platform by using multiple data-storage and data-communication technologies.

Reference

[1] K.Bakshi, 2012, "Considerations for big data: Architecture and approach." In Aerospace Conference, 2012 IEEE, pp. 1-7.

[2] P.Zikopoulos & C. Eaton, 2011, "Understanding big data: Analytics for enterprise class hadoop and streaming data", McGraw-Hill Osborne Media.

[3] D.Agrawal,S. Das, & A.El Abbadi,2011, "Big data and cloud computing: current state and future opportunities", Proceedings the 14th International Conference on Extending Database Technology ACM, 530-533.

[4] P.C.Chen, T. DOKIC, & M.KEZUNOVIC 2014, "The Use of Big Data for Outage Management in Distribution Systems" Proceedings International Conference on Electricity Distribution (CIRED) Workshop.

[5] NoSQL Architecture Web Site: http://www.sitepen.com/blog/2010/05/11/nosqlarchitecture/

[6] A. Jacobs, 2009 ,"The pathologies of big data." Communications of the ACM, Vol.52 no.8, 36-44.