Annals of Emerging Technologies in Computing (AETiC)
Vol. 9, No. 2, 2025

Research Article

Investigation into the Integration of Intranet and Extranet Data in Data Centers

Zhenyu Gao[1,2,*], Jian Cao[1,2], Yuxiao Zhao[1,2], Guangrui Peng[1,2], Wei Huang[1,2], Jiajia Wu[1,2], Jing Lin[3] and Xueliang Guo[3]

[1]Nari Group Corporation, State Grid Electric Power Research Institute, China
[2]Nanjing NARI Digital Service Co. Ltd, China
daixiguxing@163.com; caojian@sgepri.sgcc.com.cn; zhaoyuxiao@sgepri.sgcc.com.cn; pengguangrui@sgepri.sgcc.com.cn; huangwei9@sgepri.sgcc.com.cn; wujiajia@sgepri.sgcc.com.cn
[3]Big Data Center of State Grid Corporation of China
jing1-lin@sgcc.com.cn; xueliang-guo@sgcc.com.cn
*Correspondence: daixiguxing@163.com

Received: 18th February 2025; Accepted: 27th March 2025; Published: 1st April 2025

Abstract: Research on data storage and its optimal utilization has been a crucial area of study for decades, with recent advancements significantly improving efficiency. However, integrating heterogeneous data sources in computing presents persistent challenges, including security vulnerabilities, excessive resource allocation, and prolonged processing times. The complexity of managing intranet and extranet data integration within enterprises further complicates seamless data utilization. Businesses primarily rely on two conventional approaches—SQL integration and service integration—to merge and access data efficiently. While these methods reduce operational bottlenecks and enhance data accessibility, they still face limitations in scalability, flexibility, and real-time data processing. To address these challenges, Calcite's versatile and adaptable architecture provides a powerful solution by enabling the independent creation and enhancement of SQL with complex structures and sophisticated logic. This research introduces a novel design framework for integrating "intranet and extranet data in a data center" through fusion computing. Unlike traditional models, the proposed architecture eliminates the need to transfer large primary datasets, focusing instead on generating computed results while maintaining rapid data querying at the terabyte scale within seconds. By leveraging fusion computing principles, this approach minimizes data redundancy, enhances computational efficiency, and strengthens security protocols. Additionally, it enables real-time analysis, reducing processing latency and improving decision-making capabilities for businesses and organizations managing vast data networks. The fusion computing framework also offers adaptability to dynamic data environments, ensuring it remains relevant for future technological advancements. Furthermore, the proposed framework fosters better interoperability between disparate data sources, allowing organizations to optimize their data storage and retrieval processes with minimal infrastructure modifications. The experimental results indicate that this model significantly outperforms conventional methods in terms of processing speed, security, and computational resource efficiency. The findings of this study contribute to the evolving landscape of data integration, offering practical implications for large-scale data centers, cloud computing services, and enterprise-level data management systems.

Keywords: *Apache Calcite; Cross-Domain Querying; Data Fusion; Data Security; Fusion Computing; Intranet-Extranet Integration; SQL Optimization*

---

1. Introduction

Safe and effective integration of intranet and extranet systems has become vital given the fast digitalization of corporate infrastructure, especially in sectors involving sensitive operational data. The capacity to do cross-domain data analysis without sacrificing security or adding latency is crucial as companies run across both internal departments and outside partners more and more. Current

Zhenyu Gao, Jian Cao, Yuxiao Zhao, Guangrui Peng, Wei Huang, Jiajia Wu, Jing Lin and Xueliang Guo, "Investigation into the Integration of Intranet and Extranet Data in Data Centers", *Annals of Emerging Technologies in Computing (AETiC)*, Print ISSN: 2516-0281, Online ISSN: 2516-029X, pp. 66-77, Vol. 9, No. 2, 1st April 2025, Published by International Association for Educators and Researchers (IAER), DOI: 10.33166/AETiC.2025.02.005, Available: http://aetic.theiaer.org/archive/v9/v9n2/p5.html.

technologies, however, sometimes struggle to manage the constraints placed by physically separated networks, which results in fragmented data silos and ineffective operations.

Within the framework of State Grid Corporation's digital development, several business lines have embraced different data storage solutions to satisfy different operational requirements. Although the great availability of storage technologies generates possibilities, their variations provide difficulties for consolidated data analysis. When trying joint searches across intranet and extranet domains, companies sometimes find limited data collaboration and trouble linking operational processes.

Although intranet and extranet systems differ in user roles, access control, and security mechanisms—as shown in Table 1—they must still be integrated to ensure unified enterprise data operations [1, 2]. The integration becomes increasingly complex as security requirements grow [3], especially in high-security environments where networks must remain physically separate [4, 5]. Existing solutions like QR code-based data exchange [6] only partially address these challenges.

Table 1. The features of Intranet and Extranet

Titles	Intranet	Extranet
User Information	Members of the specific firm's proprietary	Group of closely related firms Shared in closely trusted held circles
Access	Private	Semi-private
Security Mechanism	Firewall, encryption	Intelligent firewall, encryption, various document security standards

This study investigates a universal, fast, and secure joint query function for heterogeneous data sources across intranet and extranet environments, supported by a self-developed isolation device for information security. Although several products are available, none adequately support the application scenarios of cross-domain fusion computing in enterprise data centers, nor do they address the query inefficiencies caused by “data islands” created by different business lines.

2. Literature Review

The spread of many data sources as State Grid Informatization develops calls for hybrid computing solutions able to manage distributed, large-scale datasets from both Intranet and Extranet environments [6]. Under these conditions, fusion computing gets difficult and drives the creation of several tools, including Spark, Flink, and Presto. By storing intermediate data in memory, an open-source, memory-based engine called Spark accelerated computation speed, so lowering disk I/O and allowing non-shuffle operations via a Directed Acyclic Graph (DAG) model [7]. Through its DataStream API, Flink, a distributed computing engine built for stream and batch data processing, abstracted distributed data streams and supported sophisticated operations, enabling quasi-real-time processing [8]. Designed for interactive analytics, Presto [9] is a distributed SQL query engine that lets direct querying of data from many relational and non-relational databases without moving it to separate analytical systems. Although every tool addressed particular computing paradigms—Spark lacking full SQL support, Flink offering low-latency but resource-intensive performance, and Presto excelling in interactive querying but facing out-of-memory issues with large datasets—their emergence marked major progress in handling multi-source fusion computing [9].

Razis and Mitropoulos [10] presented a technology-oriented online loan core banking system that automated the entire loan life cycle, from creditworthiness assessment to repayment. Its interoperable design integrated external and internal banking elements, reducing errors and freeing resources for business improvement. Using the balanced scorecard method, the study evaluated the system's performance through predefined KPIs, confirming its effectiveness and value. To investigate main difficulties and solutions in big data integration, Rozony *et al.* [11] systematically reviewed 150 peer-reviewed publications. Important problems found were security, scalability, data quality, and semantic heterogeneity. Semantic alignment was often accomplished using ontology-based models, but they suffered from scalability. Machine learning shown promise in automating data quality and schema matching, although depending on high-quality training data. Although scalable integration was supported by distributed computing models like Hadoop and Spark, they also required large infrastructure. Blockchain assured safe integration options but lacked scalability; cloud platforms have given flexibility but also create privacy concerns. The review underlined the need of more research in unstructured data management, cross-domain harmonizing, and real-time integration. Bartolucci *et al*. [12] introduced a fusion-based quantum computation model for photonic systems, utilizing entangling measurements (fusions) on small entangled resource states for fault tolerance. This model manages probabilistic photonic gates and errors via quantum error correction, achieving higher error thresholds than previous methods, including tolerating 10.4% photon loss perfusion. The modular architecture requires less classical processing than traditional photonic quantum computing approaches. Using Attribute-Based Encryption (ABE), Veeresh and Parvathy [13] examined safe data sharing in cloud computing, noting its limited security but high computational cost and fine-grained control. They underlined the need for group cybersecurity initiatives and the part cyber insurance plays in reducing risks. They suggested encrypting content before cloud upload and a system that generates safe private keys for managed user access and control in order to improve data security and economy of cost. Olabanji [14] focused on integrating Cloud Computing with Python and SQL to enhance data security and automate control processes. The paper highlighted that while Cloud Computing transformed resource management, it also brought security issues, such as data breaches and unauthorized access. SQL was particularly crucial in automating control processes related to data security, including resource provisioning, access control, and database management. By integrating SQL, organizations could streamline these processes, improve data security, and reduce the likelihood of human error. Despite facing challenges like skill requirements and data privacy concerns, solutions such as Snowpark and Dataiku facilitated the integration, ultimately strengthening security measures and operational efficiency. Wang and Peng [15] examined the role of integrating intranet and extranet systems in enhancing knowledge management within manufacturing through fusion computing. They highlighted how new information technologies facilitate this integration, including big data, IoT, digital twins, and cyber-physical systems. Fusion computing enables seamless data exchange and processing across intranet and extranet platforms, improving knowledge acquisition and management. By leveraging this integration, organizations can efficiently collect and analyze data from internal and external sources, supporting the creation of intelligent, data-driven manufacturing environments. The integration of these technologies allows for smarter decision-making and more effective knowledge management, driving innovation in digital factories. Using a federated learning framework for cross-domain data analytics, Zhang *et al*. [16] suggested safe model training free from exposing sensitive data. Their privacy-preserving showed 12% improvement in model accuracy in trials. Among the difficulties, though, were system complexity and node-based overhead in communication.

Overall, the reviewed literature highlights a broad range of techniques and technologies addressing secure integration, analytics, and data interoperability in distributed environments. Common challenges include scalability limitations, privacy concerns, system complexity, and lack of real-time capability. Few studies propose unified frameworks that bridge heterogeneous networks while ensuring low-latency and high-security data exchange. Hence, there remains a clear research gap in designing lightweight, secure, and interoperable solutions that support dynamic cross-domain data querying in hybrid infrastructures. The present study addresses this gap by introducing a secure query mechanism supported by Calcite's optimization and a novel isolation protocol.

Combining heterogeneous data sources between physically separated intranet and extranet networks creates issues, including real-time query limits, high data transfer overhead, and secure cross-domain access. Current systems such as Spark, Flink, and Presto either rely on data movement or assume homogeneous networks, which creates security concerns and latency. Furthermore, confusing user access and application development is the absence of a consistent interface. This work presents a scalable, safe, efficient joint query mechanism based on a self-developed isolation device and Calcite's SQL optimization features, addressing these problems.

To guarantee safe cross-domain data access between intranet and extranet environments, the isolation layer of the system combines encryption, authentication, and access control systems. Although these characteristics are essential for data security and integrity, they could cause small latency during query execution, especially in cases of concurrent access on a large scale. But with asynchronous encryption techniques and lightweight, token-based authentication, performance loss was negligible in our present implementation. Designed to run apart from the core query engine, the isolation layer was meant to minimize coupling and clear processing bottlenecks. For terabytes-scale searches, the overhead is minimal; future work will concentrate on thorough benchmarking and additional optimization to assess performance under petabytes-scale workloads.

2.1. Comparison and Trade-offs of Using Apache Calcite

Because of its modular architecture, dynamic query planning features, and support of SQL parser and optimizer functionalities free of storage or execution engine requirements, Apache Calcite was selected for this work. Calcite is lightweight and more suitable for embedding into custom systems that demand flexible query optimization across heterogeneous data sources than frameworks such as Apache Spark SQL and Presto. Calcite lacks a built-in distributed execution engine, unlike Spark SQL or Presto, which would limit scalability in large-scale systems. Though it excels in managing vast datasets with distributed computing, Spark SQL adds system complexity and resource overhead. Though it offers great performance distributed SQL querying, Presto is more suited for batch analytics than real-time integration. Conversely, Calcite's simplicity and adaptability fit our system, which gives low-latency cross-domain query optimization top importance. To evaluate quantitative differences, future studies will compare Calcite's performance under like conditions against these engines.

2.2. Novelties and Contributions

Though a range of mature platforms—such as Apache Spark, Flink, and Presto—support distributed computing and cross-database querying—they are not well-suited to situations involving physically isolated intranet and extranet environments. Usually assuming a unified network environment or requiring significant data movement, these systems can cause security vulnerabilities, high resource consumption, or latency. Moreover, they lack natural ways to carry out safe collaborative searches without breaking internal data restrictions. This work offers a new fusion computing architecture that, without raw data transfer, enables fast, safe, flexible querying across intranet and extranet domains is enabled. Our method produces optimized SQL statements fit for the network context and returns only computed results for further aggregation by using a self-developed isolation device and Apache Calcite's pluggable query optimization framework. This lowers network load, greatly improves data privacy, and lets terabytes-scale searches in seconds. While preserving scalability and operational efficiency, the proposed framework closes a significant void in current systems by enabling real-time, safe data collaboration across logically and physically separated network domains.

3. The Proposed Method

The research aims to provide a universal, convenient, and rapid joint-query facility on heterogeneous data sources both within the Intranet and Extranet. Utilizing the company's in-house developed isolation device, specially tailored for information security networks, the approach provides secure, seamless, and efficient access to data from different network environments. The solution meets the increasing requirement for data integration within complex network structures, allowing users to access and analyze data from various platforms securely. Finally, the project facilitates better decision-making and operational effectiveness through the assurance of data being available, secure, and consolidated within isolated network spheres.

3.1. Key Features and Advantages of Apache Calcite for Cross-System Data Querying

Apache Calcite is an open-source dynamic data regulation approach licensed by the Apache Software Foundation and written in Java [17]. It encompasses many components that include a general database regulation system, but does not have key functions like storing and processing data; nevertheless, it is done by some specialized engines. Although the Apache Calcite architecture does not support data storage and processing, it has some practical features that are preferred over other engines compared to well-known engines, including Apache Hive, Drill, Storm, and Flink [18]. The upper-hand-specific features of Calcite are summarized as follows:

3.1.1. Open-Source Friendliness

An open-source framework written in Java helps be easily operated with many data processing engines written in Java or run in a Java Virtual machine (JVM)-based environment, particularly in the Hadoop ecosystem.

3.1.2. Multiple Data Frameworks

Calcite supports query enhancement and execution but not data storage and processing, so it is ideal for mediating between applications that utilize several data repositories and multiple data processing engines. Cross-system support: the Apache Calcite approach can run and boost queries across various query execution systems or engines and dataset back-ends such as Apache Spark, Hive, Flink, and Drill due to cross-system support.

3.1.3. Support for SQL and Its Extensions

Since many networks do not offer their query language, SQL, as an available query language, is preferred. Therefore, Calcite assists SQL dialects and extensions for those systems.

3.1.4. Reliability

Calcite is reliable since it includes a thorough test collection that can validate all elements of the system, as well as query optimizer rules and incorporation with back-end datasets.

3.2. Isolation Device for Information Security Network

At the boundary where Intranet and Extranet information conjugate, a protection measure of the "dual network isolation" is applied. The core database of a business's Extranet information is deployed in the Intranet, which can only be accessed through the isolation device of the information security network. The isolation device of an information security network is generally employed to function as a database proxy access by implementing Java Data Base Connectivity (SG-JDBC) network protocol that encrypts isolation, authenticates security, provides access control, and filters security-related SQL statements. It also supports external interactions of various databases, which improves the access efficacy of the business system to the database and meets the security protection demands caused by an explosive growth in the access to the internal database of a company's Extranet business system.

More specifically, the primary interactive function of the isolation device in information network security (logical type) is to provide database proxy access for business systems. The extranet business can only utilize the dedicated SG-JDBC driver to access the intranet database through a private security communication protocol. During the interaction process, the message on the Extranet side is completely discarded, and only the part of the payload data that needs to be stored is forwarded to the Intranet and stored in the database. Therefore, the service vulnerabilities of an Intranet are not experienced for the Extranet, which can realize high isolation robustness. Also, it supports two-way deployment. Generally, the isolation device of the information security network (logical type) is deployed in the headquarters, and provincial network companies utilize it. Figure 1 depicts the deployment structure of the isolation device of the information security network (logical type).

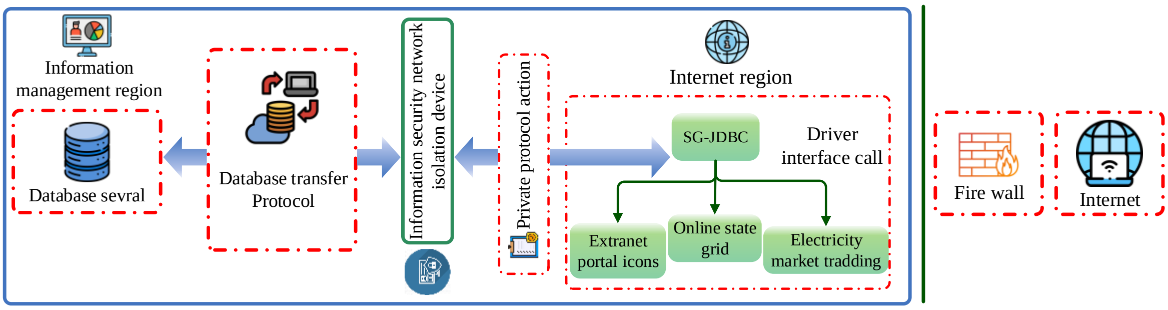

Figure 1. A deployment structure of isolation device in information security network (Logical Type)

3.3. A Design of the Overall Architecture

Based on a company's self-developed isolation device for the information security network, the research aims to provide a universal, convenient, and fast joint-query function for heterogeneous data sources on the Intranet and Extranet. However, the following four core problems remain to be resolved.

3.3.1. Access Method

The research considers the differences between different business systems and provides a relatively unified and universal interface integration method for external needs, thus reducing the difficulty of developing and applying related business systems.

3.3.2. Sensorless Call

The research aims to provide a joint Intranet and Extranet query service for a business without changing users' habits. Hence, users do not need to distinguish the region where the data source is located when it is used; they only need to focus on their business logic. Moreover, difficulties related to implementations will be significantly reduced.

3.3.3. Intranet and Extranet Interaction

Direct access is not possible due to the physical deployment of the isolation device of the information security network between the Intranet and Extranet. Thus, the research aims to design a logic that enables safe and reasonable Intranet and Extranet interactions.

3.3.4. Massive Data

TB or even PB-level data are shared in the Open Data Processing Service (ODPS) data in data centers. How to get fast association queries from massive data should be considered before running the research and its design.

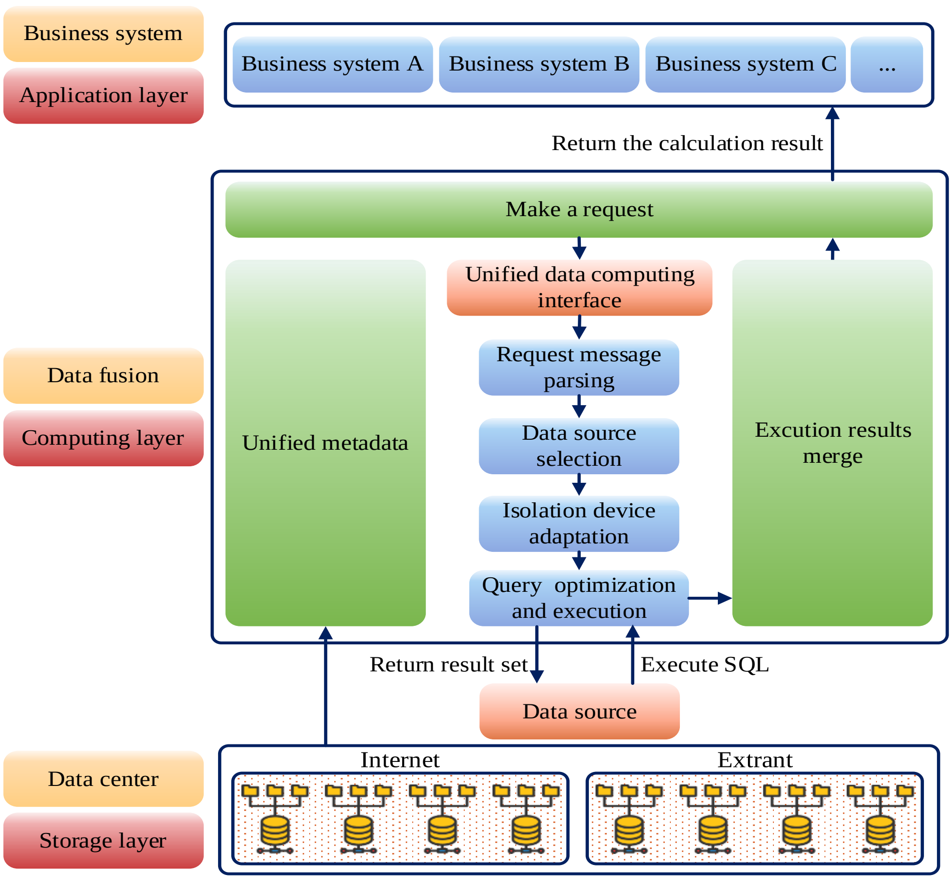

Figure 2. An architectural design of fusion computing based on a data center

To resolve those issues, the State Grid Big Data Center researches the empowerment of data fusion in data centers. Storage is commonly employed in a data center as the research object of ODPS. The manuscript intelligently proposes a design concept for the fusion computing of "Intranet and Extranet data in a data center." When multiple data from the Intranet and Extranet are combined in calculations, the proposed design will not move the original data but only share the calculated results and realize TB-level data query in seconds. Besides, it provides a new solution to promote the integration of intranet and extranet data in the data center, and data security is ensured.

Fusion computing of "Intranet and Extranet data in the data center" involves three layers from bottom to top: business system application, fusion computing of data, and data storage center. Figure 2 depicts a business system application that initiates a request in SQL or Restful through a unified interface module for data calculation. After a business request message is obtained, the layer in charge of computing data fusion combines the pre-registered data source information in the unified data source management module to distinguish different data sources from internal and external sides and parses them. After successfully parsing, two SQL statements containing only internal or external data sources are obtained and transmitted to their corresponding network regions through isolation devices. They are submitted to their respective data center for execution after optimization is run. The set containing results returned by the data center is harvested and transmitted to the region to which a business belongs through the isolation device for aggregation. The aggregated result completes the association and merges the results in a set-in memory, and finally, the calculations are returned to the business.

4. Implementation

Designed and tested in a data centre environment using a cluster of servers fitted with Intel Xeon processors, each with 128 GB RAM, and SSD-based storage arrays to guarantee fast I/O operations. The proposed system has A 10 Gbps internal network backbone that helps to support effective data distribution between nodes. Low latency and stable performance in this arrangement help terabytes-level data processing. Performance bottlenecks may thus develop when handling petabyte-scale data or when the result sets exceed available memory, since the system now depends on in-memory computation for generating result sets. Under such conditions, memory saturation can lead to slower query answers or task failures.

4.1. A Brief Explanation of the Calcite Framework's Competencies

The research is implemented based on the Calcite framework, which mainly includes seven modules: The JDBC Driver module supports applications employing the JDBC client. The SQL Parser and Validator module is responsible for SQL parsing and validation. Expressions Builder works for its own SQL parsing and validation docking framework [19]. The operator Expressions module helps process relational expressions. Metadata Providers are employed to support external custom metadata. Pluggable Rules define optimization rules. The Query Optimizer, the core, focuses on query optimization. Figure 3 depicts the specific architecture.

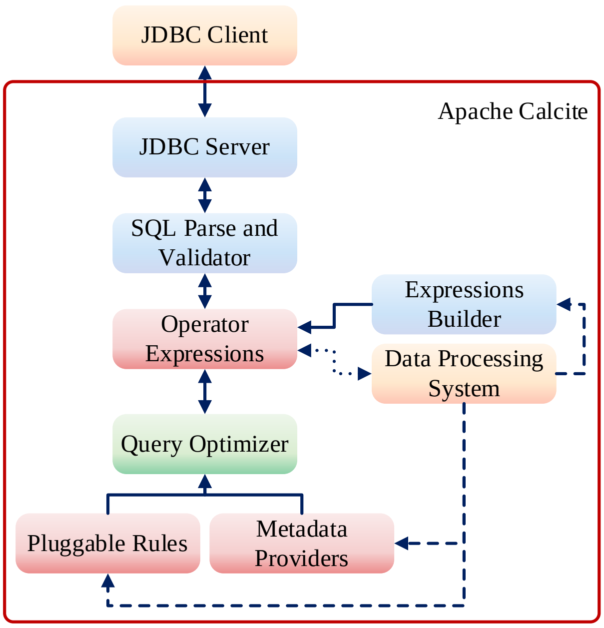

Figure 3. An architecture diagram of a Calcite component

A Calcite is designed to be flexible and pluggable, and each module complies with customization. Therefore, the research mainly implements a customized development of Intranet and Extranet data based on Calcite architecture.

The adaptation and expansion of multiple data sources, batch and stream integration, and other functions are realized by inheriting and rewriting the interface of the data source adapter when the internal and external ODPS data sources of business-related query data centers, SQL optimizer interface, and planner rule interface are implemented as an example. Due to the unique network environment of the Intranet and Extranet, the Intranet and Extranet are physically isolated, so the customized SQL rules are required to be devised based on the isolation device principle of a secure network to generate SQL optimization statements that conform to the unique network environment of the Intranet and Extranet. The optimized SQL statements will be executed on their respective data sources, and the execution results will be uniformly collected in the memory of the computing layer of the data fusion for association, merging, and other types of calculations. Then, the calculated results will be returned to the business system. The key technologies are described as follows:

4.1.1. The Adaptation of the ODPS Data Source

Currently, Calcite only supports some familiar data sources, such as MySQL, Oracle, and Postgres. The research is underway to support the adaptation of ODPS data sources [20]. According to ANSI 2003 syntax, general syntax specifications are developed. For special syntax in ODPS data sources, it is extended in the form of Parser plug-ins. The library table definitions of different databases can be flexibly converted into the platform structure, which can also perform conversion definitions between different data types (incredibly complex data types), thereby supporting the ODPS dialect by employing the technologies of library table mapping and field type mapping. **SQL Parsing:** SQL Parsing is mainly used for parameter verification and format conversion of business data in a request. When the business initiates a request, the SQL parsing module obtains the SQL passed by a company, parses it according to the registration information of the data source, and splits it into two SQL statements for the Intranet and the Extranet. Each SQL only involves the data source of the region where it is located. The segmented SQL statements are inserted into the middle database and pooled by the Intranet and the Extranet computing engines before execution.

4.2. Illustration

To illustrate, a query of the ODPS database and its parsing process is regulated as follows:

$$\begin{aligned}&\text{SELECT u.id, name, age, sum(price) FROM users AS u}\\&\text{Join orders AS o ON u.id} = \text{o.user\_id}\\&\text{WHERE age} >= 20 \text{ AND age} <= 30\\&\text{ GROUP BY u.id, name, age ORDER BY u.id}\end{aligned} \tag{1}$$

The users' table is in the central management information area of the ODPS of the data, and the table of orders is in the central Internet area of the ODPS of the data. The parsed SQL can be divided into the following two SQL statements, namely, Expressions (2) and (3), as follows:

The SQL must be passed to the management information area to execute.

$$\text{SELECT id, name, age FROM users WHERE age } >= 20 \text{ AND age} <= 30 \tag{2}$$

SQL that needs to be executed in the Internet area:

$$\text{SELECT price, user\_id FROM orders} \tag{3}$$

5. SQL Optimization

The optimizer included in Calcite can support both the system and SQL. If the system's SQL is weak, as if it were ODPS, the ODPS optimizer grows weak even though it supports SQL. After ODPS parses SQL into a syntax tree, Calcite can restore it to complete SQL statements and optimize.

Before passing any query, the optimizer module knows which models and tables are in the query planner. The SQL optimization process includes identifying the operations in SQL requests, the corresponding database tables, and the comprehensive identification of SQL request statements. The Calcite optimizer module adopts its own lexical and grammatical analysis to pinpoint comprehensively and parse SQL statements and pinpoint operators and predicates when SQL requests and trigger rules are cyclically implemented. When an expression that any process can no longer adjust is generated, it is terminated [21].

Expression (4) presents an example of querying data by conducting a join on the internal Splunk ODPS table and the products of an internal MySQL table.

SELECT p.'product_name,' COUNT(*) AS c FROM 'ODPS.' 'Splunk' AS s
JOIN 'MySQL.'' products' AS p ON s.'product_id' = p.'product_id.'
WHERE s.'action' = 'purchase' GROUP BY p.'product_name'
ORDER BY c DESC (4)

Figure 4 depicts the execution process before SQL optimization is run.

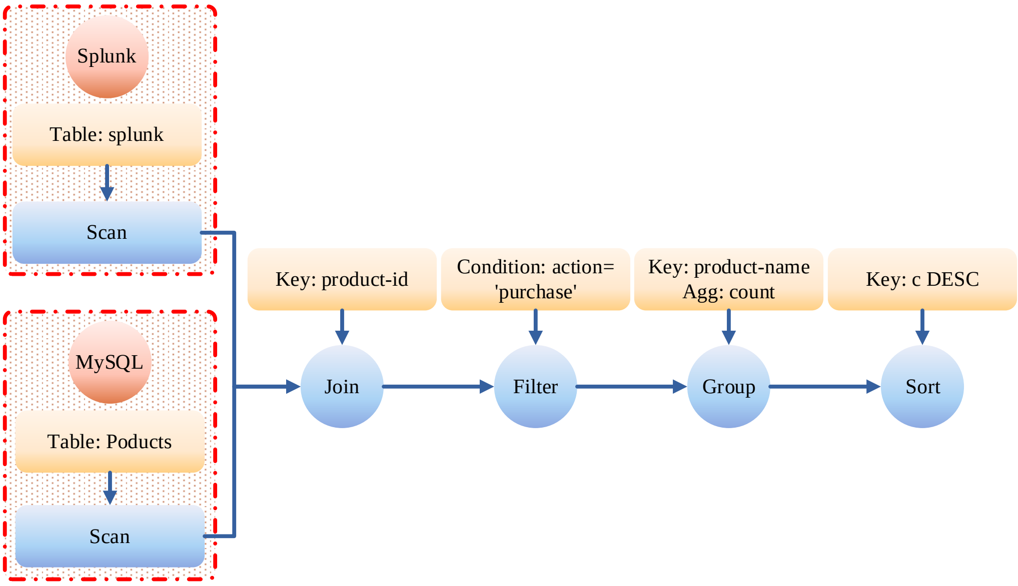

Figure 4. An execution process before running the optimization

The query engine module assigns the query conditions after optimization. Then, the query performance is optimized, and the query results become more consistent. Figure 5 depicts the optimized execution process.

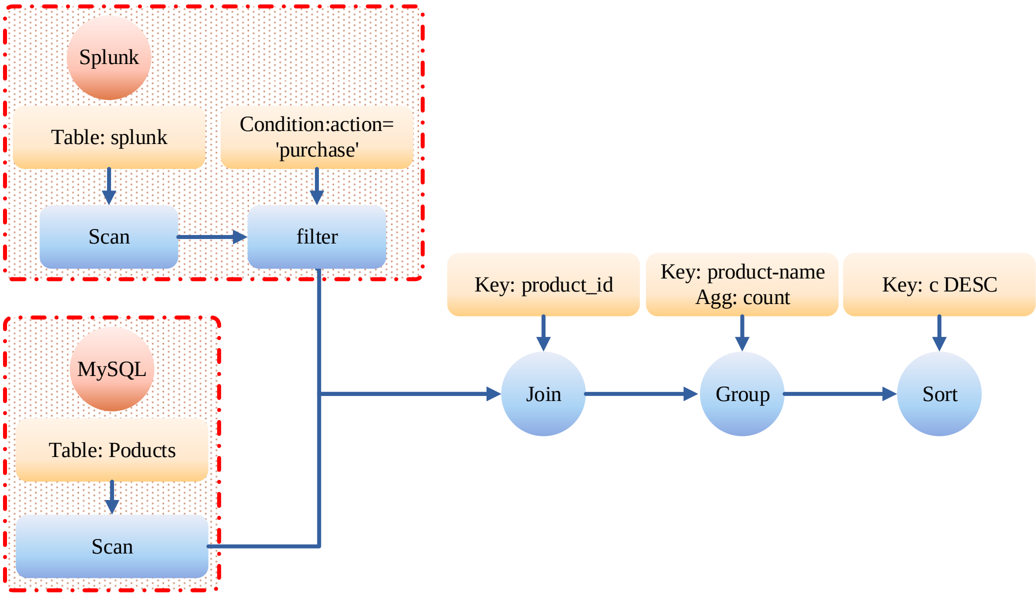

Figure 5. The flow of optimized execution

6. SQL Execution and Merged Result Sets

The optimized SQL statements for the Intranet and the Extranet will be submitted to the ODPS database in the data center. Then, the execution results will be returned to the memory of the respective regional nodes. The area where the business initiates requests determines whether the Intranet or the Extranet node merges the result sets. If a request is issued on the Extranet, the result set obtained on the Intranet is transmitted to the memory of the Extranet node in the form of a stream through the isolation device. Then, it is associated and merged with the result set attained in the memory of the Extranet. A process of fusion computing is shown in Figure 6.

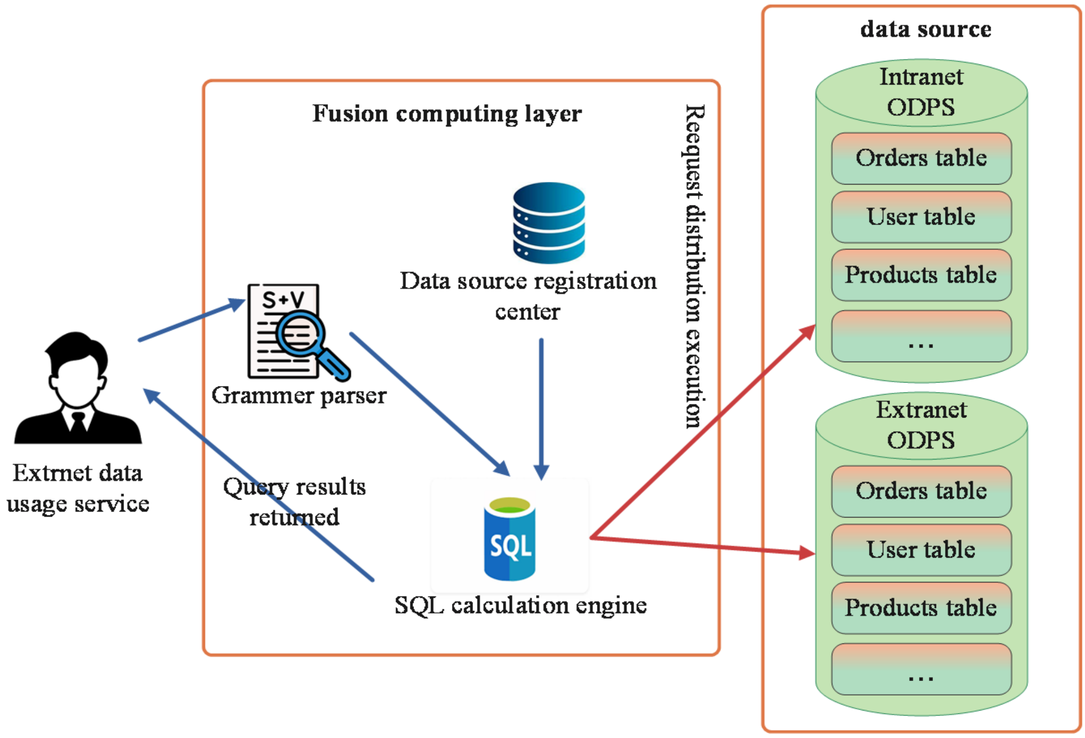

Figure 6. A process of fusion computing

This article focuses on improving data management with separate intranet and extranet platforms in business settings. Companies can enhance the efficiency and security of data integration from internal and external sources by adopting the suggested fusion computing architecture. This system mitigates network congestion, minimizes resource utilization, streamlines data processing, and facilitates expedited data retrieval. Moreover, Apache Calcite's adaptable framework permits the creation of intricate SQL procedures customized to particular business requirements, empowering firms to manage extensive, diverse data more efficiently and securely.

Although the suggested fusion computing architecture presents notable advancements in integrating intranet and extranet data, it is essential to acknowledge certain limits. A drawback is a dependence on Apache Calcite, which, while adaptable, may not accommodate all varieties of data sources or specific query needs without significant customization. Moreover, the system's ability to effectively process massive data sets, particularly those surpassing terabyte scales, could be limited by the hardware and network capabilities accessible in particular data center locations. Another constraint is the intricacy of setting up and managing the isolation device, necessitating specialist expertise to guarantee adequate security and performance. Subsequent studies should focus on overcoming these restrictions by investigating methods to enhance interoperability with various data sources, maximize the consumption of hardware and network resources, and streamline the configuration and maintenance procedures to encourage wider adoption.

7. Conclusion

As the State Grid Corporation undergoes its digital transformation, it is developing new business lines that require different data storage methods to accommodate diverse storage processes. The convenient accessibility and wide availability of various storage sources offer fresh prospects for companies. Nevertheless, several storage formats possess unique attributes that can provide specific benefits or drawbacks, contingent upon the situation. This article aims to streamline the intricacies associated with the analysis of intranet and extranet data, resulting in a decrease in network load and resource utilization. With SQL integration and service integration, enterprises can effectively reduce the obstacles to concurrent merging and utilizing intranet and extranet data. Calcite's versatile and adaptable architecture enables the independent creation and extension of SQL, allowing intricate and specific logic. The paper presents a design concept for fusion computing of "intranet and extranet data in a data center." The main focus is on sharing computed results instead of transferring the original data. This approach allows for terabyte-level data searches to be executed within seconds. The research focuses primarily on the ODPS database in data centers, with future intentions to improve the data source's support capabilities. Although the data source can process operations at the terabyte level, the existing system can only handle tiny query result sets due to memory computation limitations, which restrict the size of the data results. In the future, further research will be conducted to overcome this issue and provide support for more extensive data queries. Furthermore, investigated will be performance benchmarking of Apache Calcite against other SQL optimization engines, including Apache Spark SQL and Presto to better evaluate trade-offs in scalability, flexibility, and execution efficiency.

CRediT Author Contribution Statement

Zhenyu Gao: Conceptualization, Methodology; Jian Cao: Validation, Formal Analysis; Yuxiao Zhao: Investigation; Guangrui Peng: Writing—Original Draft; Wei Huang: Project Administration; Jiajia Wu: Investigation; Jing Lin: Supervision, and Xueliang Guo: Writing—Review & Editing.

References

[1] Matthew N. O. Sadiku and Cajetan M. Akujuobi, “Intranets and Extranets”, In Fundamentals of Computer Networks, 1st ed. Singapore: Springer Nature, 30th August 2022, ch. 6, pp. 71–77, Print ISBN: 978-3-031-09416-3, Online ISBN: 978-3-031-09417-0, DOI: 10.1007/978-3-031-09417-0_6, Available: https://link.springer.com/chapter/10.1007/978-3-031-09417-0_6.

[2] Gerasimos Razis and Sarandis Mitropoulos, “An integrated approach for the banking intranet/extranet information systems: the interoperability case”, *International Journal of Business and Systems Research*, Print ISSN: 1751-200X, Online ISSN: 1751-2018, pp. 330–357, Vol. 16, No. 3, 4th March 2022, Published by Inderscience, DOI: 10.1504/IJBSR.2022.122531, Available: https://www.inderscienceonline.com/doi/abs/10.1504/IJBSR.2022.122531.

[3] Steffen Herbold, Alexander Trautsch, Fabian Trautsch and Benjamin Ledel, “Problems with SZZ and features: An empirical study of the state of practice of defect prediction data collection”, *Empirical Software Engineering*, Print ISSN: 1382-3256, Online ISSN: 1573-7616, p. 42, Vol. 27, No. 2, 18th January 2022, Published by Springer Nature, DOI: 10.1007/s10664-021-10092-4, Available: https://link.springer.com/article/10.1007/s10664-021-10092-4.

[4] Kornel Terplan, Eds., *Intranet performance management*, 1st ed. New York, USA: Taylor & Francis, 2022, Available: https://www.taylorfrancis.com/books/mono/10.1201/9780203750254/intranet-performance-management-saba-zamir-kornel-terplan.

[5] Bo Ma, Zanyou Su, Kuncheng Zhuo, Meng Li, Donghai Huang *et al.*, “Research and Development of Enterprise Data Security Transmission Management System”, In *Proceedings of the 2021 IEEE International Conference on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)*, 30th September-3rd October 2021, New York City, NY, USA, Online ISBN: 978-1-6654-3574-1, DOI: 10.1109/ISPA-BDCloud-SocialCom-SustainCom52081.2021.00230, pp. 1710–1716, Published by IEEE, Available: https://ieeexplore.ieee.org/abstract/document/9644846.

[6] Xin Jin, Fengyi Li, Xiaodong Li, Weihan Tian, Biao Wang *et al.*, “Visual-based data exchange system for internal and external networks in physical isolation”, *Cognitive Robotics*, Online ISSN: 2667-2413, pp. 134–144, Vol. 1, January 2021, Published by ScienceDirect, DOI: 10.1016/j.cogr.2021.08.002, Available: https://www.sciencedirect.com/science/article/pii/S2667241321000136.

[7] Ersheng Pan, Siwei Liu, Jianqin Liu, Qingru Qi and Zun Guo, “The state grid corporation of China’s practice and outlook for promoting new energy development”, *Energy Conversion and Economics*, Online ISSN: 2634-1581, pp. 71–80, Vol. 1, No. 2, 12th November 2020, Published by The Institution of Engineering and Technology (IET), DOI: 10.1049/enc2.12007, Available: https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/enc2.12007.

[8] Shanjiang Tang, Bingsheng He, Ce Yu, Yusen Li and Kun Li, “A survey on spark ecosystem: Big data processing infrastructure, machine learning, and applications”, *IEEE Transactions on Knowledge and Data Engineering*, Print ISSN: 1041-4347, Online ISSN: 1558-2191, pp. 71–91, Vol. 34, No. 1, 24th February 2020, Published by IEEE, DOI: 10.1109/TKDE.2020.2975652, Available: https://ieeexplore.ieee.org/abstract/document/9007378.

[9] Rong Gu, Han Yin, Weichang Zhong, Chunfeng Yuan and Yihua Huang, "Meces: Latency-efficient rescaling via prioritized state migration for stateful distributed stream processing systems", In *Proceedings of the 2022 USENIX Annual Technical Conference*, 11-13th July 2022, Carlsbad, CA, USA, Online ISBN: 978-1-939133-29-13, pp. 539–556, Published by USENIX ATC 22, Available: https://www.usenix.org/conference/atc22/presentation/gu-rong.

[10] Gerasimos Razis and Sarandis Mitropoulos, "An integrated approach for the banking intranet/extranet information systems: the interoperability case", *International Journal of Business and Systems Research*, Print ISSN: 1751-200X, Online ISSN: 1751-2018, pp. 330–357, Vol. 16, No. 3, 4th March 2022, Published by Inderscience, DOI: 10.1504/IJBSR.2022.122531, Available: https://www.inderscienceonline.com/doi/abs/10.1504/IJBSR.2022.122531.

[11] Farhana Z. Rozony, Mst N. A. Aktar, Md Ashrafuzzaman and Ashraful Islam, "A Systematic Review of Big Data Integration Challenges and Solutions for Heterogeneous Data Sources", *Academic Journal on Business Administration, Innovation & Sustainability*, Online ISSN: 2997-9552, pp. 1–18, Vol. 4, No. 04, 2nd October 2024, Published by All Academic Research, DOI: 10.69593/ajbais.v4i04.111, Available: https://allacademicresearch.com/index.php/AJBAIS/article/view/111.

[12] Sara Bartolucci, Patrick Birchall, Hector Bombin, Hugo Cable, Chris Dawson *et al.*, "Fusion-based quantum computation", *Nature Communications*, Online ISSN: 2041-1723, p. 912, Vol. 14, No. 1, 17th February 2023, Published by Nature, DOI: 10.1038/s41467-023-36493-1, Available: https://www.nature.com/articles/s41467-023-36493-1.

[13] Veer Veeresh and Rama L. Parvathy, "Python-Based Free and Open-Source Web Framework Implements Data Privacy in Cloud Computing", In *Proceedings of the International Conference on IoT Based Control Networks and Intelligent Systems*, 28th November 2023, Singapore, Print ISBN: 978-981-99-6585-4, Online ISBN: 978-981-99-6586-1, pp. 717–729, Published by Springer Nature, DOI: 10.1007/978-981-99-6586-1_48, Available: https://link.springer.com/chapter/10.1007/978-981-99-6586-1_48.

[14] Samuel O. Olabanji, "Advancing cloud technology security: Leveraging high-level coding languages like Python and SQL for strengthening security systems and automating top control processes", *Journal of Scientific Research and Reports*, Online ISSN: 2320-0227, pp. 42–54, Vol. 29, No. 9, 13th September 2023, DOI: 10.9734/JSRR/2023/v29i91783, Available: https://journaljsrr.com/index.php/JSRR/article/view/1783.

[15] Hongwei Wang and Gongzhuang Peng, "The Merging of Knowledge Management and New Information Technologies", In Collaborative Knowledge Management Through Product Lifecycle: A Computational Perspective, 1st ed. Singapore: Springer Nature, March 2023, ch. 6, pp. 229–283, Print ISBN: 978-981-19-9625-2, Online ISBN: 978-981-19-9626-9, DOI: 10.1007/978-981-19-9626-9_6, Available: https://link.springer.com/chapter/10.1007/978-981-19-9626-9_6.

[16] Yiwei Zhang, Jie Liu, Jiawei Wang, Lu Dai, Fan Guo *et al.*, "Federated Learning for Cross-Domain Data Privacy: A Distributed Approach to Secure Collaboration", *arXiv*, arXiv:2504.00282, Online ISSN: 2331-8422, 31st March 2025, Published by arXiv, DOI: 10.48550/arXiv.2504.00282, Available: https://arxiv.org/abs/2504.00282.

[17] Mark Dodds and Khuzaima Daudjee, "Apache Ignite+ Calcite Composable Database System: Experimental Evaluation and Analysis", *Computer Science Bibliography*, Online ISSN: 2367-2005, 10th March 2025, Published by Open Proceedings, DOI: 10.48786/edbt.2025.77, Available: https://dblp.org/rec/conf/edbt/DoddsD25.html.

[18] Yupeng Fu and Chinmay Soman, "Real-time data infrastructure at uber", In *Proceedings of the 2021 International Conference on Management of Data*, 18th June 2021, New York, NY, United States, Online ISBN: 978-1-4503-8343-1, pp. 2503–2516, Published by SIGMOD, DOI: 10.1145/3448016.3457552, Available: https://dl.acm.org/doi/abs/10.1145/3448016.3457552.

[19] Xiaohan Li, Bowen Yu, Guanyu Feng, Haojie Wang and Wenguang Chen, "LotusSQL: SQL engine for high-performance big data systems", *Big Data Mining and Analytics*, Print ISSN: 2096-0654, Online ISSN: 2097-406X, pp. 252–265, Vol. 4, No. 4, 26th August 2021, Published by YUP, DOI: 10.26599/BDMA.2021.9020009, Available: https://ieeexplore.ieee.org/abstract/document/9523499.

[20] Maiara G. Flausino, Nádia P. Kozievitch, Keiko V. O. Fonseca and Eunice Liu, "Open Data Portals-A case study of Challenges and Opportunities", In *Proceedings of the Brazilian e-Science Workshop (BreSci 2024)*, 14th October 2024, Florianópolis, Brazil, Online ISSN: 2763-8774, pp. 104–111, DOI: 10.5753/bresci.2024.244096, Available: https://sol.sbc.org.br/index.php/bresci/article/view/30587.

[21] Md M. Rahman, Siful Islam, Md Kamruzzaman and Zihad H. Joy, "Advanced Query Optimization in SQL Databases for Real-Time Big Data Analytics", A*cademic Journal on Business Administration, Innovation & Sustainability*, Online ISSN: 2997-9552, pp. 1–14, Vol. 4, No. 3, 24th June 2024, Published by All Academic Research, DOI: 10.69593/ajbais.v4i3.77, Available: https://allacademicresearch.com/index.php/AJBAIS/article/view/77.