

# Research on Mixed Transaction Analytical Data Management Oriented to Data Middle Platform

Ai-hua Zhou<sup>1,\*</sup>, Li-peng Zhu<sup>1</sup>, Meng-han Xu<sup>2</sup>, Sen Pan<sup>1</sup>, Jun-feng Qiao<sup>1</sup>, Hong-bin Qiu<sup>1</sup>, Song Deng<sup>3</sup>

1.Global Energy Interconnection Research Institute Co. LTD,

Artificial Intelligence on Electric Power System State Grid Corporation Joint Laboratory (GEIRI), Beijing, China

2.State Grid Jiangsu Electric Power Co. LTD. Information & Telecommunication branch, Nanjing, China

3.Institute of Advanced Technology, Nanjing University of Posts & Telecommunications, Nanjing, China

zah831@163.com, roc@163.com, xumenghan@js.sgcc.com.cn, yingzi108@126.com,

1318558905@qq.com, cnqihongbin@163.com, dongsong@njupt.edu.cn

\*corresponding author

**Abstract**—To solve the problem of non-synchronization between enterprise application development and data development, this paper puts forward the concept of data middle platform, which combines the two data processing mechanisms of online analytical processing (OLAP) and online transaction processing (OLTP), so that faster and better data services can be provided to the foreground business. On this basis, this paper summarizes the research status of the related technologies of the data middle platform, including the architecture of the data middle platform and the key technologies of constructing the data middle platform. In-depth analysis of the business scale and business characteristics of OLTP and OLAP in various application scenarios, focusing on the technical difficulties in the application process of OLTP and OLAP in the application scenario. Finally, it summarizes the challenges faced by the basic research from three aspects: the construction of data middle platform, data quality assurance, and the application of mixed-thing analytical data management.

**Index Terms**—middle data platform, data processing, online transaction processing, online analytical processing

## I. INTRODUCTION

Traditional data processing mechanisms are roughly divided into two categories: online transaction processing (OLTP) and Online Analytical processing (OLAP). Online transaction processing (OLTP) is the main application of traditional relational databases. The basic feature of OLTP is that it can quickly respond to user operations. When user data is received by the business foreground, user data can be quickly transmitted to the computing center for data processing. The corresponding processing results are given according to the user data in a short time, which are mostly used in banks and e-commerce transactions. Online analytical processing (OLAP) is the main application of the data warehouse system. The basic characteristics of OLAP are designed based on the complex analysis operations of a large amount of data. OLAP focuses on providing decision support and can carry out query processing with a large amount of data according to the requirements of analysts, to provide intuitive query results.

At present, with the rapid development of the times, the data is in a blowout growth. Due to the limited computing power of

the software, the production system can not meet the functional requirements of historical data query and statistics, otherwise, it will lead to a large amount of data queries, which will have a serious impact on normal transactions, so the general enterprise divides the business system into OLTP and OLAP, OLTP to deal with the latest transaction business. OLAP is used for the complex analysis of massive data. However, the data development and application development of enterprises are often out of sync, the market changes to the needs of business, and the development of applications precipitates into a large amount of data, the speed of the two cannot be consistent, so the traditional architecture of separation of OLTP and OLAP cannot meet the business needs of today's enterprises.

The data middle platform is a data service factory, which connects the foreground and the background, completes the hierarchical segmentation and horizontal solution of the data, separates the public business entrance, and provides corresponding data services through API, which can quickly and flexibly respond to the needs of the front and back end of the business. The integration of OLTP and OLAP data processing makes OLTP and OLAP change from parallel relationship to vertical relationship so that they can respond to the needs of the market and users at a high speed. This paper deeply analyzes the business scale and business characteristics of OLTP and OLAP in various application scenarios, focuses on combing and analyzing the technical difficulties in the application process of OLTP and OLAP, and explores the requirements and application methods of mixed transaction analysis data management.

The rest of this paper is organized as follows. Section 2 analyzes the current research situation of data processing at home and abroad. Section 3 describes the establishment and application of data middle platform. Section 4 addresses the analysis of traditional online transaction processing and online analytical processing technology. Section 5 analyzes HTAP business and technology. Section 6 gives the prospect and challenges of data middle platform technology. The conclusion is given in the last Section.

This work was supported by the Science and Technology Project of State Grid Corporation of China (No.5700-202058480A-0-0-00)

978-1-6654-2655-8/21/\$31.00 ©2021 IEEE

## II. RELATE WORK

Gao et al. [1] established data middle-station framework to explore the hot spots of data assets, and realizes the hot spots of data assets in electric power enterprises. Wu et al. [2] elaborated on the structure and function of data middle platform by comparing the characteristics of data middle platform, data warehouse, and data platform. Lyu et al. [3] proposed a data middle platform architecture based on microservices, which could precipitate enterprise-level data resources and model service capabilities, to avoid repeated development of the same functional code and repeated model training. Ding et al. [4] proposed a domain model to build enterprise data services based on the relationship between the new intermediary platform model and digital transformation under data sharing services. Wu et al. [5] connected the wearable sports physiological parameter detection data processing platform based on the medical Internet of Things, and built a real-time interactive medical Internet of Things platform. Lu et al. [6] studied SQL-on-Hadoop and MPP technology in the OLAP and OLTP hybrid scenario to meet the business needs of enterprises. Zhang et al. [7] used a data store based on distributed lineage to solve the data storage layer and distributed architecture that can support OLAP and OLTP. Nabibayova et al. [8] examined the possibility of using OLAP technology in the field of foreign policy. Wang et al. [9] designed a discriminative admission control mechanism for a database that shares everything. To meet the requirements of enterprises, Wang et al. [10] proposed a power equipment condition monitoring data model based on connectionless hierarchical coding to provide better scalability and high performance, thereby improving the analysis performance of OLAP. In [11], [12], authors introduced the content of data sharing, data processing, and processing between the production department and management department of the enterprise. Du et al. [13] analyzed the characteristics of monitoring data, proposing a cloud storage platform can be used for the management of massive distributed line monitoring data. Francia et al. [14] proposed a MOSES processing framework for metadata processing in a big data platform, providing a wide range of functions for different types of users of the platform.

It can be seen from the above literatures that with the advent of the era of big data, data has slowly evolved into a precious strategic resource like oil. Data occupies an increasingly important position in all walks of life. A large number of researches on data middle platform and mixed transaction analysis data management have been carried out at home and abroad, and data middle platform has gradually moved from theoretical research to applied practice. Current research is mainly focused on providing data API services, maximizing the value of data assets, combining traditional OLTP and OLAP, and realizing HTAP data processing. However, different enterprises have different data requirements. The data applications of enterprises are constantly being updated. The middle-office system is constantly changing according to the needs

of the enterprise. The lack of a common general data middle-office system architecture for reference makes it impossible to widely use the data. Social applications. Therefore, building a common data middle platform based on mixed transaction analysis data management is essential to improve enterprise productivity.

## III. DATA MIDDLE PLATFORM

### A. Description and Application of Data middle platform

To better understand the data platform, we must first distinguish between the data platform and the big data platform, the data platform is not a platform, nor is it a simple system. Based on the data warehouse and big data platform, the data middle platform transforms the required data analysis into data API services and then provides them to the business more accurately and efficiently. In essence, it is a cross-business theme business system built on the data warehouse [15], as shown in Fig.1.

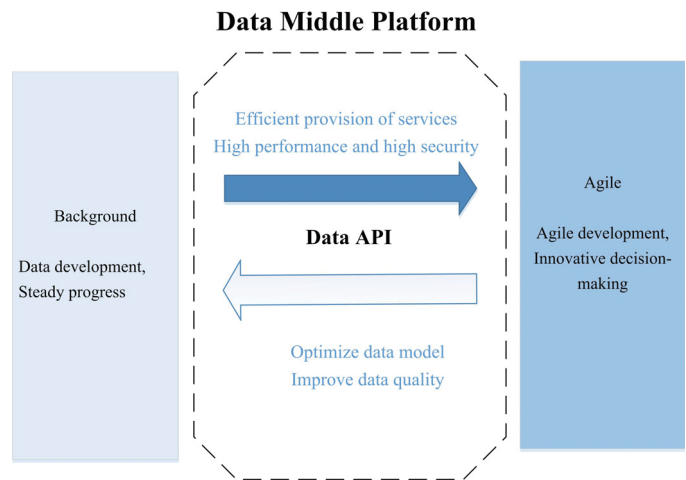


Fig. 1. Data middle platform concept diagram.

Generally, the data middle platform provides data through API for data analysis and processing, rather than directly giving the database to the foreground for its development and use. As for the process of generating Data API, how to make the production process faster, more accurate, more clear, and reliable, these are the ability to build around the data middle platform. Industry 4.0 advocated by modern society is to intelligent and digitize industrial production, construct a new type of industrial structure system, manage it with data, make use of the powerful overall planning of data, and strengthen and improve business operation by computing function and big data prediction. In essence, the data middle platform is the data service factory of the enterprise, which completes the processing process from data to value.

### B. Construction of Data middle platform

The construction of a data middle platform mainly starts from four parts: data model, data assets, data governance, and data service [18]. First of all, the plan needs to be made

for the overall situation, and then select the technology type and evaluate the configuration of the cluster according to the classification of the filtered data.

#### (1)Data Model

The data model, that is, the model in the data warehouse develops the model hierarchically according to the data warehouse specification, realizes the standardization of data, and mostly uses dimensional modeling. In addition, some widely used data mining models can also be precipitated into the data middle platform. From the above, we can see that the models in the data middle platform have extensive interoperability.

Data modeling is generally divided into two steps: determining the approved fact table and analyzing the life cycle of the business is the hub of the business. In the definition of indicators, whether to cover all the indicators in this topic, to determine which indicators can be calculated by addition, subtraction, multiplication and division, and so on.

#### (2)Data Assets

We have built some models in the above data warehouse, but these models cannot be called assets, they can be called assets only after the data island has been opened up. Need to standardize the index library, these indicators can be combined to meet the personalized needs of external personnel. The basis of asset management is to do a good job of metadata management, including collected interface information, model information, index definition, job consanguinity, data storage and access.

#### (3)Data Governance

Data governance is not only the management of the data content itself but also the active management of the users, producers and owners of the data. Compared with data management, data governance has a larger scope and a relatively perfect system. Data governance ensures that the data of data assets are preserved intact, used accurately, updated timely, and consistently.

#### (4)Data Service

In the data middle platform system, giving consistent data services to the outside is an important part of data state establishment. Data services need to include the following capabilities: data interface standardization, providing a unified online query view of data services, allowing developers to access data services quickly and easily; data development visualization: providing visual configuration of service interfaces, developers only need to configure SQL to produce API, reducing the technical requirements of interface development, and facilitating maintenance and interface management. Give business analysts easy algorithm analysis, including model management, visual choreography process, algorithm model release.

### IV. ONLINE TRANSACTION PROCESSING AND ONLINE ANALYTICAL PROCESSING TECHNOLOGY

Online transaction processing class (OLTP) and online analytical processing class (OLAP) are two categories of data processing processes. With the development of data, these two services have been unable to meet the complex business needs

of the front end, so the data middle platform has integrated OLTP and OLAP, giving birth to the hybrid transaction analysis (HTAP) to provide faster, efficient and more accurate data processing for the foreground business. OLTP data is periodically synchronized into the OLAP system through ETL tools. The problem of data transmission delay will also arise. If the data obtained by OLAP labels, then the analysis results will lag accordingly without timeliness, so the support used as a basis for decision-making is not enough. OLAP needs the data generated by OLTP as the source, the data transfer program between two different databases is complex, and the time cost is high, so it needs mixed transactions to deal with and analyze, and schedule OLTP and OLAP tasks uniformly.

OLTP refers to highly transactional systems, generally highly available online systems, mainly small transactions and small queries. We usually evaluate the system by looking at the number of Transactions and Execute SQL execution per second [16]. In such a system, a single database often processes more than hundreds or thousands of Transactions per second, and the execution of Select statements is thousands or even tens of thousands per second. The typical online transaction processing systems in daily life are e-commerce systems, banks, securities. With the continuous enlargement of the market and the increasing competition in the industry, it can no longer meet the needs of enterprises to manipulate a large number of historical data in the OLTP database only through traditional OLTP analysis and reporting tools. Therefore, OLAP applications are established, which can enable decision-makers to make accurate and effective management decisions by making full use of these data.

OLAP is the core part of the data warehouse. The so-called data warehouse is an analytical database for a large amount of data formed by OLTP, which is used by decision-makers to deal with business intelligence, decision support and other important decision information. The data warehouse contains a lot of important data information, but it is difficult for the development decision maker to use the unanalyzed data directly, so before the decision-maker receives the data, the data needs to be processed by relevant steps.

### V. HTAP BUSINESS AND TECHNICAL KEY ANALYSIS

HTAP is a system that combines OLTP and OLAP business processing at the same time. In 2014, Gartner gave a strict definition of HTAP. The hybrid transaction and analytical processing (HTAP) is a new application architecture, which breaks the gap between online transaction processing and online transaction analysis as shown in Fig.2. Therefore, it supports more information and "real-time business" decisions. HTAP breaks the gap between online transaction processing and online transaction analysis, so it can switch back and forth between transactional database scenarios and analytical database scenarios to meet users' real-time business decisions. Compared with two independent systems, this architecture not only does not need very complex and expensive ETL operations but also can carry out real-time and efficient analysis and processing of the latest data. This ability to analyze data in



real-time and efficiently is a powerful weapon for enterprises to compete in the future. The whole key technology includes the following parts.

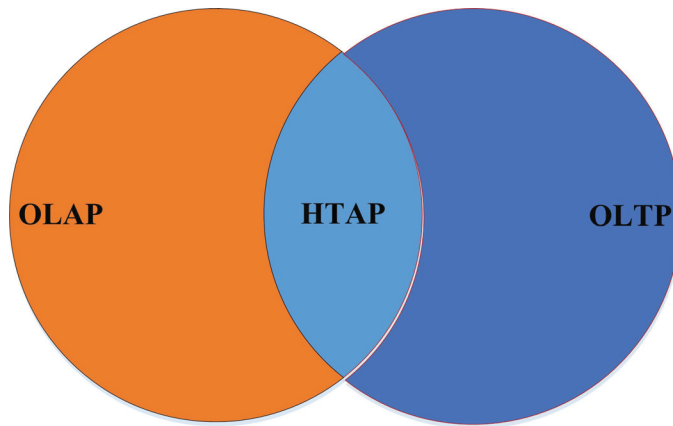


Fig. 2. HTAP structure.

#### (1) Row storage

In the face of traditional relational databases such as Oracle's Oracle DB, MySQL, IBM's DB2, Microsoft's SQL Server, etc., row storage (Row-based) is generally used [11]. In a database that uses row storage, data is stored in a logical storage unit based on row data, and the data in a row exists in a continuous storage form in the storage medium.

#### (2) Columnar storage

Relative to row storage, some emerging databases such as Hbase, HP Vertica, EMC Greenplum, and other distributed databases use column storage methods. In the database based on column storage, the data is stored according to the column-based logical storage unit, and in the storage medium, the data in a column exists in the form of continuous storage.

The traditional row database is stored according to rows. In the case of maintaining a large number of indexes and materialized views, the area cost of time processing and space storage is very high. Different from the row database, the data of the column database is stored according to the column, each column is stored separately, and the data is the index. Access to only the columns involved in the query greatly reduces the system I/O, processing one column per line, consistent data types and similar data characteristics also greatly reduce the process and time spent in compression.

#### (3) MPP

MPP (Massively Parallel Processing), that is, massively parallel processing, in the database non-shared cluster, each node has its own disk storage system and memory system. MPP distributes tasks to multiple servers and nodes in parallel, merges each node to complete the calculation of the distributed tasks, and finally connects the results on each node to get a complete result. Here is an example of a typical MPP product Greenplum architecture, as shown in Fig.3.

### VI. PROSPECTS AND CHALLENGES

Data Central platform is the best practice for the Internet to achieve data intelligence in the past 10 years, and it is the

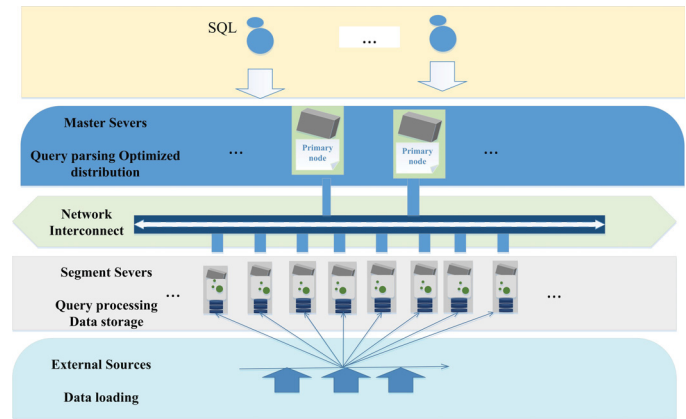


Fig. 3. Greenplum structure chart.

technical precipitation accumulated in the process of realizing Internet data intelligence, as well as big data's ability to continuously integrate data and exert value on the business side. The data middle platform can help enterprises solve the problem of non-synchronization between data development and application development, integrate OLTP and OLAP data processing, and improve the productivity of enterprises. Nowadays, the data middle platform focuses on real-time computing, the mobile terminal of upper application and intelligence.

(1) Data processing is gradually developing in the direction of a real-time trend, which combines real-time computing with historical data to realize the integrated development of streaming batches.

(2) BI insight analysis is the most important visualization method of data analysis in the data middle platform. In the field of digital BI, its terminal must consider the end-to-end adaptation, more in the way of digital indicator Kanban, rather than like the PC end to highlight the rich presentation effect and historical indicators.

(3) After building the data middle platform, enterprise users can accumulate rich data assets, which are the basis for the implementation of algorithms and artificial intelligence. The common artificial intelligence application scenarios for data users are sales volume forecast, traffic forecast, recommendation algorithm of the corresponding population and other related scenarios. These are scenarios that provide direct assistance to business decisions.

In the future, the data middle platform will gradually occupy the core position in the enterprise. However, data middle platform will also face the challenges of data middle platform construction, data quality assurance and the application of mixed thing analytical data management.

(1) The data middle platform is a platform with data processing capabilities, which can encapsulate the data capabilities of the enterprise into the platform and quickly provide the work used by the business foreground. Therefore, how to find and identify valuable business scenarios and maximize the value of the data middle platform is the biggest problem in the construction of the data middle platform.

(2) When the data middle platform adopts OGG data replication mode to realize the data access of the source-side business system, it will lead to the inconsistency between the data access and the source side. After a comprehensive analysis, there may be the following problems. Firstly, the maintenance of the source-side business system changes the data table structure, and the data middle platform does not receive immediate feedback for linkage processing, which leads to the interruption of data access. Secondly, part of the OGG configuration process of the source business system is still the responsibility of the source business system operation and maintenance manufacturers, and the data middle platform does not have the authority to monitor the sourcing process in real-time. Once a fault occurs in the configuration process, it is unable to locate the fault link quickly and accurately, which affects the quality of data access.

(3) The complexity of operation and maintenance becomes very complex due to the increasing number of HTAP database services and diverse technologies, and it will become difficult and complex to manage services and scheduling in distributed systems.

## VII. CONCLUSIONS

This paper summarizes some related concepts involved in the data middle platform, including data model, data assets, data governance, data services, traditional data processing OLTP and OLAP and mixed transaction analytical data management. This paper summarizes the research status of data middle platform application construction, online analytical processing, online transaction processing and analytical data management of mixed things at home and abroad. This paper deeply analyzes the business scale and business characteristics of OLTP and OLAP in each application scenario, focuses on combing and analyzing the technical difficulties in the application process of OLTP and OLAP in the application scenario, and explores the requirements and application methods of mixed transaction analysis (HTAP) data management. The future development trend of the data middle platform station and the technical difficulties faced by the data middle platform station at the present stage are summarized.

## ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their comments and constructive suggestions that have improved the paper. The subject is sponsored by the Science and Technology Project of State Grid Corporation of China (No.5700-202058480A-0-0-00).

## REFERENCES

- [1] Q. Gao, R. Ran, X. Liu, Y. Liu, P. Hu, M. Huang, and J. Qi, "Visualization analysis method of data assets hotspots in power enterprises based on data middle platform," in *2021 International Conference on Electronics, Circuits and Information Engineering (ECIE)*. IEEE, 2021, pp. 306–309.
- [2] H. Wu, L. Shen, X. Chen, Q. Yang, Y. Cai, M. Zeng, and H. Li, "Research on the application of data middle platform technology in integrated energy business system," in *2020 IEEE 3rd International Conference of Safe Production and Informatization (IICSPI)*. IEEE, 2020, pp. 316–319.
- [3] G. Lyu, P. Liu, Y. Lu, T. Wang, and X. Kang, "A data middle platform architecture based on microservice serving power grid business," in *2021 8th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2021 7th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)*. IEEE, 2021, pp. 219–224.
- [4] Y. Ding, J. Zhang, X. Wang, S. Li, W. Gao, and S. Hu, "Research on the relationship between model of middle platform and enterprise digital transformation in business field," in *2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*. IEEE, 2021, pp. 552–556.
- [5] Q. Wu, P. Tang, and M. Yang, "Data processing platform design and algorithm research of wearable sports physiological parameters detection based on medical internet of things," *Measurement*, vol. 165, p. 108172, 2020.
- [6] X. Lu, F. Su, H. Liu, W. Chen, and X. Cheng, "A unified olap/oltp big data processing framework in telecom industry," in *2016 16th International Symposium on Communications and Information Technologies (ISCIT)*. IEEE, 2016, pp. 290–295.
- [7] K. Zhang, M. Sadoghi, and H.-A. Jacobsen, "DI-store: A distributed hybrid oltp and olap data processing engine," in *2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2016, pp. 769–770.
- [8] G. Nabibayova, "About an application of olap-technology in decision making support systems," in *2011 5th International Conference on Application of Information and Communication Technologies (AICT)*. IEEE, 2011, pp. 1–4.
- [9] D. Wang, P. Cai, W. Qian, and A. Zhou, "Discriminative admission control for shared-everything database under mixed oltp workloads," in *2021 IEEE 37th International Conference on Data Engineering (ICDE)*. IEEE, 2021, pp. 780–791.
- [10] D. Wang and Q. Zhou, "A method of distributed on-line analytical processing of status monitoring big data of electric power equipment," *Proceedings of the CSEE*, no. 19, pp. 5111–5121, 2016.
- [11] F. Lin, M. Hu, Y. Jiang, and B. Ni, "Architecture and related techniques of a power dispatching data platform," *Automation of Electric Power Systems*, vol. 31, no. 1, pp. 61–64, 2007.
- [12] G. Chang, J. Hao, D. Liu, and H. Xiong, "Design of the integrated data platform in electric power control center based on multi-agent technology," *Automation of Electric Power Systems*, vol. 32, no. 1, pp. 85–89, 2008.
- [13] W. Du and L. Zhao, "A research of massive distributed high voltage electrical online monitoring data management platform based on cloud storage," in *2013 International Conference on Information Science and Cloud Computing Companion*. IEEE, 2013, pp. 330–334.
- [14] M. Francia, E. Gallinucci, M. Golfarelli, A. G. Leoni, S. Rizzi, and N. Santolini, "Making data platforms smarter with moses," *Future Generation Computer Systems*, vol. 125, pp. 299–313, 2021.
- [15] K. C. Lin, J. Z. Shyu, and K. Ding, "A cross-strait comparison of innovation policy under industry 4.0 and sustainability development transition," *Sustainability*, vol. 9, no. 5, p. 786, 2017.
- [16] G. Li, X. Zhou, J. Sun, X. Yu, H. Yuan, J. Liu, and Y. Han, "A survey of machine learning based database techniques," *Chinese Journal of Computers*, vol. 43, no. 11, pp. 2019–2049, 2020.