Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 12, No. 2, 2025

Optimizing the Integration of Enterprise Data Warehouses and Large Language Models

Wentao Xu *

Hangzhou Huaguang Advanced Welding Materials Co., Ltd., Hangzhou, Zhejiang 311100, China
* **Corresponding author Email:** 270821734@qq.com

Abstract: With the in-depth advancement of the digital transformation strategy, enterprise-level data warehouses, as the core infrastructure for intelligent decision-making, are confronted with technical bottlenecks such as insufficient dynamic expansion capabilities and low real-time analysis efficiency. However, the rapid development of large language model technology provides a new path for the intelligence of data processing. This paper systematically explores the integration logic and technical adaptability of the two types of technologies, revealing the feasibility of upgrading the data processing paradigm through model light-weighting technology. This study reviews the application efficiency of the existing system in scenarios such as real-time analysis and resource allocation from the perspective of technical synergy, and demonstrates the development trend of the converged architecture in directions such as trusted computing and multi-modal processing. This paper provides a theoretical framework and practical reference for constructing a new generation of data infrastructure that supports dynamic optimization and intelligent decision-making, and has guiding significance for promoting the release of the value of data elements.

Keywords: Data Warehouse; Large Language Model; Efficient Fine-tuning; Intelligent Data Analysis.

1. Introduction

With the in-depth advancement of the digital transformation strategy on a global scale, enterprise-level data warehouses, as the core infrastructure supporting intelligent decision-making, are facing a critical turning point in their technological evolution. The country's "14th Five-Year Plan" for the development of the digital economy clearly states that a data element market system should be established to drive enterprise-level data platforms to upgrade towards real-time and intelligent directions. The current mainstream data warehouses, through the integration of HTAP architecture and cloud computing technology, have achieved large-scale applications in fields such as financial risk control and supply chain optimization. However, bottlenecks such as insufficient dynamic expansion capabilities and low real-time analysis efficiency have become increasingly prominent. Meanwhile, large language model technology, with its powerful semantic understanding and generation capabilities, is gradually penetrating the entire data processing process. The AI risk management framework released by the National Institute of Standards and Technology (NIST) of the United States has even listed the lightweight deployment of models as a key development area.

This study focuses on exploring the internal logic and technical path of the integration of enterprise-level data warehouses and large language model technologies, and systematically analyzes the complementary of the two types of technologies in terms of data processing paradigms, system architecture characteristics, and optimization mechanisms. By sorting out the technical bottlenecks of the data warehouse in aspects such as dynamic optimization of multi-dimensional indexes and elastic architecture design, combined with the latest progress of large language models in the fields of low-cost fine-tuning and private deployment, the feasibility of achieving intelligent upgrading of data processing through efficient parameter fine-tuning technology is revealed. This paper comprehensively reviews the application efficiency of the existing technical system in scenarios such as real-time data analysis and dynamic resource allocation from three dimensions: technical adaptability, system collaboration, and industrial implementation. It also explores the development trends of the integrated architecture in directions such as trusted computing and multimodal processing, providing theoretical references and practical guidance for building a new generation of intelligent data infrastructure.

2. Current Situation and Challenges

2.1. Core Features and Use Cases of Enterprise Data Warehouses

As the core hub of modern enterprise data assets, the core functions and application scenarios of enterprise-level data warehouses directly determine the enterprise's data-driven capabilities. From the perspective of technical architecture, a data warehouse realizes the full lifecycle management of data through a multi-level logical architecture. The data integration platform framework proposed by Chen et al. indicates that its core functions cover data collection, cleaning and transformation, and unified storage. Among them, the data collection method interfaces with heterogeneous systems through API interfaces, and combines ETL strategies to achieve the integrated processing of structured and unstructured data, ultimately forming a standardized data model that supports OLAP analysis. In terms of application scenarios, the research by Agus et al. shows that the deep integration of data warehouses with business systems such as ERP can effectively support business scenarios like inventory optimization and supply chain collaboration. For instance, a dynamic replenishment model based on historical transaction data can reduce warehousing costs by more than 15%. Furthermore, modern data warehouses also achieve the integration of real-time analysis and transaction processing through the HTAP architecture. The research by Ferreira et al. [1] indicates that this hybrid processing mode demonstrates significant advantages in scenarios such as financial risk

control and customer profiling, supporting millisecond-level decision responses and concurrent execution of complex queries. Overall, enterprise-level data warehouses are evolving from traditional data storage carriers to intelligent decision-making hubs. The deep integration of their functions and scenarios will continue to unlock the value of data assets.

2.2. Main Problems Currently Faced by enterprise-level Data Warehouses

At present, enterprise-level data warehouses face multiple challenges when dealing with complex business requirements. Kahn et al. [2] found through the study of cloud migration cases that the heterogeneous architecture of legacy systems has insufficient compatibility with the cloud environment, resulting in problems such as difficulties in integrating data silos and lengthy security review processes during the migration process. Moreover, enterprises generally lack mature practical experience in cloud-native technologies. Santhosh Gourishetti et al. [3] further pointed out that although centralized data warehouses can guarantee data consistency, their rigid architecture is difficult to support the requirements of dynamic expansion. Especially when facing real-time analysis and high-concurrency query scenarios, the performance bottleneck is significant. Furthermore, the research by Kushal Shah et al. [4] indicates that the traditional local deployment model has the drawbacks of low resource utilization and high operation and maintenance costs. Enterprises need to balance the contradiction between data governance norms and agility requirements. These problems jointly highlight the adaptability predicament of enterprise-level data warehouses between technological iteration and business development, and it is urgent to achieve a breakthrough through technological integration.

2.3. Bottlenecks and Demands in the Development of Data Warehouse Technology

At present, the development of enterprise-level data warehouse technology is confronted with core bottlenecks such as the efficiency of multi-dimensional data processing, system scalability and real-time analysis capabilities. The multi-dimensional index structure based on R-tree proposed by Yetong Wang et al. [5] has significantly improved the query efficiency and scalability of the digital certificate warehouse for property rights, but it is still difficult to cope with the requirements of high-concurrency data integration in dynamic business scenarios. The research of Jiayi Tang et al. [6] in the field of e-commerce shows that although real-time tracking of the Internet of Things and AI-driven optimization technologies can improve space utilization and processing speed, problems such as inconsistent data quality and excessive system deployment costs have restricted the implementation of the technologies. Meanwhile, the digital twin system developed by A. Pracucci et al. [7] revealed the insufficiency of data collection and analysis capabilities in complex scenarios, and the existing architecture is difficult to meet the requirements of real-time dynamic modeling. Therefore, data warehouse technology urgently needs to break through the dynamic optimization mechanism of multi-dimensional indexing, the low-cost intelligent decision-making model, and the elastic architecture that supports heterogeneous data fusion, laying the foundation for the subsequent integration with large language model technology.

3. The Current Situation and Potential of Large Language Model Technology

3.1. Basic Principles and Technical Characteristics of Large Language Models

The large language model is constructed based on the Transformer architecture, and its core principle is to achieve context-aware sequence modeling through the self-attention mechanism and position encoding. The multi-head self-attention layer of the Transformer can capture long-distance dependencies in parallel, while position coding techniques (such as RoPE) inject relative position information into sequence elements. Chen et al. [8] have shown that although RoPE has an advantage in capturing position relationships, its mathematical expression ability is still limited by the theory of circuit complexity. During the training process, layer normalization and residual connection form the basis of stable optimization. Xiong et al. [9] analyzed the Pre-LN structure and proved that adjusting the position of the normalized layer can significantly improve the training stability and reduce the preheating requirement. Furthermore, the model acquires the ability of general semantic representation through pre-training with massive corpora and realizes task adaptation in combination with fine-tuning techniques. However, Peng et al. [10] pointed out that the Transformer has inherent limitations in combinatorial logic tasks, and its attention mechanism is difficult to effectively handle complex mathematical reasoning problems. These technical features jointly constitute the dual characteristics of high efficiency and applicable boundaries of large language models.

3.2. Application of Large Language Models in Data Processing and Analysis

The application of large language models in the field of data processing and analysis is gradually breaking through the traditional paradigm and demonstrating unique intelligent value. Through the efficient fine-tuning technology of parameters, large language models can quickly adapt to the specific requirements of enterprise-level data processing. The LoRA (Low-Rank Adaptive) technology proposed by Hu et al. [11] reduces the number of training parameters to one ten-thousandth of traditional fine-tuning by freezing the weights of the pre-trained model and injecting a low-rank matrix. It maintains comparable performance to full-parameter fine-tuning in scenarios such as data classification and information extraction, while significantly reducing the consumption of computing resources. It provides a feasible path for the lightweight deployment of enterprise-level data warehouses. In the field of human-machine collaboration, the CoAnnotating framework developed by Li et al. [12] innovatively combines model uncertainty estimation and task allocation mechanisms, enabling LLMS to form dynamic complementarity with manual annotation in the task of unstructured text annotation. Experiments show that this framework can increase the comprehensive efficiency by up to 21% while ensuring the annotation quality. It has opened up new ideas for the automated processing of massive enterprise data. Trummer et al. [13] pointed out that language models have demonstrated the potential to reconstruct data processing flows. From natural language query translation, data pattern matching to exception detection and other links, models such as GPT-3 Codex can directly generate SQL code

or cleaning rules through context learning. This end-to-end processing capability is driving data analysis towards a more intelligent direction. These technological breakthroughs not only enhance the efficiency of data processing, but also endow enterprise data warehouses with a new dimension of proactive insight into business value through semantic understanding capabilities.

3.3. Low-cost Fine-tuning and Private Deployment Technologies

In the enterprise-level application scenarios of large language models, low-cost fine-tuning and private deployment technologies have become the key research directions to break through resource constraints. The MSPLoRA proposed by Jiancheng Zhao et al. [14] successfully reduces the number of trainable parameters by more than 80% by constructing a multi-scale pyramid low-rank adaptive structure and dynamically allocating parameters in three dimensions: global, intermediate and hierarchical, while maintaining the performance advantage of cross-domain NLP tasks. This method effectively solves the problem of parameter redundancy in traditional fine-tuning through a hierarchical feature capture mechanism, providing a lightweight solution for enterprise private deployment. The ElaLoRA developed by Huandong Chang et al. [15] further introduces an elastic rank adjustment mechanism, dynamically pruning and expanding the rank based on gradient importance. Under the same parameter budget, it improves the task adaptation efficiency by 37% compared with the traditional LoRA method. This adaptive resource allocation strategy can not only meet the computing power conditions of enterprises of different scales, but also ensure the flexibility of model tuning, providing an scalable technical path for private deployment scenarios. The AG-LoRA designed by the team of Qingchen Wang [16] achieves fine-tuning performance improvement on mainstream models such as LLaMA through the initialization of singular value decomposition guided by the activation mode, combined with the global rank allocation strategy, while reducing the GPU memory consumption to 62% of the traditional method. These innovative technologies systematically build a technical system for the low-cost implementation of large language models from multiple dimensions such as parameter optimization, dynamic adaptation, and resource allocation, providing a reusable methodological framework for the intelligent upgrade of enterprise-level data warehouses and promoting the efficient transformation of large model technology from laboratory research to industrial practice.

4. Conclusion

This study systematically explores the integration path of enterprise-level data warehouses and large language model technology, revealing the core bottlenecks of traditional data warehouses in real-time analysis, dynamic expansion and heterogeneous data processing. At present, data warehouses are constrained by rigid architectures and resource utilization issues, making it difficult to cope with the demands of high-concurrency queries and real-time decision-making. However, large language models, through low-cost fine-tuning technologies such as LoRA and dynamic rank allocation strategies, have demonstrated the potential to restructure data processing flows. Studies show that the integration of the two types of technologies can build an intelligent analysis architecture that supports elastic scalability and is expected to break through the shackles of the efficiency of multi-dimensional data analysis.

References

[1] Ferreira F E R R, Fidalgo R d N. A Performance Analysis of Hybrid and Columnar Cloud Databases for Efficient Schema Design in Distributed Data Warehouse as a Service[J]. Data, 2024, 9: 99.

[2] Kahn M, Mui J Y, Ames M J, et al. Migrating a research data warehouse to a public cloud: challenges and opportunities[J]. Journal of the American Medical Informatics Association : JAMIA, 2021, 29: 592 - 600.

[3] Gourishetti S. Centralized Data Warehouse vs. Data Mesh: A Comparative Analysis of Modern Data Management Paradigms[J]. International Journal For Multidisciplinary Research, 2024.

[4] Shah K. The evolution of data warehouse architectures: from on-premises to cloud-native solutions[J]. World Journal of Advanced Research and Reviews, 2025.

[5] Wang Y, Xing K, Zheng B. Multi-Dimensional Indexing and Efficient Query Optimization of Property Right Digital Certificate Warehouse Based on Big Data Analysis[C]//2024 International Conference on Telecommunications and Power Electronics (TELEPE). , 2024: 266-271.

[6] Tang J. Applications and Challenges of Warehouse Optimization Technology in E-commerce Platforms under the Background Big Data[J]. Advances in Economics, Management and Political Sciences, 2025.

[7] Pracucci A. Designing Digital Twin with IoT and AI in Warehouse to Support Optimization and Safety in Engineer-to-Order Manufacturing Process for Prefabricated Building Products[J]. Applied Sciences, 2024.

[8] Chen B, Li X, Liang Y, et al. Circuit Complexity Bounds for RoPE-based Transformer Architecture[J]. ArXiv, 2024.

[9] Xiong R, Yang Y, He D, et al. On Layer Normalization in the Transformer Architecture[J]. ArXiv, 2020.

[10] Peng B, Narayanan S, Papadimitriou C. On Limitations of the Transformer Architecture[J]. ArXiv, 2024.

[11] Hu J E, Shen Y, Wallis P, et al. LoRA: Low-Rank Adaptation of Large Language Models[J]. ArXiv, 2021.

[12] Li M, Shi T, Ziems C, et al. CoAnnotating: Uncertainty-Guided Work Allocation between Human and Large Language Models for Data Annotation[J]. ArXiv, 2023.

[13] Trummer I. From BERT to GPT-3 Codex: Harnessing the Potential of Very Large Language Models for Data Management[J]. Proc. VLDB Endow., 2022, 15: 3770-3773.

[14] Zhao J, Yu X, Yang Z. MSPLoRA: A Multi-Scale Pyramid Low-Rank Adaptation for Efficient Model Fine-Tuning[J]. ArXiv, 2025.

[15] Chang H, Ma Z, Ma M, et al. ElaLoRA: Elastic & Learnable Low-Rank Adaptation for Efficient Model Fine-Tuning[J]. ArXiv, 2025.

[16] Wang Q, Shen S. Activation-Guided Low-Rank Parameter Adaptation for Efficient Model Fine-Tuning[J]. IEEE Access, 2025, 13: 70909-70918.