Research on the Architecture of Power Grid Data Asset Management System Based on Data Asset Space Model

Wei Song¹, Mo Hai²*,¹, Jun Wang¹, Hanqing Hu³, Yuejing Zhang², Haifeng Li²

¹ Jibei Electric Power Company Limited Metering Centre, Beijing 102208, China
² School of Information, Central University of Finance and Economics, Beijing 100081, China
³ Beijing Information Science and Technology University, Beijing 100192, China

* Corresponding author: haimo_hm@163.com

Abstract. To solve the data management problems of the operation monitoring center of power grid enterprises: the lack of a unified management standard, information island and privacy protection, we propose a data asset space model of power grid-DAEVP to organize the power grid data asset more efficiently. The model can uniformly represent the heterogeneous data asset distributed in different business departments of the power grid enterprises, thus make the operation monitoring center own the control right as well as protect the privacy of the power grid data asset; At the same time, a power grid data asset algebra is proposed to solve the operation problem of the power grid data asset. A set of power grid data asset operators defined by the power grid data asset algebra can represent the behavioral semantics of the power grid data asset; in the meanwhile, the orthogonality and closure properties of the data asset operators are proved.

1. Introduction
With the development of the information system of power grid enterprises, the amount of data is increasing rapidly. The data generated by grid enterprises can be divided into four categories: the operation data, the equipment data, the customer service data and the internal operation data. The operation data refers to the voltage, current, power, frequency, charge and phase data of each node in the grid; equipment data refers to the status data and ledger data of each equipment in the grid; the customer service data refers to the basic data of customers, electricity behavior data and service data; the internal operation data refers to the data of human, finance, material, security and technology management. At present, the data stored in the grid enterprises have reached a scale of PBs and are still in a high growth. How to use massive data to upgrade the management level of grid enterprises, make the data resources served as a strategic asset of grid enterprises and enhance the application level as well as commercial value of the data, are challenges faced by grid enterprises.

The business departments of the grid enterprise have generated a large number of data in their business activities, but there exist three problems in the management of the data asset: the lack of unified management specifications, information island and privacy protection, which are described in the following:

- The data generated and managed by each business department of the grid enterprise are distributed in their own servers and the unconnected information systems of different departments, and each department provides its own specific data management functions. However, the entire grid enterprise lacks unified data management specifications, which leads to
the inconvenience of accessing and managing data of all business departments for the operation monitoring center;

- The data managed by each business department of the grid enterprise are bound by various departments. Each department actually controls the management of its own data. The data managed by each department lack of interconnection, and it is difficult for the operation monitoring center to retrieve the original data of each department;
- Each business department of the grid enterprise will mine valuable information after collecting the data they manage, and even provide the information it mines to a third party, thus it will threaten data privacy and cause data leakage.

The key reason of the above problems is: the organization and use of data are considered from the perspective of each business department of the grid enterprise, while ignoring an important subject-the operation monitoring center. The data management issue doesn’t be considered from the perspective of the operational monitoring center. Current research on the architecture of power grid data asset management system\[1-10\] does not solve the organization and operation issue of power grid data assets. To solve these two problems, the architecture of power grid data asset system based on asset space model is proposed.

We study the problem of power grid data asset management from the following two issues:

- The organization issue of the power grid data asset: how to organize the grid data asset to uniformly represent the heterogeneous data asset distributed in different business departments of the power grid enterprise, make the operation monitoring center own a control power of the grid data asset and protect the privacy of data asset;
- The operation issue of the power grid data asset: how to define a set of asset operators to represent the behavior semantic of grid data asset, and make the operation of the power grid data asset be finished by a simple combination of asset operators.

2. Power Grid Data Asset Space Model

2.1 Terminology

**Definition 1: Physical Data Resource**
A physical data resource is a kind of abstraction and encapsulation of the data stored in the database or file system of each business department of the power grid enterprise, it can be a record, a relational table or an excel file, etc. Physical data resources can be represented by a triple \(<\text{identifier}, \text{metadata}, \text{content}>\). Among them, the identifier uniquely identifies a physical data resource, the metadata is the attribute of the physical data resources, which includes: categories, tags, titles, creation time, modification time, permissions, owner and the content is the physical content of a physical data resource.

**Definition 2: Physical Data Asset**
A physical data asset is a collection of limited physical data resources, which can be represented as:

\[
\text{Physical data asset} = \{\text{physical data resource} \mid \text{identifier} \in \text{database or file system} \} \text{ and } |\text{physical data asset}| < \infty
\]

**Virtual Data Asset**
A virtual data asset is objects that are abstracted from distributed heterogeneous physical data assets and has the same format and uniform asset operations. A virtual data asset is a collection of limited physical data resources that can be represented as:

\[
\text{Virtual data asset} = \{\text{physical data resource} \mid \text{identifier} \in \text{database or file system} \} \text{ and } |\text{virtual data asset}| < \infty
\]

**Definition 4: Effective Data Asset**
An effective data asset is a view of a virtual data asset, which represents the result obtained from the operation on one or more virtual data assets obtained by asset operators. An effective data asset is a collection of limited physical data resources, which can be represented as:

\[
\text{Effective data asset} = \{\text{physical data resource} \mid \text{identifier} \in \text{database or file system} \} \text{ and } |\text{effective data asset}| < \infty
\]

**Physical Data Asset Space**
The physical data asset space \(P\) is defined as a hyperspace which is composed of all physical data assets in the physical layer, that is, the collection of all physical data assets,
and it can be represented as: \( P = \bigcup_{i=1}^{n} PA_i \).

**Definition 6: Virtual Data Asset Space** The virtual data asset space \( V \) is defined as a hyperspace which is composed of all virtual data assets in the virtual layer, that is, the collection of all virtual data assets, and it can be represented as: \( V = \bigcup_{i=1}^{n} VA_i \).

**Definition 7: Effective Data Asset Space** The effective data asset space \( E \) is defined as a hyperspace which is composed of all effective data assets in the effective layer, that is, the collection of all effective data assets, and it can be represented as: \( E = \bigcup_{i=1}^{n} EA_i \).

Physical data assets are located at the bottom level, which represent data assets distributed in each business department of the power grid enterprise, and the data model of physical data assets of different business department may be different; virtual data assets are located at the middle layer, which is a kind of general data model independent of the business departments of the power grid enterprise after the abstraction of physical data assets; effective data assets are located at the top level, which is based on the results obtained by the operations on the virtual data assets by asset operators. Actually an effective data asset does not store physical data assets, which is determined by an asset operation expression that defines the effective data asset. The definition of an effective data asset make the operation results of the virtual data assets be stored in a form of an asset operation expression. In order to get the contents of the effective data asset, the asset operation expression which defines the effective asset will be performed only when the valid data assets are accessed. Effective data assets, like virtual data assets, can participate in the operation of asset algebra. The definition of an effective data asset can contain the application requirements, thus make the effective data asset have more advanced semantics than the virtual data asset. For example, the operation monitoring center has two virtual data assets, one is the electricity information table of important customers of the marketing business system, and the other is the user electricity information table of the electricity collection system. The operation monitoring center can define an effective asset by the union operation on the two virtual assets, and only by one query on the effective asset all data of the two data tables can be seen.

2.2 Power Grid Data Asset Space Model

We propose a data asset space model of power grid-DAEVPM model to organize the power grid data asset more efficiently, which is composed of three layers: physical space layer, virtual space layer and effective space layer. The physical space layer can only provide a single system image of physical data resources of some business department of power grid enterprises, while the virtual space layer and the effective space layer can provide a single system image across the business departments of power grid enterprises. The mapping of virtual space layer as well as physical space layer doesn't distinguish the formats and sources of physical data assets, which make the distributed and heterogeneous physical data assets in physical layer be mapped to virtual data assets with a consistent format in the virtual layer, thus achieve uniformity. The effective space layer abstracts the concept of effective assets on the basis of the virtual layer, so that the operation monitoring center can manipulate the data assets at a higher level of abstraction to make the operation more convenient and easy to use. The effective layer abstracts the concept of effective assets on the basis of the virtual layer, so that the operation monitoring center can manipulate the data assets at a higher level of abstraction to make the operation more convenient and easy to use. As shown in Fig.1, the physical data asset \( PA \) can only provide a single system image for the physical data resources of the marketing department, the virtual data asset \( VA \) can provide a single system image for the heterogeneous physical data resources of the market department and the shipment inspection department, and the effective data asset \( EA \) provides a single system image for the heterogeneous physical data resources of the marketing department, the development department, the shipment inspection department, and the dispatching center by the virtual assets \( VA \) and \( VA_2 \). The virtual data asset \( VA \) is able to access the heterogeneous physical data resources of both the marketing and shipment inspection departments by physical data assets \( PA \) and \( PA_2 \), while the effective data assets \( EA \) can access the heterogeneous physical data resources of the marketing department, the
development department, the shipment inspection department and the dispatching center by virtual data assets - $V_1$ and $V_2$. It can be found from Fig. 1 that the relationship between an effective data asset and a virtual data asset is a many-to-many relationship, that is, multiple effective data assets can access the same virtual data asset at the same time, while an effective data asset can also use multiple virtual data assets simultaneously; at the same time, the relationship between the virtual data assets and physical assets is also a many-to-many relationship, i.e., multiple virtual data assets can simultaneously access the same physical data asset, and a virtual data asset can also use multiple physical data assets.

3. Power Grid Data Asset Algebra

To solve the issue of the operation of the grid data asset, we propose a kind of grid data asset algebra. According to the application requirements, the grid data asset algebra summarizes the common behavior of grid data assets and abstracts the asset operators: union, intersection, difference, selection, matching and projection. The two important properties of the asset operators: closure and orthogonality are proved. The closure property means that the input and output of each asset operator are virtual data assets or effective data assets, which makes the operation on the virtual data assets or effective data assets use the simple combination of one or more operators to complete. The orthogonality property makes each asset operator independent.

![Fig. 1. Power Grid Data Asset Space Model](image)

3.1 Union

**Definition 8: Union Operator**

Supposed there are two virtual data assets or two effective data assets, the union operation of these two objects can be defined as:

$$A_1 \cup A_2 = \{a | a \in A_1 \lor a \in A_2\}$$

In this formula, “$\cup$” is the union operator, $a$ is a physical data resource, and the result $A_1 \cup A_2$ is a new virtual data asset or effective data asset which is a collection of physical data resources belonging...
to $A_1$ or belonging to $A_2$. The union operator can be used to add new physical data resources to a virtual data asset or an effective data asset. For example, if we add two new physical data resources: $a$ and $b$ to a virtual data asset or an effective data asset $A_1$, it can be regarded as to perform the union operation on one asset $A_1$ and the other virtual data asset or effective data asset $A_2$, and $A_2$ contains physical data resources $a$ and $b$.

3.2 Intersection

**Definition 9: Intersection Operator** Suppose there are two virtual data assets or two effective data assets, the intersection operation of these two objects can be defined as:

$$A_1 \cap A_2 = \{a | a \in A_1 \land a \in A_2\}$$

In this formula, “$\cap$” is the intersection operator, $a$ is a physical data resource, and the result $A_1 \cap A_2$ is a new virtual data asset or an effective data asset which is a collection of physical data resources belonging to $A_1$ and belonging to $A_2$.

3.3 Difference

**Definition 10: Difference Operator** Suppose there are two virtual data assets or two effective data assets, the difference operation of these two objects can be defined as:

$$A_1 - A_2 = \{a | a \in A_1 \land a \notin A_2\}$$

In this formula, “$-$” is the difference operator, $a$ is a physical data resource, and the result $A_1 - A_2$ is a new virtual data asset or an effective data asset which is a collection of physical data resources belonging to $A_1$ and not belonging to $A_2$. The difference operator is used to delete a specified physical data asset from a virtual data asset or an effective data asset. For example: to delete two physical data resources $a$ and $b$ from a virtual asset or an effective data asset $A_1$, can be regarded as the difference operation of one asset $A_1$ and the other virtual data asset or effective data asset $A_2$, which contains physical data resources $a$ and $b$.

3.4 Selection

**Definition 11: Selection Operator** Suppose there is a virtual data asset or effective data asset $A$, its selection operation can be defined as:

$$\sigma_{P(a)}(A) = \{a | a \in A \land P(a)\}$$

In this formula, “$\sigma$” is the selection operator, $a$ is a physical data resource, $P(a)$ is the selection condition, the result $\sigma_{P(a)}(A)$ is a new asset, which is a collection of physical resources, the resources belong to the asset $A$ and make the selection criteria $P(a)$ true. $P$ is a logical expression, whose value is a logical value “true” or “false”. The basic form of $P$ is $X_1 \Delta Y_1$, and among them, $\Delta \in \{<,\leq,\geq,=,\neq\}$, $X_1$ is the name of an attribute, $Y_1$ is a constant. $P(a)$ is a compound logic expression consisting of basic logical expressions by the logical operator “$\land$” and “$\lor$” after a finite number of operations. For example: $P(a) = a.metatag = electricity\ table \land a.meta.create\_time > 2017.2.1$.

3.5 Matching

**Definition 12: Matching Operator** Suppose there is a virtual data asset or effective data asset $A$, its matching operation can be defined as:

$$\theta_{C(a,k)}(A) = \{a | a \in A \land C(a,k)\}$$
In this formula, “θ” is the matching operator, a is a physical data resource, \( C(a,k) \) is a matching criteria, and the result \( \theta_{C(a,k)}(A) \) is a new asset, which is a collection of physical resources, the resources belong to the asset A and make the matching criteria \( C(a,k) \) true.

### 3.6 Projection

**Definition 12: Projection Operator** Suppose there is a virtual data asset or effective data asset A, its projection operation can be defined as:

\[
\pi_{\tau_1,\tau_2,\ldots,\tau_n}(A) = \{ a_{\tau_1,\tau_2,\ldots,\tau_n} \mid a \in A \}
\]

In this formula, “\( \pi \)” is the matching operator, \( a \) is the selected physical data resource, \( \tau_1,\tau_2,\ldots,\tau_n \) is the metadata of \( a \), \( a_{\tau_1,\tau_2,\ldots,\tau_n} \) is the new view of the selected physical data resource \( a \) whose metadata \( \tau_1,\tau_2,\ldots,\tau_n \) are extracted.

### 3.7 Properties of the Asset Operators

**Theorem 1:** Suppose \( * \) is an arbitrary asset operator of grid data asset algebra system \( < A : \cup \cap \neg, \sigma, \theta, \pi > \), the asset operator \( * \) is closed for A.

**Proof:** According to the number of objects, the asset operators are divided into two classes: unary asset operator and binary asset operator. Given a non-empty grid data asset collection A and one unary asset operator \( * \in \{ \sigma, \theta, \pi \} \), \( *: A \rightarrow A \) exists. According to the definition of unary operation closure, any unary asset operator is closed for A. Given a non-empty asset collection A and one binary asset operator \( * \in \{ \cup, \cap, \neg \} \), \( *: A \times A \rightarrow A \) exists. According to the definition of binary operation closure, any binary asset operator is closed for A.

Therefore, any asset operator is closed for A. The closure of the asset operator guarantees that the input and output of the asset operator are all assets, which makes the system concept less and the processing mechanism unified. This brings an important feature for system design: simplicity.

**Theorem 2:** Suppose \( * \) and \( \Theta \) are two arbitrary asset operators \( < A : \cup \cap \neg, \sigma, \theta, \pi > \), the asset operators \( * \) and \( \Theta \) are orthogonal.

**Proof:** The union, intersection, difference, selection and projection operation have been proved to be orthogonal in the relational algebra, while the matching operation performs searching based on the content keywords, which cannot be represented by the combination of the five operators: union, intersection, difference, selection and projection.

Therefore, none of the six operators can be replaced by the combinations of other operators. The orthogonality of asset operators eliminates the mutual influence among the operators and makes operators independent on each other. The change on any operator will not influence other operators, which greatly improves the reusability and scalability of the operators.

### 4. Conclusions

In this paper, we propose a grid data asset space model-DAEVP model to solve the organizational problems of grid data assets. The DAEVP model divides the organization of grid data assets (DA) into three independent layers: physical layer(P), virtual layer(V) and effective layer(E). The key idea of the DAEVP model is to distinguish two concepts: physical data assets and virtual assets. The core principle of the DAEVP model is: do not operate on the physical data assets of the physical space layer. The physical data assets are mapped into virtual data assets of virtual space layer, and are operated indirectly by operating on virtual data assets of the virtual space layer. In virtual space layer the organization and management of grid data assets is independent of each department of the power grid
enterprise, so as to make the operation monitoring center have a control of power grid data assets and a protection of the data privacy of power grid data assets. The asset algebra including six asset operators is constructed to solve the operation problem of power grid data assets, and the orthogonality as well as closure property of the asset operators are proved.

Acknowledgment
The work is supported by State Grid Corporation of China Science and Technology Project, Young Teachers Development Foundation of Central University of Finance and Economics (Project Number: QJJ1634), Opening Foundation of Network and Data Security Key Laboratory of Sichuan Province(Project Number: NDSMS201604), Program for Innovation Research in Central University of Finance and Economics.

References
[1] Li Qian, Bai Xiao-Ming, Zhang Lin, et al. Data Asset Management and Data-Based Operation of Power Suppliers, East China Electric Power, 2014, 42(3): 487-490.
[2] Heng Xing-Chen, Zhang Shi-Jun, Chen Feng, et al. Grid Data Asset Management Platform Based on SOA, Computer System and Applications, 2016, 25(9): 67-72.
[3] Gu Bin, Peng Tao, Che Wei. Study on the whole life cycle management system of enterprise operation data assets, Enterprise reform and management, 2017, 3: 4-6.
[4] Bai Xiao-Jun, Liu Xin, Li Yu-Qin, et al. Research on the enterprise data asset management system based on the unified data center of electric power, Computer products and circulation, 2017, 7: 282-283.
[5] Chen Yi-Bo, Chen Qian, Gui Jian-Xin. Research on Power Grid Operation Analysis and Decision System Based on Big Data Technology, Power information and communication technology, 2015, 13(8): 128-131.
[6] Wei Xiao-Jing, Li Yi-Song, Zhou Chun-Lei, et al. Operation and monitoring practice of power enterprises based on big data, Proceedings of the 2016 annual conference on power industry informatization, 2016.
[7] Huang Hai-Cao, Chen Shun-De. Study on life cycle management of data assets of power grid enterprises, Management Observer, 2015, 1(564): 62-66.
[8] Yang Jun, Chen Wei, Wu Chun-Yang. Research on data asset management system of power grid enterprise, China electric power enterprise management, 2014, 1: 186-187.
[9] Liu Shi-Min, Zhu Ji-Yang, Zhang Meng-Meng, et al. Electronic test, 2017, 7: 70-71.