

Design of a source data management system for data middle platform

Zhi Yao¹

¹ Guangdong Provincial Key Laboratory of Optical Information Materials and Technology & Institute of Electronic Paper Displays, South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou, China
2020023804@m.scnu.edu.cn

Abstract. As a mature mechanism for enterprise management to use and process data in recent years, data middle platform has been widely used and developed by enterprises with its unique data screening and management organization mode. Because the data middle platform can establish a proprietary business model and system structure according to different enterprises, it can also have strong applicability. This paper makes a structural analysis on the current popular data middle platform, and makes an innovative management design from the source data management level to optimize the traditional data middle platform, so as to realize a more efficient and centralized management system.

Keywords: Data middle platform, Source data management, Structural optimization.

1 Introduction

One kind of data middle platform is a systematic project integrating various management, systems and control factors. Due to its flexibility, variability and applicability, the project can face a variety of enterprises and institutions, and can be a hot research and development direction in related fields at home and abroad.

At present, China's data middle platform industrial chain can be divided into three parts: upstream, middle and downstream. The participants in the upstream are mainly providers of basic services such as network, hardware and cloud services. The midstream is various types of middle platform enterprises, and the downstream is customers in different industries. The gradual improvement of Digital China Taiwan industrial chain will be conducive to the large-scale development of the whole industry.

The development of the data middle platform industry is facing many challenges, such as high requirements for real-time data, high computing costs, negative cases that raise questions about the value of the middle platform and reduce customer confidence. In addition, the homogenization problem in the data middle platform and the market competition have become the constraints of its development, but at the

same time, such challenges have also become the research and development direction of the innovative development of the data middle platform.

In addition, the application and development of the data middle platform in the enterprise also provides more requirements for the data center, which makes the data middle platform more suitable for the enterprise business scenario. The data middle platform should further broaden the application scenarios of enterprise digital middle platform, so that enterprises can operate in a smarter way and realize intelligent response and business innovation.

2 Construction ideas and structure optimization

2.1 Data asset management platform

According to our goal and the structure of the data center, we conducted a lot of simulation, discussion and analysis, and finally obtained the overall construction idea as shown in Fig. 1:

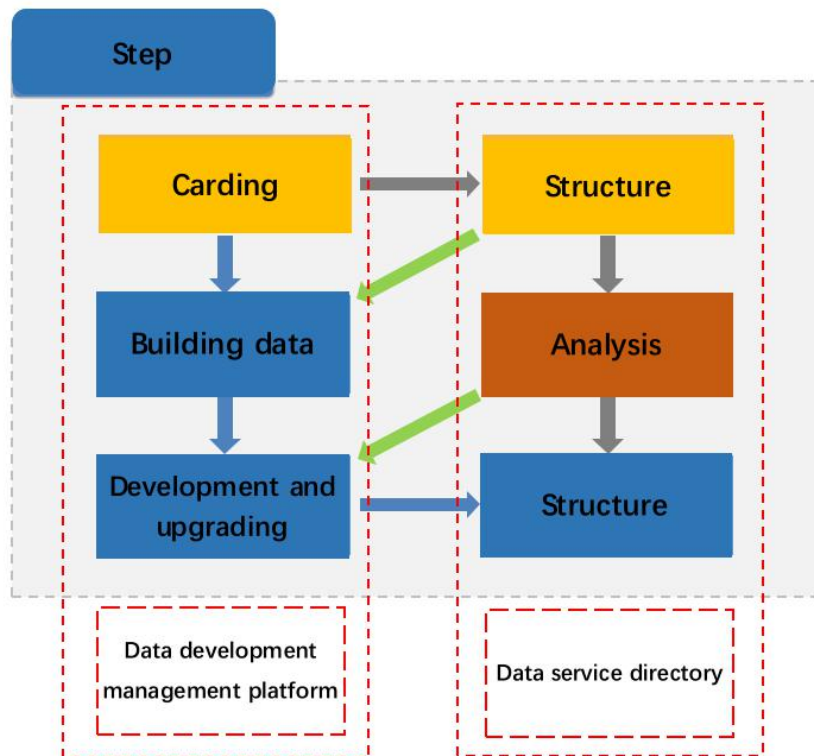


Fig. 1. The overall construction idea.

2.2 Multi source data aggregation

Starting from the specific application scenario and according to the user needs of the scenario, census all kinds of data types involved in the scenario, such as structured data, unstructured data, industrial equipment data, etc., locate their storage location and mode, collect and store them through different processing processes to the data asset management platform to form a data source entity.

Among them, structured data needs to be designed through the global database model to form a standard and authoritative unique data table; Unstructured data (such as video, pictures, etc.) shall be stored through feature processing based on scene requirements to form feature files, and provide a unified storage mode of "data (pictures, audio, etc.) - labels" and the ability of rapid positioning and retrieval; Industrial equipment data mainly refers to the data provided by the equipment with interface and acquisition conditions. The system is mainly responsible for: Being able to collect industrial equipment data with data access conditions; Standard interface definitions can be provided for equipment data encapsulation according to interface definitions.

At the same time, the system shall provide ODS layer to carry data source entities. ODS layer should have strict logic control with other DWD, DM and other layers. For example, DWD data must come from ODS layer, and DM must be processed by DWD before entering.

The multi-source data aggregation structure is shown in Fig. 2.

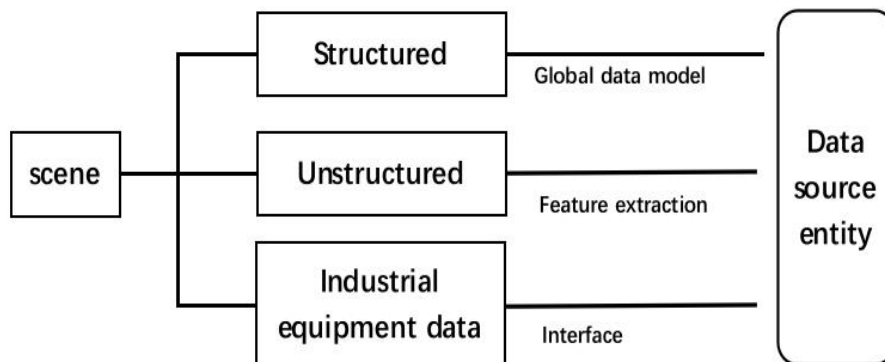


Fig. 2. Multi source data aggregation structure.

2.3 Multi source data aggregation process model

The multi-source data aggregation process model is shown in Fig. 3.

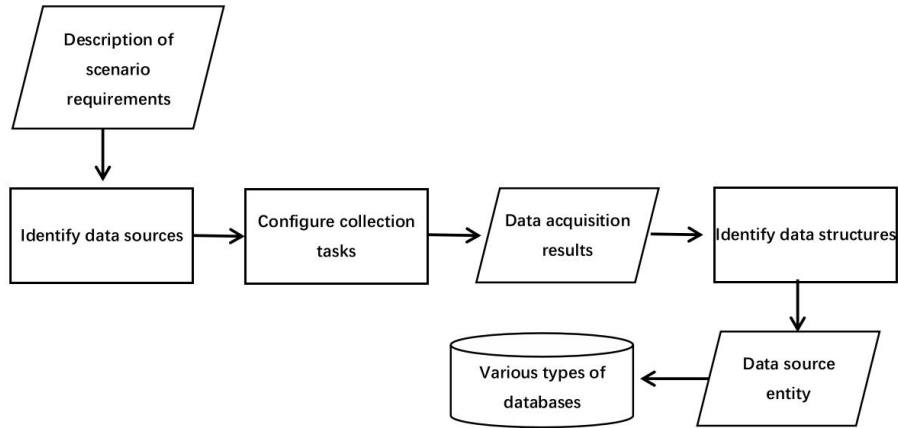


Fig. 3. Multi source data aggregation process model.

3 Source data management presentation

3.1 Data acquisition

Structured data source management. It is required to support the registration management of the data source information of the relational database instance.

- It provides registration of Oracle and MySQL relational database instance address information, database access user name and password information, as well as classification description information such as the system and user department of the database instance;
- It provides alias naming for structured data sources. Provides alias naming for structured data sources;
- It provides the functions of modifying, deleting and maintaining the registration information of relational data sources;

Structured data source management. It is required to support organizing relational structured data sources by project.

- It supports querying data sources by alias and display them in a list;
- It supports querying data sources according to the system name of the database and displaying them in a list;
- It supports querying data sources according to the user department of the database and displaying them in a list;
- It supports to view the registration details of the selected structured data source;
- It supports the functions of selecting, adding and deleting structured data sources according to the project;
- When the unauthorized relational data source is selected for use, the application authorization mechanism is provided. The application is approved by the owner of the

relational data source (project owner, data source creator and platform administrator, which are set according to business needs). After the approval authorization is passed, the applicant obtains the access token and obtains the use right of the data resource

3.2 Multi source heterogeneous data acquisition

It supports the collection of structured and unstructured data source data, and stores the collected data in the target data storage of several sets.

Structured data collection. It supports Oracle and MySQL relational database table data collection.

- It supports full table and full volume data collection of relational database;
- It supports incremental data collection of relational database;
- It supports data collection for relational databases, select tables and select table fields;
- The collected relational data can be bound with the target relational storage database resources to store the collected data;
- The storage retention time of the collected data in the relational storage of the data center to realize the automatic timing or manual clearing mechanism

Structured data collection. It is required to support the monitoring of relational data collection process.

- It supports the monitoring of relational data collection tasks, and obtains the task execution progress, task duration, task execution status and other information in real time for list display;
- It supports querying collection tasks by data source name and task creation time;
- It supports the log view function of executing tasks;
- It supports the function of redoing failed tasks. It needs to support automatic redo attempt. Automatic redo attempt (three times by default) after any failure, manual redo operation shall be carried out manually after solving the problem

4 Conclusion

Through the data source management system proposed in this paper, we can realize flexible and effective data storage, data transmission and security means. In terms of static data storage, it should support the permission management of data storage folders and single files, in terms of dynamic permission control, it can support role-based access permission management, and can conduct secret flow for secret information and personnel handling corresponding information.

In addition, our system will also establish a perfect role authorization management with complete authorization function. Considering the use of user volume, data volume, server load, load balancing, resource elastic loading and other performance optimization strategies, our system will reduce the responsibility of the platform as

much as possible, reduce the user operation response time limit, and enhance the user's experience of using the database.

References

1. Nour, M.: A dynamic open access construction product data platform[J]. *Automation in Construction* 19(4), 407-418 (2010).
2. He, X.: Construction and Application of Higher Vocational Schools Evaluation Data Acquisition Platform. *Higher Education Development and Evaluation* (2009).
3. Huang, Z., Liu, Q.: Construction of Data Mining Application Platform Based on the Service Models of Cloud Computing. *Telecommunications Science* (2012).
4. Dai, W.J., Wang, X.F.: Design and Construction of Data Integration Platform Based on XML and BizTalk Technology. *Computer Technology and Development* (2008).
5. Zhang, X.X., Zhang, L.: Construction of Platform for Decision-making Management and Data Center in Hospitals. *Chinese Medical Equipment Journal* (2012).
6. Shan, Z., Wu, J.P., Zhu, Y.P.: Construction of Distributed and Heterogeneous Data Sharing Platform. *IEEE* (2009).
7. Zhang, S.: Design and construction of a multi-radar data fusion simulation platform. *Foreign Electronic Measurement Technology* (2010).