Exploration of Oilfield Data Fusion Mode Based on ToGAF Framework

Ying Qiao and Hongyue Qiu

ABSTRACT

In the process of digitalized management of oilfield striding towards intelligent construction, how to effectively fuse oilfield data with multisource, heterogeneous, massive and distributed characteristics has become one of the problems to be solved urgently. In this paper, various data fusion models are compared, and the oilfield data fusion model based on ToGAF framework is proposed. This model combines master data management (MDM) with enterprise service bus (ESB) organically, realizes the function of oilfield data-driven business processing, and truly enables enterprises to pay more attention to their core business rather than IT system, so as to enhance the comprehensive management ability of enterprises.

1. INTRODUCTION

As early as 1980, the United States put forward the concept of massive data. With the development of information technology and intelligent devices, more than 2.5 trillion bytes of data are produced globally every day, these big data are driving the shift to a data-driven society[1][2]. For oilfield, its data come from many fields in the development process, such as drilling and seismic. In the process of data acquisition, the acquisition equipment used by oilfield enterprises and the software for managing these data are different, and because of the intervention of some third-party software using self-standard data, the production of massive multi-source heterogeneous data cannot be avoided. In order to solve this problem, this paper compares various data fusion models, and proposes an oilfield data fusion model based on ToGAF framework. The model can effectively improve the efficiency of business management of oilfield enterprises and accelerate the pace of oilfield intelligent construction.
2. COMPARISON OF MULTIPLE DATA FUSION MODES

2.1 Data Fusion Model in Federated Database

Federated Database System (FDBS) is a collection of cooperative but independent unit databases (CDBS). Federated database system is often used as a tool for data fusion, because of the distributed and heterogeneous characteristics [3][4]. Data in federated databases can be stored in multiple databases by various distributed ways, and the data between different databases can be queried by "federal-style" technology, so as to realize "data fusion" by "cross-database". As shown in Figure 1 is five-tier framework of federated database system: (1) Local schema: It is expressed by local data model. (2) Component schema: Derivatives for translating local patterns into FDBS common data models. (3) Export schema: A data subset of valid members in a federated database system. (4) Federated schema: Integration of multiple export schema. (5) External schema: It can be modified individually according to the actual use of users.

Figure 1. Five-tier framework of federated database system.

Figure 2. Data fusion mode structure diagram of data warehouse.

Figure 3. Data fusion model based on middleware integration.

Figure 4. Master data management framework.
TABLE I. COMPARISON OF MULTIPLE DATA FUSION MODES.

<table>
<thead>
<tr>
<th>Mode</th>
<th>Advantage</th>
<th>Disadvantage</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Federated Database</td>
<td>integration is the feature of database and there is interoperability between databases</td>
<td>data is scattered and difficult to maintain.</td>
<td>a small number of autonomous databases with good autonomy.</td>
</tr>
<tr>
<td>Data Warehouse</td>
<td>its data is centralized and topic-oriented and it help enterprises to make decision.</td>
<td>data is stored locally, occupying disk space. It can't reflect the change of integrated data in real time.</td>
<td>enterprises requiring decision analysis.</td>
</tr>
<tr>
<td>Middleware Integration</td>
<td>it can integrate unstructured data sources and has strong independence.</td>
<td>data is read-only and the interoperability between databases is not supported.</td>
<td>data fusion of structured and unstructured data.</td>
</tr>
</tbody>
</table>

2.2 Data Fusion Model in Data Warehouse

Data warehouse is a new information technology based on relational database, parallel analysis processing and distributed technology. Data warehouse has four major properties, namely: (1) Subject-oriented: data selection is generally related to the decision-making that enterprises need to make at present. (2) Integrated: data mainly comes from data sets obtained by data integration of business subsystems. (3) Time-variant: operate data as needed at the time. (4) Unmodifiable: generally does not modify historical data [5][6]. Figure 2 shows the architecture diagram of data fusion mode in data warehouse, which is composed of four parts: data source, data processing, data storage and data application. Through ETL tool operation, data warehouse eliminates heterogeneity of data from multiple data source systems, then loads data into data warehouse, after that enterprise managers establish corresponding application management through these cleaned and integrated data. In fact, data fusion mode in data warehouse is to convert multi-source heterogeneous data into common data.

2.3 Data Fusion Model in Middleware Integration

Middleware is located in the upper layer of operating system software, network and database, and in the lower layer of user application software. There are many kinds of middleware, mainly including database middleware, remote procedure call middleware [7][8]. A data fusion model based on middleware integration proposed in this paper is to support data type conversion between heterogeneous data sources. It belongs to database middleware. As shown in Figure 3, each wrapper implements the processing of multi-source heterogeneous data sources to enhance the degree of data fusion and integration, and then achieves the scheduling operation between users and data by the universal data access interface provided by the mediator.

2.4 Comparison of Multiple Data Fusion Modes

All three data fusion modes have solved the problems of sharing with intercommunication between system and system, data and data, data and system to a certain extent, but there are also differences and how to use should depend on the specific situation.
As shown in TABLE I, The federated database is oriented to the integration of multiple database systems, which can realize the interoperability between heterogeneous data sources and databases, but its data is relatively scattered. Comparatively speaking, the data of data warehouse is centralized and subject-oriented, and it is more convenient to provide decision support for enterprises. Compared with the data warehouse, the middleware model mainly integrates structured and unstructured data sources, and its biggest advantage is that the integrated data source is completely autonomous, and the management is more flexible and convenient, however, the disadvantage is that the data is usually read-only, while the data in the above two ways are readable and writable.

### 3. DATA FUSION MODEL BASED ON TOGAF FRAMEWORK

In order to better integrate enterprise information resources management with big data, this paper proposes a data fusion model based on ToGAF framework: ToGAF is an open enterprise architecture defined by The Open Group, which divides enterprise architecture into four interrelated domains: business architecture, data architecture, application architecture and technology architecture. [9][10][11].

#### 3.1 Master Data Management based on ToGAF Framework

Master data is the core data of enterprise information management. Master data management is a set of standards and schemes for generating and maintaining master data[12]. The master data management proposed in this paper combines the data architecture with technology architecture based on the ToGAF architecture to solve the problem of enterprise data decentralization and to escort enterprise management.

##### 3.1.1 MANAGEMENT SOLUTION OF MASTER DATA

Master data management is divided into business management and technology management. Business is the core of an enterprise. In the process of master data business management, we should do "Four Normalizations": (1) Standardization of Classification: support a variety of flexible classification. (2) Systematization of Coding Rules: meet flexible coding requirements. (3) Specification of Attribute define: determine the ownership and maintenance of master data attributes. (4) Practicability of Approval Process: define standard and feasible master data approval process. Master data management should attach equal importance to business and technology management. Master data technology management should do: (1) Master Data Acquisition and Maintenance: confirm the definition of the master data and find the target data for collection. (2) Quality Assurance of Master Data: operate standardization and deduplication of collected data. (3) Clear Management Process: give permission according to administrator level. (4) All-round Sharing of Management: such as all-round sharing of publish mechanism, integration standards management etc.
3.1.2 MANAGEMENT FRAMEWORK OF MASTER DATA

Perfect master data travel can raise the benefits of enterprises. This section presents a master data management framework, as shown in Figure 4. It is mainly divided into three parts: data integration platform, master data platform and master data service. Data integration platform integrates all kinds of data to provide data resources for the management of the master data platform. The master data platform manages the master data and provides support for the master data service, such as supply and marketing management, asset management, financial management, etc. This paper puts forward the "trilogy" of master data management platform, that is, to achieve the "Four Normalizations" of data, the simplification of process and the agility of business. "Four Normalizations" of Data: data standardization, data governance, data integration and data architecture. Process simplification requires consistent business views, internal and external application collaboration, end-to-end management, etc. Business agility should be able to achieve business intelligence analysis, custom query analysis, etc.

3.2 Enterprise Service Bus Based on ToGAF Framework

Enterprise Service Bus (ESB) is used to implement accurate, efficient and secure transmission of different messages and information in enterprise applications [13]. The enterprise service bus proposed in this paper combines the business architecture, data architecture, application architecture and technology architecture, which aims to provide an interconnection platform for oilfield enterprises.

3.2.1 CORE IDEA OF ENTERPRISE SERVICE BUS

The ESB connects to multiple systems and sends data from each system to where it is needed. Its functions include service packaging, data mapping and so on. The core idea of enterprise service bus is to take components as the core, process as the main line, protocol as the link, and data as the carrier. As follows:(1) Component centric: each component provides external services, and it can also consume services provided by other components or traditional programs.(2) Take process as the main line: scheduling process according to actual needs, business processing and data processing.(3) Take Agreement as Link: processes bind services through interfaces, and the implementation of service binds to interfaces. Binding interfaces to specific protocols in real time without affecting the process itself. (4) Take Data as Carrier: data transfer, data conversion, data publication.

3.2.2 MAIN FUNCTIONS OF ENTERPRISE SERVICE BUS

Enterprise Service Bus (ESB) is a distributed architecture that provides collaboration, control and management between services by means of "soft bus". It is also a middleware mode that can realize unification and connection of services, applications and resources and provide a standard integration mode for applications and services[9]. Its main functions are as follows: (1) Integration of multiple systems by data conversion and protocol conversion. (2) Build new business processes by
packaging legacy system services. (3) Eliminate technical differences between different applications (Cross-operating system and language).

Figure 5. Framework for the combination of master data management and ESB.

Figure 6. The architecture of master data and enterprise business system.

Figure 7. Data fusion model based on ToGAF framework.

3.3 Data Fusion Model Based on ToGAF Framework

As shown in Figs. 5 and 6. The framework and system architecture for the combination of master data management and enterprise service bus proposed in this paper aims to make data-driven business and improve the efficiency of business management in oilfield enterprises. The data fusion model based on the ToGAF framework is shown in Figure 7. The model is divided into four parts, including framework vision, framework requirements, and framework design and framework implementation. The framework vision requires a comprehensive considerations, clear framework principles, business, technology strategies and stakeholders. Framework requirements should clarify the specific needs of the enterprise, establish corresponding constraints and propose assumptions. Framework design mainly consists of three parts: business framework, information resource framework and technology framework. Information resource framework and technology framework provide support for business framework. The three frameworks are different, but they
all contribute to enterprise business processing. The model links oilfield enterprises from data to business, adding bricks and tiles to business management.

4. CONCLUSIONS

In order to solve the problem of data fusion in Oilfield, so as to run-up the intelligent management of oilfield enterprises, this paper takes data fusion as the core and aims at improving the efficiency of oilfield enterprise management, compares various data fusion modes, and puts forward an oilfield data fusion model based on ToGAF. The model combines master data management with enterprise service bus organically, making data-driven business the central point of oilfield intelligent management. On the one hand, it solves the problem of multi-source heterogeneous data fusion in oilfield, on the other hand, it effectively improves the efficiency of business management in oilfield enterprises, and truly enables enterprises to pay more attention to their core business rather than IT system, so as to enhance the comprehensive management ability of enterprises and strengthen the core competitiveness of enterprises.

REFERENCES