Research on Data Platform Construction Methods for Multi-Business Group Enterprises

Mingming Dong [1, a], Jianjun Lin [1, b], Zuyi Hao [1, c], Bo Li [1, d], and Lihua Gao [1, e]

[1] Network Information Department, China North Vehicle Research Institute, Beijing, China.

[a] ming_ming_dong@163.com, [b] 11267216@qq.com, [c] 1305225022@qq.com;

[d] type_2013@163.com, [e] gaolihua_buaa@qq.com.

Abstract. With the rapid development of digital technologies, accelerating digital construction and transformation has become an urgent priority for enterprises. The core of digital transformation lies in data-driven business innovation and growth. However, many multi-business group enterprises face significant challenges in unified data management and circulation across subsidiaries due to independent operations of various business units and the absence of a unified data platform. This hinders the effective utilization of data empowerment. Addressing the needs of multi-business group enterprises for unified data storage, sharing, and data empowerment, this paper analyzes common problems and their causes in data management within such groups. It proposes a construction approach for a group-level unified data platform, designs the platform's functional architecture, and outlines a data sharing and exchange model for subsidiaries. This provides a reference for building a unified group data platform, aiming to foster continuous innovation and enhance the competitiveness of group enterprises.

Keywords: Data Platform; Digital Transformation; Data Assets.

1. Introduction

Data and information are the economic lifeblood and strategic assets of enterprises. In the context of informatization, information systems provide the framework, but data embodies the core business value [1]. In 2020, data was elevated to the status of the fifth fundamental factor of production, alongside land, labor, capital, and technology. Leveraging and mining data enables enterprises to promptly adjust business development directions and optimize resource allocation amid increasingly intense global competition, thereby further enhancing market competitiveness[2]. Within multi-business group enterprises, various business units are often operated independently by subsidiaries, with the group headquarters primarily fulfilling a management function. Differences in core businesses and data standards among subsidiaries lead to significant challenges in group data management and application, such as dispersed data storage, difficulties in exchanging and sharing data between systems, and a lack of unified paradigms for system operation and management [3-5]. To address these issues, building a group-level unified data platform has become the foundation for realizing data applications in multi-business group enterprises[6-7]. This paper focuses on researching data platform construction methods for multi-business group enterprises. It proposes a platform construction approach, designs the functional architecture of the group data platform, and outlines data sharing models for subsidiaries. This provides support for the construction of a unified group data platform, aiming to ensure smooth internal data sharing and efficient application within multi-business group enterprises, ultimately achieving goals such as cost reduction and efficiency improvement, business innovation, activation of data assets, and risk prevention.

2. Analysis of Current Data Management Status

In recent years, with the continuous advancement of information and digital technologies, subsidiaries within multi-business group enterprises have successively built a series of information systems, digital R&D tools, and digital production lines. Taking a R&D and manufacturing group as an example, these systems include OA (Office Automation), project management, PDM (Product

Data Management), ERP (Enterprise Resource Planning), MES (Manufacturing Execution System), etc. This has essentially achieved comprehensive digital coverage across all stages, including product R&D, testing and verification, production, procurement and logistics, sales, and maintenance. Furthermore, data generated throughout product R&D, testing, sales, and support has evolved from solely focusing on results to encompassing the entire lifecycle, and from sparse sampling to comprehensive coverage. Data has become a core element of enterprise operation and management, providing a rich source for a unified data platform. However, due to differences in business modules and varying levels of digital maturity among subsidiaries within multi-business group enterprises, persistent problems remain in data quality, data product development capabilities, and data interaction channels. This results in difficulties in data circulation and sharing between subsidiaries, hindering the effective role of data empowerment. For instance, group leadership cannot grasp product R&D progress in real-time; production data cannot effectively feedback to design units; and large volumes of test data and batch production data lack suitable tools for management and mining, failing to effectively support product iteration. The underlying causes are summarized as follows:

1) Data silos problem: Data is stored and managed in a dispersed manner across subsidiaries within the group. Cross-unit data sharing and management are difficult, preventing unified development and utilization. Data silos are severe.
2) Data trust problem: A lack of trust exists among subsidiaries within the group, leading to reluctance to share proprietary data.
3) Data quality problem: The level of data governance varies significantly among subsidiaries within the group. Common group-wide master data codes and other data standards are inconsistent, resulting in low data quality.
4) Data product development Problem: Some subsidiaries within the group have not established data platforms and lack data development environments and capabilities.
5) Data interaction problem: There is a lack of secure and reliable data interaction channels between heterogeneous data systems across subsidiaries within the group.

To address the needs of multi-business group enterprises for data integration, sharing, empowerment, and the formation of high-quality data assets, constructing a unified group data platform is essential. Such a platform would aggregate product lifecycle data, manage common group data assets, and create a group data pool, efficiently solving the above problems. By enabling data to feedback into product R&D, production, sales, and maintenance, and assisting group leadership in rapid decision-making, the platform can lay the foundation for data-driven digital transformation, driving continuous innovation and enhanced competitiveness across the entire group.

3. Data Platform Construction Approach

Focusing on the entire product lifecycle within multi-business group enterprises, and guided by the core concept of comprehensive data integration management and application, a unified group data platform should be constructed leveraging next-generation information technologies. Through unified collection, management, governance, and application of group data assets, the platform aims to achieve online management of the project lifecycle and synergistic interaction between R&D, production, sales, and maintenance data. This will enhance the efficiency and coordination of overall business processes within the multi-business group enterprise, ultimately achieving objectives such as cost reduction and efficiency improvement, business innovation, activation of data assets, and risk prevention. The group data platform can adopt a “1+M+N” construction approach, as shown in Fig.1. “1” represents the “group-side data platform”, which can be deployed on the group's public cloud or within the group headquarters. It acts as the central node, collecting and aggregating data from all subsidiaries, other group-side business systems, and group-shared master data. Data collected from subsidiaries is stored in partitioned and isolated areas. After

governance, it forms group-shared data assets, providing unified services externally. "M" represents "data collection tools" deployed at each subsidiary. These tools are responsible for collecting and aggregating data from various business systems and equipment within the subsidiary and sending it to the group-Side data platform. "N" represents the "self-built data platforms" of each subsidiary. Subsidiaries with their own platforms can send data directly to the group-side data platform. For subsidiaries that have already built or plan to build their own data platforms, the M-type collection tool is not deployed. Additionally, manual data entry accounts on the group-side data platform are provided to each subsidiary for inputting data not managed by existing business systems.

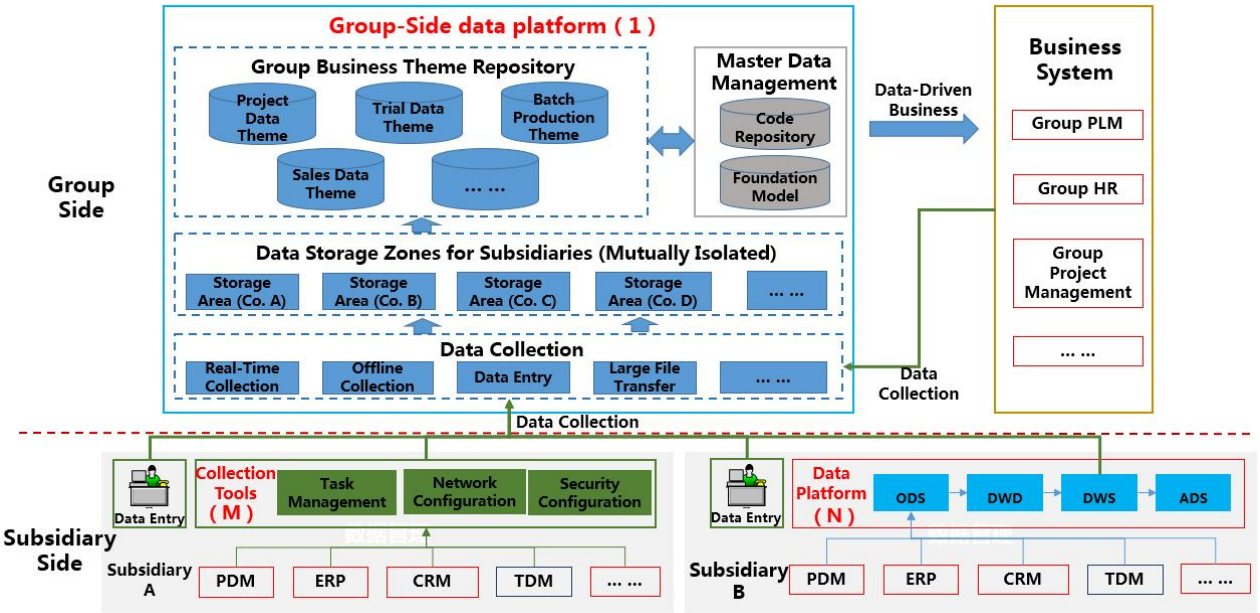

Fig. 1 Group data platform construction approach

The group-side data platform and data collection tools are centrally constructed by the group. Subsidiaries are responsible for building their own data platforms if required. Concurrently with the construction of functional modules on the group-side data platform, group-wide data governance efforts should be coordinated. This includes unifying master data and data exchange standards, and planning data application scenarios to improve construction benefits. Subsidiaries with better foundational conditions and clear data application scenarios can pilot the application first, accumulating implementation experience. This ensures the group data platform effectively solves practical problems and improves the operational efficiency of both the group and its subsidiaries.

4. Data Platform Functional Architecture

4.1 Group-Side Data Platform Architecture

Following the "1+M+N" approach, the group-side data platform must encompass full data lifecycle management functions. It primarily consists of three major functional modules: data foundation platform, data management, and data application, as shown in Fig.2. Data foundation platform provides secure, reliable, low-cost, and elastically scalable big data storage and computing services. It supports one-stop creation, management, deployment, operation, and monitoring. Key functions include: data source management, distributed deployment data storage & computing, data calculation, SQL support, offline batch processing, real-time stream computing, ad-hoc query, operation & maintenance management, account & permission control. Data management module provides full lifecycle management capabilities for data collection, storage, governance, and service. It integrates all aspects of data management, enabling synergy among various data resources and supporting diverse business applications. It supports flexible collection methods to aggregate

massive heterogeneous data. Based on this, data governance enhances data quality and unifies data definition rules. Processed, classified, and integrated business data is organized into layered subject-oriented data asset libraries based on business themes. Data service interfaces support the implementation of BI and AI application scenarios. Data application module: A BI and AI development platform with capabilities for visual BI dashboard creation, full lifecycle machine learning management, and intelligent algorithm training and execution. Through functions like visual modeling, operator management, and model management, it enables one-stop dashboard creation, AI model training, and deployment management. Leveraging this module, hidden business patterns and development trends within data can be discovered, unlocking the powerful potential of data, enhancing user decision-making capabilities, and empowering business management and operations to achieve goals like process optimization and business model innovation. This module primarily consists of: visual modeling, drag-and-drop operator components, data mining model library, platform management.

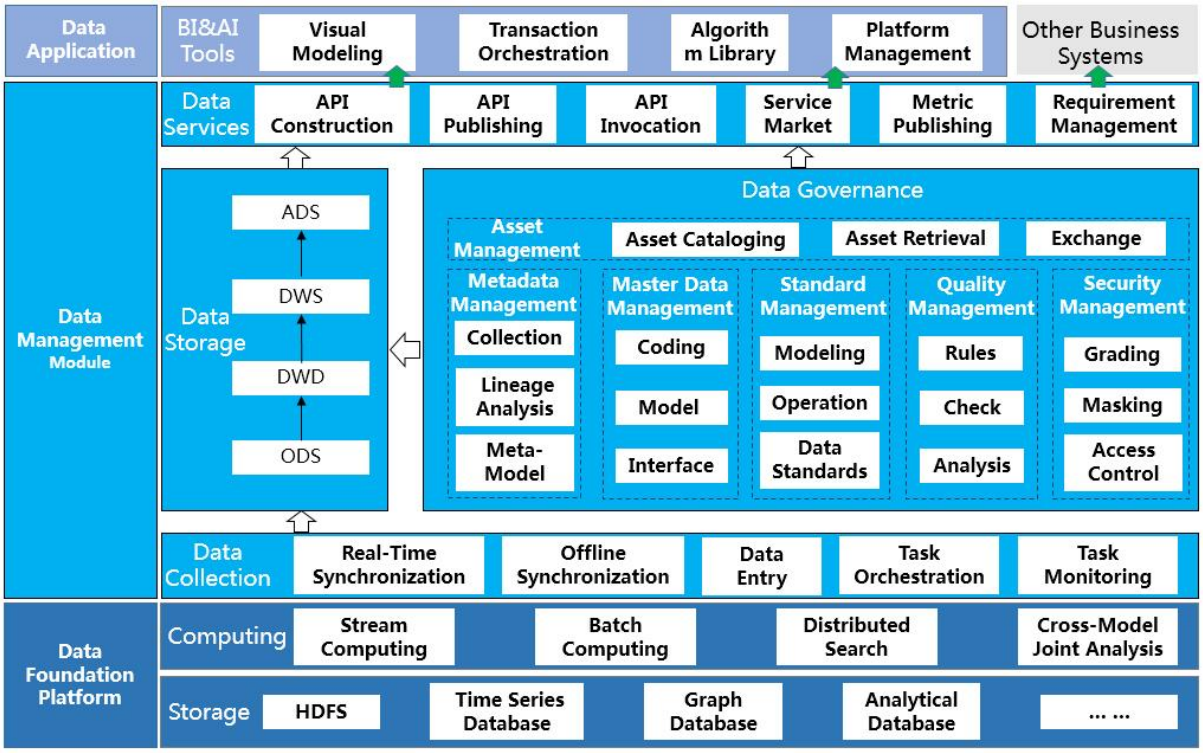

Fig. 2 Functional architecture of the group-side data platform

4.2 Data Collection Tool Architecture

The data collection tool is responsible for collecting data from various business systems or devices within group subsidiaries. It should support both offline and real-time data collection and primarily includes functions for network security configuration, collection task orchestration, and monitoring, as shown in Fig.3

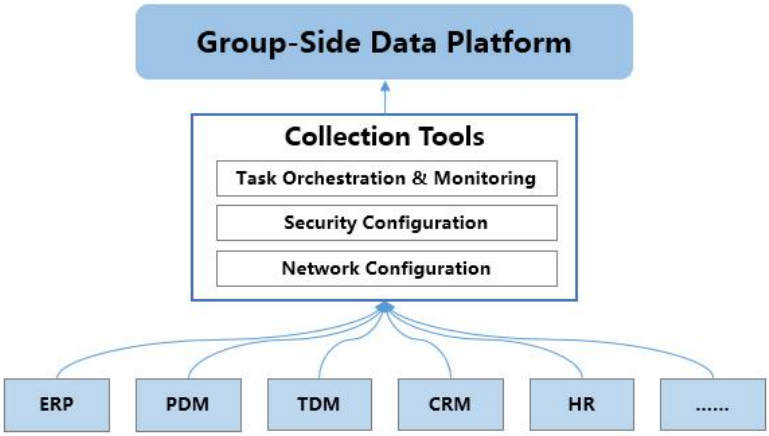

Fig. 3 Functional architecture of the data collection tool

1) Network configuration: Supports configuring collection addresses for different types of databases and business systems.
2) Security configuration: Supports configuring different data access permissions and data transmission mechanisms.
3) Task orchestration & Monitoring: Supports graphical drag-and-drop orchestration of data collection tasks, real-time monitoring of task progress, and both offline and real-time collection modes.

5. Data Sharing and Exchange Model

After group data is collected, stored, and governed, sharing and exchange between data providers and consumers are conducted based on the group data asset catalog, ensuring controlled data permissions. As shown in Fig.4, the group platform administrator can build data services using aggregated data on the platform based on clearly defined group business needs. Subsidiaries can also act as data providers, extracting valuable data from their own business systems, processing it into data services, compiling data catalogs, and synchronizing them to the group data platform. The group data platform aggregates catalogs from all units to form the group's master catalog, which is published for querying and browsing by all units.

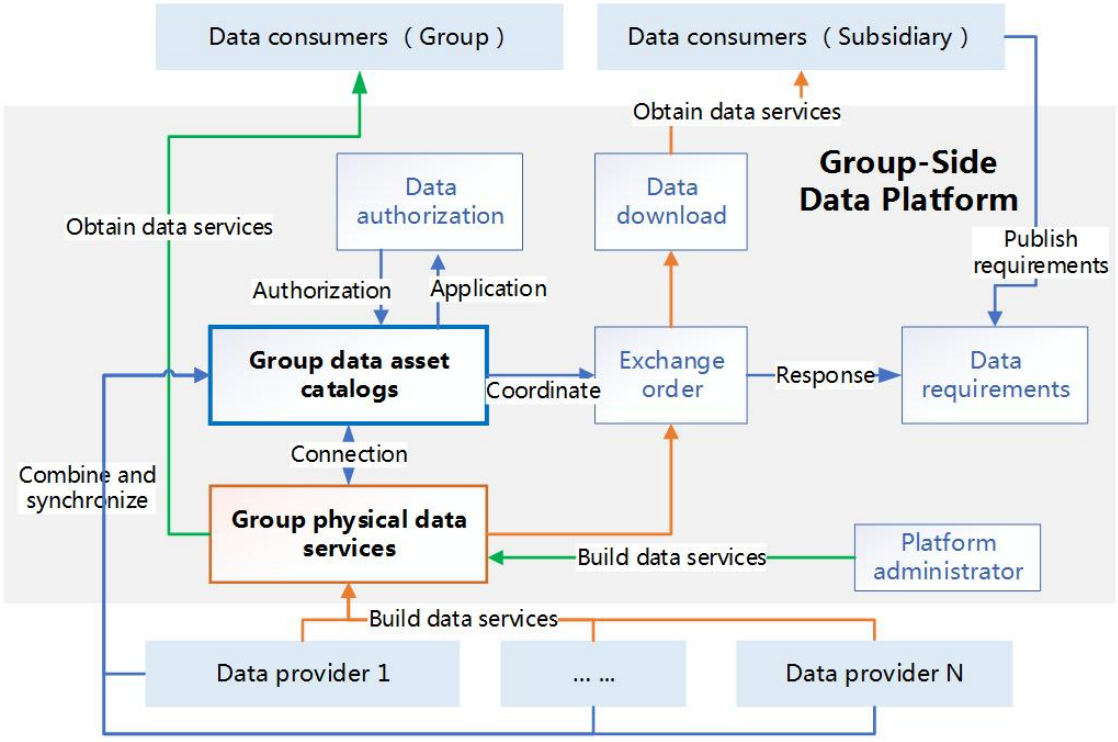

Fig. 4 Data sharing and exchange process

Data services built by the group platform administrator based on explicit group business needs can be used directly by authorized personnel without platform approval. Other data consumers can publish their data requirements on the group data platform or search for useful data services through the platform catalog. The group data platform facilitates the connection and response between providers and consumers through a data exchange order, executing activities like data application, authorization, viewing, and download to complete the data transaction process. It also records the transaction. Physical data services can be delivered via the platform or provided offline directly. In the catalog construction process, group experts propose unified standards and specifications. Subsidiaries process their own catalogs according to these asset catalog standards and regularly aggregate them to the group platform. For subsidiaries adopting the “N mode” (with self-built data platforms and data asset processing capabilities), they can process asset catalogs on their own platform and then synchronize them to the group platform. For subsidiaries adopting the “M mode” (lacking data processing capabilities), they need to connect business data required for the catalog directly to the group-side data platform using the data collection tool. They then utilize the data processing capabilities of the group-side data platform to process assets and form the catalog, as shown in Fig.5.

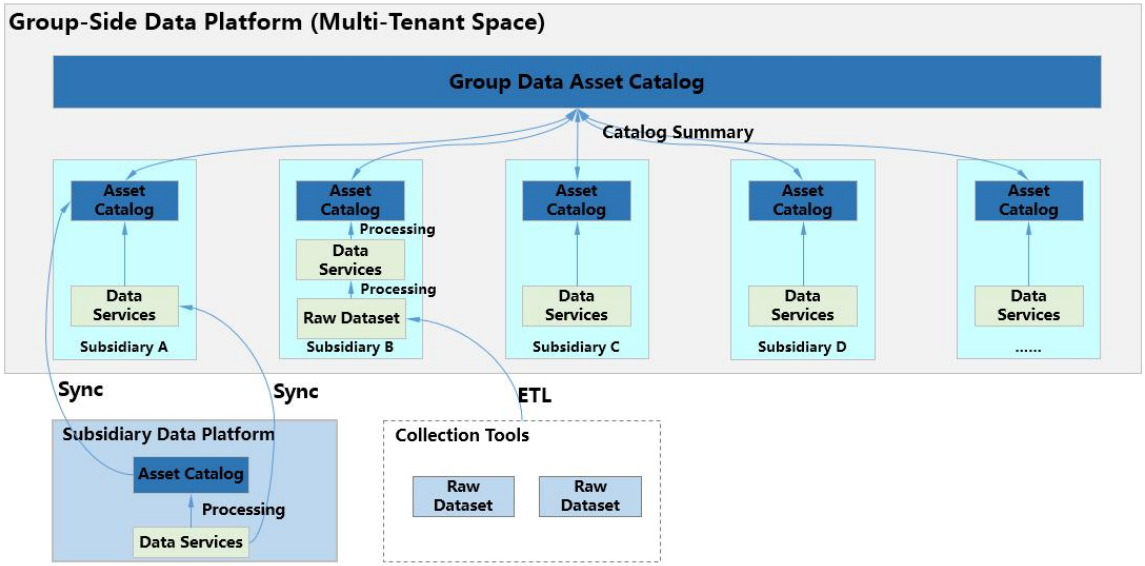

Fig. 5 Data asset catalog construction methods

6. Summary

With the continuous improvement of informatization levels among subsidiaries within multi-business group enterprises, data is no longer just a by-product of applications but the core driver of new businesses and applications, representing a strategic asset for the group. However, many such groups currently lack a unified data platform, leading to dispersed data storage and management across subsidiaries, difficulties in data circulation and sharing, and challenges in effectively leveraging data empowerment. Focusing on data integration and sharing within multi-business group enterprises, this paper analyzed existing problems and causes in data management. It proposed a construction approach for a group data platform, designed its functional architecture, and outlined data sharing models for subsidiaries. This provides support for building a unified group data platform, aiming to assist multi-business group enterprises in activating their core data assets and achieving digital transformation and upgrading.

References

[1] LIU Wen. Business-Driven Data Governance Method for Aerospace Enterprises. Information Technology and Informatization, 2019, (5):62-68.

[2] Lu Jingjing. Research and Application of Material Master Data Management in Large Enterprises. Lanzhou: Lanzhou University, 2016.

[3] HUANG Liang, Fu Dezhi, GOU Fan. Enterprise Data Governance Sharing Approach and Framework for Multi-Business Groups. Network Security Technology & Application, 2024, (10):53-56.

[4] TIAN Jiang. Practice and Thoughts on Distributed Data Sharing Governance. Software and Integrated Circuit, 2021, (04):40-43.

[5] Zhan Xiufeng, Shao Xiaowei, Rong Zhengguan, Wang Jijun. Research and Practice on Group Data Indicator Management. Technology Wind, 2024, (23):135-138.

[6] Zheng Yaodong, Li Xufeng, Chen Heping. Exploration of Data Middle Platform Construction Methods for Industrial Enterprises. Modern Information Technology, 2020, 4(21):130-133.

[7] LI Yuan, QIN Kun, CAO Yun. Research on Data Middle Platform Architecture for Military Enterprises. Computer Knowledge and Technology, 2021, 17(8): 25-26.