

A Blockchain-based Transparent Data Ecosystem for Cross-regional Economic Collaboration: Architecture and Scalability Analysis

Xuefeng Yang¹, Weizu Lu^{1*}

¹Baiyin Highway Development Center of Gansu Province, Baiyin 730900, Gansu, China

Abstract

INTRODUCTION: The rapid development of the digital economy has intensified the demand for efficient and trustworthy cross-regional data sharing, yet current systems suffer from data silos, trust deficits, and scalability limitations, constraining the circulation of data elements across regional boundaries. **OBJECTIVES:** This study proposes a blockchain-based transparent data ecosystem architecture that addresses trust establishment and intelligent analytics requirements, providing technical infrastructure for data-driven cross-regional economic decision-making in the digital economy context. **METHODS:** The architecture comprised four layers: a blockchain layer employing hierarchical grouped Practical Byzantine Fault Tolerance (PBFT) consensus to reduce communication complexity from $O(n^2)$ to $O(n\sqrt{n})$, a smart contract layer implementing risk-aware access control policies, a data storage layer combining on-chain hash indexing with off-chain distributed storage, and a data analytics layer utilizing Graph Attention Networks (GAT) for supply chain prediction. Experimental evaluation was conducted on Hyperledger Fabric for scalability testing, with the SupplyGraph benchmark dataset for analytics validation. **RESULTS:** The proposed consensus optimization achieves 453 TPS at 100 nodes, a 206% improvement over standard PBFT, with communication overhead reduced by 90%. GAT reduces prediction error by 36.95% compared to baseline methods, with attention weight analysis confirming meaningful association patterns aligned with supply chain domain knowledge. **CONCLUSION:** The proposed architecture integrates trustworthy data sharing with intelligent analytics within a unified framework. The hierarchical consensus design aligns with governance structures inherent in regional economic integration, while embedding analytics within the blockchain environment enables auditable analysis processes. The results demonstrate that the architecture meets the performance requirements of large-scale cross-regional transaction systems in the digital economy environment, facilitating efficient data element circulation across administrative boundaries.

Keywords: Blockchain, Cross-regional Collaboration, Scalable Consensus, Graph Attention Network, Supply Chain Analytics, Data Ecosystem

Received on 06 February 2026, accepted on 16 July 2026, published on 01 October 2026

Copyright © 2026 Xuefeng Yang *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](#), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetsis.15396

*Corresponding author. Email: Iwz2026@yeah.net

1. Introduction

The continued increase in globalization has enhanced cooperation between different regional economies, making

the exchange of data between upstream and downstream businesses and governments is crucial for informing

economic decisions. In the conventional regional economic integration projects, such as the Yangtze River Delta and the Guangdong-Hong Kong-Macao Greater Bay Area, cross-regional collaboration has created urgent demands for data sharing efficiency and reliability. The results of the empirical analysis based on panel data from 31 Chinese provinces for the period from 2007 to 2023 show that the application of blockchain technology in regional digital economy development is significant, and this effect is more noticeable in more open regions [1]. However, contemporary cross-regional data sharing practices face many challenges that affect the improvement of collaboration efficiency and the quality of decisions.

Data silos represent one of the main challenges to facilitating inter-regional collaborations. The presence of distinct data systems is a result of separate development across disparate administrative regions, industry sectors, and corporate entities. As a result, there is a limitation in the flow of useful economic information. While all parties involved recognize the benefits of data sharing, it is still important to deal with issues concerning risk aversion, privacy compliance, and accountability. [2]. This problem is worsened by the presence of trust deficits. Without trusted intermediaries, the data providers cannot guarantee that the data will not be misused or tampered with, and the data consumers cannot guarantee the authenticity and integrity of the data they receive. The existing centralized data exchange architecture does not meet the requirements of multi-party collaboration because of the risks associated with single-point failures and the lack of trust [3].

Insufficient scalability is one of the fundamental problems facing contemporary data-sharing systems. With an increase in the number of participating regional entities and data exchange operations, these systems need to handle large-scale concurrent requests. Scalability has been identified as one of the major bottlenecks facing the adoption of mainstream blockchain technology, characterized by low throughput, increased transactional latency, and storage costs [4]. Analysis from both intra-chain and cross-chain points of view shows that existing single solutions are unable to meet multiple requirements for performance, security, and decentralization [5]. Moving on to the consensus mechanism, the Practical Byzantine Fault Tolerance (PBFT) algorithm is used in consortium blockchain networks because of its property of deterministic finality. However, its $O(n^2)$ communication complexity makes it difficult to scale the system with an increasing number of nodes. Optimization methods proposed for overcoming this problem are node grouping, trust evaluation, and dynamic committee selection [6].

Blockchain technology, characterized by decentralization, immutability, and traceability, has been found to be an efficient way to address trust issues in data sharing between organizations. The analytical framework, which includes five dimensions—business, social, economic-financial, technological, and policy—helps to elucidate the mechanisms behind blockchain technology in various development domains. The analysis also indicates that developing countries can benefit from leapfrog

development through this technology [7]. For the domain of supply chain management, the major success factors that contribute to blockchain technology adoption include organizational, collaborative, technological, and governance issues [8]. As pointed out earlier, consortium blockchains have a middle ground between the other two types of blockchains. They have the benefit of providing data-sharing situations within organizations that incorporate both decentralization and access control. The use of smart contracts provides the technological basis for enforcing data access policies.

Moreover, from the perspective of data analytics, Graph Neural Networks (GNNs) offer powerful tools to model the intricate associations existing in economic networks. Notably, supply chain networks are inherently graph-based since companies or products act as nodes while transactions or collaborations serve as edges. GNNs can effectively capture implicit dependencies that cannot be easily captured using existing methods. Several studies have demonstrated that GNNs can accurately predict hidden connections between entities within the supply chain by identifying relationships that do not appear in the dataset but exist in reality [9]. The use of GNNs in predictive modeling improves the accuracy of forecasting and risk management by integrating information from neighboring entities. Multi-source heterogeneous data fusion is a prerequisite for regional economic analysis. Economic data from different regions is generally obtained from different departments and data systems, exhibiting significant differences in formats, granularities, and quality. Data heterogeneity, modeling, and uncertainty quantification are key challenges faced in current scenarios [10].

Although considerable progress has been made in blockchain technology-based data sharing, consensus optimization, and intelligent data analysis, several research gaps remain. In terms of architectural design, most existing works focus on optimizing each technical aspect in isolation and do not offer integrated frameworks for coupling blockchain-based trust building and deep learning-based analytical techniques. At the scalability level, further research is needed on consensus optimization strategies aimed at facilitating cross-regional economic collaboration in a blockchain consortium setting. At the data analytics level, existing GNN research literature mainly focuses on supply chain risk prediction and link prediction, whereas the literature on GNN and blockchain-based data sharing in support of regional economic collaboration remains limited.

Based on the above analysis, this study proposes a blockchain technology-enabled transparent data ecosystem architecture to support cross-regional economic collaboration. This data ecosystem architecture is designed to address the triple challenges of data silos, trust deficits, and scalability limitations concurrently while supporting regional economic decision-making with intelligent data analytics. The principal contributions of this study are summarized below. About architectural design, a layered system architecture is presented, in which the blockchain

layer, smart contract layer, data storage layer, and data analytics layer are integrated, enabling the entire functional chain from trustworthy data storage to intelligent analysis. In terms of scalability optimization, a consensus mechanism enhancement strategy based on hierarchical grouping and an on-chain/off-chain coordinated storage strategy is presented, achieving notable performance improvement under large-scale node conditions. Concerning intelligent analytics, the Graph Attention Network (GAT) is applied to solve the problem of collaboration prediction within the supply chain, thus justifying the application possibility of the data analytics layer. As for experimental validation, the scalability of the architecture is examined in terms of different numbers of nodes in the Hyperledger Fabric blockchain platform, while the practicality of the data analytics component is validated through an open-access supply chain benchmark dataset.

2. System Architecture and Scalability Design

2.1. Overview of System Architecture

The proposed architecture of the system uses a structure that is represented by four layers that are set out in hierarchical order. The layers are the blockchain layer, the smart contract layer, the data storage layer, and the data analytics layer. The blockchain layer acts as the base layer of trust, where the distributed ledger is maintained. The smart contract layer enables the enforcement of data access policies and collaboration rules through the execution of contracts. The data storage layer adopts the on-chain/off-chain coordinated approach for optimizing data storage efficiency while ensuring data integrity verification. The data analytics layer integrates heterogeneous data from various sources and enables in-depth modeling and predictive analysis of supply chain network relationships using GNN models.

As shown in Figure 1, a series of standardized interfaces is employed to connect these layers. Upon receipt of a data-sharing request from a regional participant, permission verification is performed by the smart contract layer. Subsequently, the data storage layer executes a corresponding operation, and the data are sent to the blockchain layer for confirmation. The data analytics layer periodically retrieves the updated data from the data storage layer, processes it with the help of data fusion, and then inputs the processed data into the GNN for analysis and prediction.

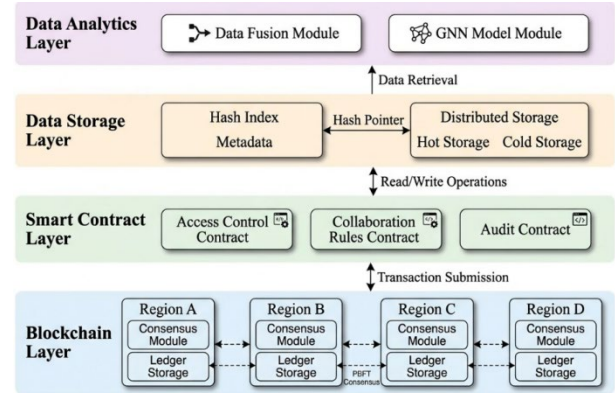


Figure 1. Overall architecture of the blockchain-based transparent data ecosystem

2.2. Blockchain Layer Design

Scenarios for cross-regional economic cooperation usually involve tens to hundreds of participating nodes, placing extremely high demands on transaction confirmation time. The PBFT algorithm has been identified as a popular solution for consortium blockchain environments due to its deterministic finality and relatively low resource requirements. However, its $O(n^2)$ communication complexity results in considerably reduced consensus efficiency as the number of nodes increases. To overcome this problem, various optimization techniques have been proposed. These techniques include grouping methods based on node feature-based trust evaluation [11] and node partitioning techniques based on spectral clustering [12].

A hierarchical grouping method is introduced in this study to improve consensus efficiency. The nodes are divided into multiple consensus groups based on their regional affiliation and business relevance. Then, the PBFT consensus algorithm is executed individually within these groups, followed by cross-group consensus among the representative nodes to achieve global consistency. The division of consensus groups considers geographic proximity, transaction frequency, and the historical collaborative relationships among the nodes. The association degree between nodes v_i and v_j is defined as:

$$w_{ij} = \alpha \cdot g_{ij} + \beta \cdot f_{ij} + \gamma \cdot h_{ij} \quad (1)$$

where g_{ij} , f_{ij} , and h_{ij} represent the geographic proximity score, normalized transaction frequency value, and historical trust degree, respectively. The weight coefficients satisfy $\alpha + \beta + \gamma = 1$. From the association degree matrix, the system utilizes spectral clustering to group the nodes into k consensus groups, where the number of nodes in the group is within a reasonable range in terms of the intra-group efficiency and the fault-tolerance ability of the system.

The primary node election for intra-group consensus incorporates a reputation mechanism to reduce the

probability of malicious nodes becoming primary nodes. The reputation value of the node v_i is dynamically updated based on historical consensus participation:

$$R_i^{(t+1)} = \lambda \cdot R_i^{(t)} + (1 - \lambda) \cdot p_i^{(t)} \quad (2)$$

where $R_i^{(t)}$ represents the reputation value after the t -th round of consensus, $p_i^{(t)}$ represents the behavior score for the current round considering message response timeliness and verification result correctness, and λ is the decay factor. The primary node is randomly selected from the top m candidate nodes ranked by reputation value, balancing both security and fairness.

During the cross-group consensus phase, representative nodes from each group form an upper-level consensus committee that executes a simplified PBFT process. Given a system with n nodes partitioned into k consensus groups, the improved overall communication complexity is $O(n^2 / k + k^2)$, which reaches the optimal value of $O(n\sqrt{n})$ when $k = \sqrt{n}$. Table 1 compares the theoretical communication complexity of the conventional PBFT and the enhanced approach with respect to different scales of nodes, and it is evident that the benefit of the enhanced approach increases with the scale of the nodes.

Table 1. Theoretical communication complexity comparison between traditional PBFT and the improved scheme

Number of Nodes (n)	Traditional PBFT (n^2)	Improved Scheme ($n^2 / k + k^2$, $k = \sqrt{n}$)	Reduction Ratio
64	4,096	576	85.9%
100	10,000	1,100	89.0%
256	65,536	4,352	93.4%
400	160,000	8,400	94.8%

2.3. Smart Contract Layer Design

The smart contract layer incorporates three basic contract types to enable data ecosystem operations. These are access control contracts, collaboration rules contracts, and audit contracts.

Access control contracts are responsible for managing data access permissions and identity verification. Building upon the conventional role-based permission verification approach, a risk-aware approach is proposed to dynamically manage access control. When an access request is met, the contract calculates the risk score for the current access:

$$r = \omega_1 \cdot r_{history} + \omega_2 \cdot r_{context} + \omega_3 \cdot r_{sensitivity} \quad (3)$$

where $r_{history}$ represents behavior risk based on historical access records, $r_{context}$ represents current contextual risk factors such as access time and device, and $r_{sensitivity}$ represents the sensitivity level of the target data.

The coefficients ω_1 , ω_2 , and ω_3 serve as weight parameters. If the risk score exceeds a certain threshold, then the contract denies access or requires further verification. This approach is based on various methodologies for risk quantification in healthcare data access control systems [13].

Collaboration rules contracts define the business logic of inter-regional data sharing in an event-driven system architecture. The contracts define three modes of collaboration. The data exchange model facilitates the sharing of data among peer agents by ensuring that both agents meet certain requirements before cooperation. The data aggregation model allows several data sources to feed data into a single data sink, and the agreement monitors the performance of all parties involved. Through the data subscription method, consumers can subscribe to specific types of data updates. Data transmission is automatically triggered once data meeting certain criteria becomes available for release.

These contracts use the indexing mechanism based on the hierarchy, which allows for multidimensional data lookup in terms of period, entity identification, and classification of resources. These contract designs use a modular structure, where the separation of logic and business parameters allows for changing parameters without re-deployment. They also use an index hierarchy system that allows searching in different dimensions, such as time frame, entity ID, and resource type. The contract structure has a modular design where logic and business parameters are separated; thus, business parameters can be changed without deploying the software again. The contract upgrade process uses the multisignature protocol, where the approval of more than two-thirds of the nodes is required for executing the contract upgrade process.

2.4. Data Storage Layer Design

The data storage layer uses a hybrid on-chain/off-chain mechanism to improve storage performance while ensuring the integrity of the data [14]. Specifically, the on-chain storage mechanism consists of storing hash values and indexes of the data, which requires little storage space but can provide sufficient assurance for integrity validation. The off-chain hot storage layer relies on a distributed file system for keeping the recently used data to facilitate read and write operations. The off-chain cold storage layer keeps the historical data using cheaper storage devices.

To implement data integrity verification, Merkle tree structures are used. The batches of data that are submitted are arranged as a Merkle tree, where leaf nodes represent the hash value of each data, whereas the root nodes are stored in the blockchain. The verification mechanism

works by using the Merkle proof from the leaf node to the root node, allowing recalculation of the root node's hash value. The type of Merkle tree used in this architecture supports dynamic updates, allowing the addition of new data items without rebuilding the tree entirely, with a time complexity of $O(\log n)$ [15]. The off-chain storage uses a content-addressed distributed file system in which data is uniquely identified and located based on content hash values, thus inherently supporting data deduplication. The system maintains data replicas across various geographically distributed storage nodes, with the number of replicas being dynamically adjusted based on data importance and access frequency. To meet the multi-dimensional query requirements in cross-regional collaboration, the system generates multi-level index information in the off-chain storage layer. The first-level index is organized by data source region, while the second-level index is organized by data type and time range. The index information is stored synchronously on-chain to ensure integrity.

2.5. Data Analytics Layer Design

The data analytics layer is the core innovation that distinguishes this architecture from traditional data sharing schemes based on blockchain technology. The data processing workflow includes two steps: data fusion and graph modeling. In the data fusion step, a feature-level data fusion approach is adopted. Given that the system receives inputs from M data sources, each data source m yields a unified-dimension feature vector $\mathbf{h}^{(m)} \in \mathbb{R}^d$ after passing through a feature extraction network. The fused feature $\mathbf{z}$ is computed through attention-weighted summation:

$$\mathbf{z} = \sum_{m=1}^M \alpha_m \mathbf{h}^{(m)}, \quad \alpha_m = \frac{\exp(f_{att}(\mathbf{h}^{(m)}))}{\sum_{j=1}^M \exp(f_{att}(\mathbf{h}^{(j)}))} \quad (4)$$

where f_{att} is a parameterized attention scoring function. This mechanism enables the model to adaptively learn the contribution of different data sources to specific analytical tasks.

The graph modeling phase constructs the fused data into a graph structure $G = (V, E, \mathbf{X}, \mathbf{A})$, where nodes can represent analytical objects at different granularities, such as regional economic entities, enterprises, or products, and edges represent economic associations or collaborative relationships between entities. This study adopts the GAT as the core analytical model, with the node feature update rule defined as:

$$\mathbf{h}_i^{(l)} = \sigma \left(\sum_{j \in \mathbf{N}(i)} \alpha_{ij}^{(l)} \mathbf{W}^{(l)} \mathbf{h}_j^{(l-1)} \right) \quad (5)$$

where $\mathbf{N}(i)$ represents the neighbor set of node i , $\mathbf{W}^{(l)}$ is a learnable weight matrix, and $\alpha_{ij}^{(l)}$ represents attention coefficients. In contrast to Graph Convolutional

Networks (GCN), where the weights of the aggregators are fixed, GAT can adaptively learn the weights of the neighboring nodes, which can improve the ability of the model to handle heterogeneous relations in the supply chain network.

GNNs have been widely applied in supply chain and economic network analysis. Empirical results suggest that GNN-based models considering inter-enterprise relationships significantly outperform basic models that do not account for structural relationships in supply chain risk forecasting [16, 17]. Abushaega et al. [18] used GNNs and federated learning to solve multi-objective optimization problems in supply chains, achieving a mean squared error of 0.198, thus providing a methodological reference for this study.

Model outputs vary depending on the type of analysis. For node-level analysis such as demand forecasting and risk assessment, predictions are made directly from the final-layer node features. On the other hand, for graph-level analysis such as regional economic condition prediction, node features are combined hierarchically to produce graph-level representations. The experiments use demand forecasting to validate the analytics layer.

3. Experiments and Analysis

3.1. Experimental Setup

The experimental design includes two parts: blockchain scalability assessment and GNN analytics layer performance evaluation. The blockchain experiments are designed to verify the effectiveness of the proposed hierarchical consensus strategy across different node scales, while the GNN experiments are intended to verify the effectiveness of the proposed data analytics layer in supply chain cooperative prediction tasks.

The experimental blockchain platform is based on Hyperledger Fabric 2.4. Hyperledger Fabric is a leading blockchain platform for enterprise consortium blockchain networks, featuring modular architecture and pluggable consensus and channel functionality [19,20]. The test environment is deployed on a cloud server cluster, where each blockchain node runs on a separate virtual machine instance with 4 CPU cores and 8 GB of memory. The network latency among nodes is set according to cross-regional collaboration scenarios, with intra-region latency set to 5-15 ms and inter-region latency set to 30-80 ms. The parameters are set based on typical network transmission characteristics in cities within the Yangtze River Delta region. Performance testing utilizes Hyperledger Caliper, a widely used blockchain benchmarking tool that supports customizable transactional workloads and multi-dimensional metric measurement. The test transaction types include data upload, data query, and permission modification operations, simulating typical operation patterns in cross-regional data sharing.

The GNN experiments are based on the SupplyGraph dataset [21], one of the few publicly available benchmarks for supply chain GNNs. The dataset is based on real-world operational data from a large fast-moving consumer goods enterprise and includes 40 product nodes and four types of relationships: product group, subgroup, production plant, and storage location. The node features include four temporal dimensions—production volume, sales orders, distribution delivery, and factory issues—across 221 time points, spanning from January to August 2023. Despite the small scale, the veracity of the data, coupled with its public availability, makes it a suitable benchmark dataset for supply chain GNN research. In the context of cross-regional economic collaboration, demand forecasting accuracy is essential for coordinating regional production plans. Therefore, demand forecasting is selected as a validation task for the GNN analytics layer. GNN models are implemented using PyTorch and PyTorch Geometric. The training process uses the Adam optimizer with a learning rate of 0.001 for 200 epochs. Table 2 presents a summary of key configuration parameters in the experimental environment.

Table 2. Experimental configuration

Component	Parameter	Value
Blockchain Platform	Version	Hyperledger Fabric 2.4
	Block Size	50 transactions
	Block Timeout	2 s
	Node Configuration	4 vCPU, 8GB RAM
	Intra-region Latency	5-15 ms
	Inter-region Latency	30-80 ms
GNN Module	Framework	PyTorch 2.0 + PyG 2.3
	Dataset	SupplyGraph
	Nodes / Edge Types	40/4
	Time Steps	221 (Jan-Aug 2023)
	Optimizer	Adam (lr=0.001)
	Training Epochs	200

3.2. Blockchain Scalability Evaluation

The scalability tests for blockchain measure performance along three dimensions: throughput, consensus latency, and communication overhead. Experiments were carried out for four node configurations: 16, 32, 64, and 100 nodes, each repeated five times and averaged to reduce random fluctuations. The control group used standard PBFT for consensus, while the experimental group used hierarchical grouped PBFT as proposed in this study. The selection of group count k follows two principles: first, ensuring that the number of nodes per group n/k is an integer to simplify implementation; second, approaching the theoretical

optimal value. $\sqrt{n}$ as closely as possible to balance intra-group and cross-group communication overhead.

Throughput is quantified as the number of committed transactions per second (TPS). The testing mechanism uses an open-loop loading mode, where transactions are dispatched at constant rates and the number of successfully confirmed transactions per unit time is measured. In a 16-node network, traditional PBFT achieves a throughput of 856 TPS, while the hierarchical approach achieves 894 TPS, showing only a small difference. This is because the low overhead associated with consensus communication in small-scale networks limits the potential benefits of hierarchical grouping. As the number of nodes increases, the gap between these two schemes increases. In a 32-node network, the standard PBFT protocol achieves 612 TPS, while the hierarchical scheme achieves 743 TPS, representing an improvement of 21.41%. In a 64-node network, the standard PBFT protocol decreases to 287 TPS, whereas the hierarchical approach reaches 594 TPS, a 106.97% improvement. In a 100-node network, the standard PBFT protocol reaches 148 TPS, whereas the hierarchical approach reaches 453 TPS, a 206.08% improvement. The degradation of the throughput capacity due to higher node numbers is inherent in the design of distributed consensus algorithms. However, the hierarchical scheme greatly mitigates this effect by limiting the number of active participants in each iteration of consensus formation. Figure 2 shows the throughput comparison curves for different node scales.

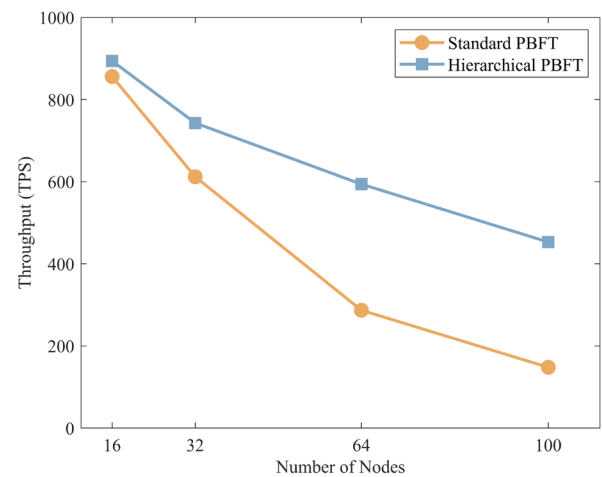


Figure 2. Throughput comparison under different node scales

Consensus latency is defined as the time interval from when a transaction is sent until it is confirmed. In a 64-node system, standard PBFT shows an average consensus latency of 1823 ms with a standard deviation of 287 ms, while the hierarchical approach shows an average latency of 694 ms with a standard deviation of 143 ms, representing a 61.93% reduction. The reduction in latency is mainly due to the hierarchical model used, reducing the number of

nodes involved in consensus in every round. For example, in a network comprising 64 nodes and 8 groups, intra-group consensus involves only 8 nodes, while cross-group consensus involves only 8 representative nodes. Unlike PBFT, where all 64 nodes take part in reaching consensus, this process considerably lowers the number of messages exchanged between senders and receivers.

The overhead of communications in terms of messages is computed using the number of messages that are sent during a single consensus operation. The conventional PBFT algorithm exhibits a complexity of $O(n^2)$ communications in terms of its prepare and commit phases, as each node broadcasts messages to all other $n-1$ nodes within both phases, resulting in a total number of $2n(n-1)$ messages. The total number of messages in the hierarchical approach comprises both intra-group and inter-group components. Given a system with n nodes partitioned into k consensus groups with n/k nodes per group, the intra-group two-phase message count is $k \times 2 \times (n/k) \times (n/k - 1)$, and the cross-group two-phase message count is $2 \times k \times (k - 1)$. Taking 64 nodes with $k = 8$ as an example, the intra-group message count is $8 \times 2 \times 8 \times 7 = 896$, the cross-group message count is $2 \times 8 \times 7 = 112$, totaling 1008 messages. This represents an 87.50% reduction compared to the standard scheme's $2 \times 64 \times 63 = 8064$ messages. Table 3 presents a comparison of performance metrics under different node scales.

Table 3. Performance comparison between standard and hierarchical PBFT

Nodes	Scheme	k	Throughput (TPS)	Latency (ms)	Messages per Consensus
16	Standard	—	856	423	480
16	Hierarchical	4	894	387	120
32	Standard	—	612	847	1984
32	Hierarchical	4	743	516	472
64	Standard	—	287	1823	8064
64	Hierarchical	8	594	694	1008
100	Standard	—	148	3156	19800
100	Hierarchical	10	453	1087	1980

The results show that the hierarchical clustering method gives significant gains when dealing with nodes in large numbers of nodes. The throughput decreases by 82.71% when using the PBFT technique with nodes from 16 to 100, while it decreases by 49.33% when using the hierarchical method. Moreover, concerning the communication overhead, the hierarchical approach lowers the message counts by 75.00% compared to the traditional approach with 16 nodes, while it lowers the message counts by 90.00% for 100 nodes. As indicated by the findings, the optimization effect of the hierarchical approach on

communication overhead increases with more nodes. The capability proves essential in situations where economic collaboration takes place across regions, as it allows the system to maintain high levels of performance despite an increasing number of regions and organizations involved.

3.3. GNN-based Analytics Evaluation

The data analytics layer evaluation uses demand forecasting tasks to assess the effectiveness of GNN models. Demand forecasting aims to predict future demand quantities using historical sales order data, a fundamental task in supply chain collaborative planning. The testing procedure follows a sliding window framework, which utilizes observations from the preceding 7 time steps to predict demand at the next time step. The dataset is split into 70% for training, 15% for validation, and the remaining 15% for testing. This data split ensures that the evaluation does not measure the model's ability to memorize historical data.

Four models are compared in this study. The Multi-Layer Perceptron (MLP) acts as a baseline, ignoring graph information entirely. GCN uses fixed-weight aggregation of neighbor information. GraphSAGE uses a sampling-based approach to information aggregation. GAT employs attention mechanisms to dynamically determine neighbor weights. MLP has two hidden layers with sizes of 64 and 32, respectively. GCN, GraphSAGE, and GAT are based on a two-layer architecture with a hidden size of 64. GAT employs four attention heads. The same training settings are used for all models for fair comparison. The performance metrics used are Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), where lower values indicate better performance.

Table 4 presents the performance of each model on the demand forecasting task. MLP achieves an MAE of 0.847 and RMSE of 1.124, serving as a baseline that reflects predictive capability without utilizing graph structure information. GCN reduces MAE to 0.612 and RMSE to 0.873, representing improvements of 27.74% and 22.33%, respectively, compared to MLP. The results demonstrate that relational information among products in the supply chain network improves predictive performance. The GraphSAGE algorithm obtains an MAE value of 0.583 and an RMSE value of 0.841, showing some improvement compared to GCN. On the other hand, GAT shows better performance than other algorithms with an MAE value of 0.534 and RMSE value of 0.762, giving a percentage improvement of 36.95%, and 32.21% respectively, compared to MLP and percentage improvement of 12.75%, and 12.72% respectively compared to GCN. The main strength of GAT is the ability of the attention function to differentiate between the importance levels of neighboring nodes. Since the degree of relationships between different products in the supply chain network is variable, the use of fixed-weight aggregation does not suffice.

Table 4. Performance comparison on the demand forecasting task

Model	MAE	RMS E	MAE Improvement vs MLP	RMSE Improvement vs MLP
MLP	0.847	1.124	—	—
GCN	0.612	0.873	27.74%	22.33%
GraphSAGE	0.583	0.841	31.17%	25.18%
GAT	0.534	0.762	36.95%	32.21%

To further explore the performance of the network, the distribution of the attention weights generated by the GAT is analyzed. About the SupplyGraph dataset, since edges are connected according to production plants, products that are manufactured from the same plant have high demand correlation because they share raw materials or production capacity limitations. The distribution of attention weights suggests that GAT models give higher weights to the edges connecting nodes corresponding to items manufactured by the same plant (average 0.31) compared to those connecting nodes from different plants (average 0.14). The above observation correlates well with the existing literature in the area of supply chain management, indicating that there is some association learning within the model. The error distribution of different models is shown in Figure 3.

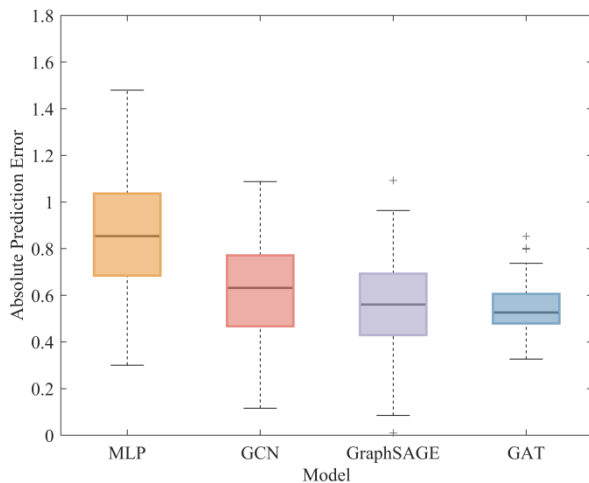


Figure 3. Prediction error distribution on the test set

The scale of the SupplyGraph dataset is small, as it contains only 40 nodes and eight months of data. The experiments presented in this paper primarily serve as a proof of concept and highlight the benefits of using GNNs over traditional approaches in supply chain prediction problems. In actual cross-regional collaborative scenarios, the number of nodes will far exceed that of this experiment. Therefore, the GNN model architecture may need to be

adapted, potentially through the inclusion of sampling techniques or hierarchical aggregation, to reduce the computational burden associated with larger graph scales.

3.4. System Integration Analysis

This section discusses the coordination effect and significance of each layer through an integrated perspective. The blockchain layer provides data operation records and validation capability, which can achieve up to 453 TPS at 100 nodes. This meets the necessary performance criteria for conducting data exchange in cross-regional cooperation. The smart contract layer enables fine-grained permission management using risk-based access control with millisecond-level execution delay for each contract, thus avoiding any possibility of a bottleneck being formed within the system. The data analytics layer has reduced the error in forecasting demand by 36.95% using the GAT model, thus offering a reliable decision-making process for collaboration planning.

As opposed to the current approaches, what makes the proposed architecture unique is that it involves the seamless convergence of reliable data sharing and intelligent analysis capabilities. The traditional approaches to data sharing using blockchains, as suggested by Geng et al. [3] and Tan et al. [14], are mainly focused on storing data and do not have any inherent analytic functions. Therefore, users need to transfer the data from one system to another for analysis purposes, thus increasing the likelihood of data leakage. The proposed approach enables analysis within a trusted environment by integrating the data analytics layer into the architecture. The analysis results can then be recorded and audited. Table 5 compares the characteristics of the proposed architecture with other related solutions across different aspects.

Table 5. Feature comparison with related approaches

Feature	Centralized Platform	Standard Fabric	Geng et al. [3]	Proposed
Decentralized Trust	No	Yes	Yes	Yes
Scalable Consensus	N/A	Limited	Limited	Hierarchical PBFT
Risk-aware Access Control	Basic	Contract-based	Contract-based	Dynamic Risk Scoring
On-chain/Off-chain Storage	N/A	Manual	Supported	Merkle-indexed
Built-in Analytics	External	No	No	GNN-based
Cross-region Coordination	Manual	Manual	Supported	Protocol-level

4. Discussion

The hierarchical grouping strategy used in large-scale node scenarios derives its performance benefits not only from reducing communication complexity but more fundamentally from its alignment with the governance

structures inherent in cross-regional collaboration. In regional economic integration areas like the Yangtze River Delta and the Guangdong-Hong Kong-Macao Greater Bay Area, participating entities naturally form collaborative clusters defined by administrative, industrial, or supply chain affiliations. The interaction intensity among entities within clusters far exceeds that between entities from different clusters. The grouping strategy described in this study aligns the hierarchical structure of collaboration with the consensus mechanism design. Findings on PBFT protocol optimization show that the decrease in throughput with increasing node numbers is mainly due to the exponential growth of messages in fully connected communication structures. The grouping strategy addresses this limitation by limiting the scope of each round of consensus [22]. Research on inter-agency collaboration shows that trust establishment, willingness to share resources, and coordination mechanism design play vital roles in effective collaboration [23]. The hierarchical consensus framework reduces coordination costs by limiting cross-group communication while maintaining the flexibility of efficient intra-group decision-making. This is consistent with existing regional collaboration models, which exhibit characteristics of “layered governance and hierarchical coordination”. From this perspective, the design of consensus protocols should not only focus on achieving optimal performance metrics; it should also consider compatibility with the governance structures of the application contexts.

The improvement in prediction accuracy achieved by GAT over conventional methods is evident. However, the real significance of the data analytics layer lies in its ability to audit the data analysis process itself. In conventional data-sharing models, the data consumer performs analysis after acquiring the data. This makes the entire analysis process, including parameters and interpretation, non-transparent. Data providers have no way to monitor proper data usage. In cross-border or cross-administrative-region scenarios, the lack of data sovereignty and accountability is considered a major hindrance to data flow [24]. The system design includes analytical modules in a secure environment so that requests for analysis, execution of models, scope of data, and output are recorded in a ledger. Research related to applying smart contracts in access control shows that including logic in contracts not only simplifies permissions but also forms an authorization record [25]. It further suggests that, in any case where analysis takes place at the level of the data consumer, the data provider could track the pattern of use through audit trails, and thus establish an accountability chain in a technical sense. The concept of “analysis as audit” might be more successful in enabling information sharing than strategies based on improving prediction alone.

However, the issue of “whether sharing is possible” is addressed by technical design. The equally critical issue of “whether there is willingness to share” in cross-regional cooperation needs more extensive institutional and governance structures. Studies analyzing blockchain technology’s role in fostering economic growth indicate

that the success of the technology is dependent on adjusting regulatory structures, developing the digital abilities of participating organizations, and creating effective systems for benefit sharing [26]. While the design provides fine-grained access controls and audit capability, it does not consider economic aspects like pricing of data, revenue sharing, and punishment for breaches. Studies on adaptive security governance show that the smart contract allows modifying security policies based on the risks incurred during runtime, making the system evolve its governance policies independently [27]. This method provides one possible way of incorporating the incentives of the economy into the framework. From a practical standpoint, technical means alone are not enough to break down data silos. There is also a need to create data ownership regulations, regional coordination institutions, and responsibility allocations. The possibility for further study lies in combining smart contracts with the use of economic incentives, including contract-based transactions for data exchanges or revenue sharing based on contribution, such that some elements of governance are embedded into the system.

Several limitations should be acknowledged. In terms of data analysis, the feature-level fusion method assumes that the datasets have been preprocessed and are of equivalent quality. However, there are considerable differences among different regions regarding how economic data is structured, the level of its granularity, as well as timeliness and reliability. Regarding the existing multi-source data heterogeneity literature, issues related to assessing data quality, detecting data conflict, and uncertainty quantification remain unresolved [28]. Thus, the architecture requires upgrading in those aspects. Specifically, in the experiment, the SupplyGraph database contains only 40 product nodes and 8 months of time series data. Compared to actual economic collaborations between multiple regions that have thousands of companies and years of history behind them, there is an order-of-magnitude gap. This approach of scaling model performance from small data sets to large scales requires additional verification. Concerning platform choice, the experimentation framework is based on Hyperledger Fabric, primarily because of its maturity level and the rich set of tools available for setting up consortia blockchain networks. Comparisons between Fabric and private Ethereum suggest that each platform has advantages and disadvantages in terms of throughput, latency, and resource usage, depending on the specific use case [29]. The generalizability of the findings presented in this study also requires validation on other platforms. Additionally, the current integration of the data analytics layer and the blockchain layer is loosely coupled, and the analytical process is not directly secured. Future studies should focus on tighter integration of these two layers, such as incorporating model inference into verifiable computations, as well as implementing privacy-preserving computation techniques for handling sensitive commercial data [30].

5. Conclusion

The need for enhanced interregional economic collaboration necessitates the establishment of a dual-oriented data sharing infrastructure, with the initial focus on developing trusted data exchange mechanisms to address trust issues stemming from organizational boundaries, and subsequently developing sophisticated analytical capabilities to extract value from fragmented, multi-source data. These two requirements have always been treated independently in previous studies, creating a fragmented approach in which systems are either “trustworthy but not intelligent” or “intelligent but not trustworthy.” This study aims to address both problems within one integrated architecture. The main approach is to extend blockchain trust mechanisms from the data storage layer to the data analytics layer, making the analytical process itself a trusted behavior that can be audited and traced.

The hierarchical grouped consensus strategy serves to demonstrate that scalability optimization does not have to occur at the expense of decentralization. A fundamental component of this strategy is identifying and utilizing intrinsic hierarchical structures within the application context. Administrative structures, upstream-downstream linkages in industrial supply chains, and spatial proximity in regional economic alliances constitute the natural grounds upon which the hierarchy-based consensus model can be divided. The superiority of GAT in predicting demand justifies the need for using network structure data to understand economic systems. Supply chains and the regional economic system should not be treated as mere collections of independent economic actors; rather, the connections between such actors carry significant predictive value, and GNNs provide an effective means of harnessing such value.

Looking toward the future, the use of both technologies would expand to include more than just the loosely-coupled architecture mentioned in this study. With technological advancements such as verifiable computation and privacy-preserving machine learning, there is a growing expectation that analysis will be conducted through more secure execution environments, achieving deep integration where “computation equals consensus.” At the application level, the technical feasibility of data sharing ultimately requires coordination with governance mechanisms and economic incentives to translate into actual collaborative action. The questions of how to encode more complex collaboration rules within smart contracts, how to devise incentive-compatible mechanisms that facilitate data flow, and how to balance transparency with commercial privacy protection are core concerns in assessing the development potential of cross-regional data ecosystems beyond the technical prototype stage. The continued evolution of regional economic integration is creating pressing needs for this research while suggesting promising directions for technical innovation.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions that helped improve this manuscript.

Competing Interest

The authors have no relevant financial or non-financial interests to disclose.

Data Availability

The blockchain scalability experiments were conducted on a self-deployed Hyperledger Fabric test environment, and the corresponding configuration parameters are fully described in this article. The GNN analytics experiments utilized the publicly available SupplyGraph dataset [21], which is accessible at <https://github.com/CIOL-SUST/SupplyGraph>.

Ethical Approval

Not applicable.

Informed Consent

Not applicable.

Author Contribution

Xuefeng Yang: Conceptualization, visualization, funding acquisition, project administration, and investigation; Weizu Lu: Conceptualization, methodology, project administration, and writing-original draft preparation.

References

- [1] Chi M, Li L, Zhong J. The role of blockchain technology in shaping the digital economy: An exploration of financial development. *Financ Res Lett*. 2025;84:107783.
- [2] Susha I, Rukanova B, Zuiderwijk A, Gil-Garcia JR, Gasco Hernandez M. Achieving voluntary data sharing in cross sector partnerships: Three partnership models. *Inf Organ*. 2023;33(1):100448.
- [3] Geng Q, Chuai Z, Jin J. Cross-organizational data exchange based on consortium blockchain with consistency guarantee. *J Supercomput*. 2024;80(12):18199-18236.
- [4] Sanka AI, Cheung RCC. A systematic review of blockchain scalability: Issues, solutions, analysis and future research. *J Netw Comput Appl*. 2021;195:103232.
- [5] Chen B, Ma L, Xu H, Ma J, Hu D, Liu X, Wu J, Wang J, Li K. A Comprehensive Survey of Blockchain Scalability: Shaping Inner-Chain and Inter-Chain Perspectives. *arXiv preprint*. 2024;arXiv:2409.02968.
- [6] Yuan F, Huang X, Zheng L, Wang L, Wang Y, Yan X, Gu S, Peng Y. The Evolution and Optimization Strategies of a PBFT Consensus Algorithm for Consortium Blockchains. *Information*. 2025;16(4):268.
- [7] Cunha PR, Soja P, Themistocleous M. Blockchain for development: a guiding framework. *Inf Technol Dev*. 2021;27(3):417-438.

- [8] Hastig GM, Sodhi MS. Blockchain for Supply Chain Traceability: Business Requirements and Critical Success Factors. *Prod Oper Manag.* 2020;29(4):935-954.
- [9] Kosasih EE, Brintrup A. A machine learning approach for predicting hidden links in supply chain with graph neural networks. *Int J Prod Res.* 2022;60(17):5380-5393.
- [10] Li X, Dunkin F, Dezert J. Multi-source information fusion: Progress and future. *Chin J Aeronaut.* 2024;37(7):24-58.
- [11] Wang Y, Zhong M, Cheng T. Research on PBFT consensus algorithm for grouping based on feature trust. *Sci Rep.* 2022;12(1):12515.
- [12] Yao Z, Fang Y, Pan H, Wang X, Si X. A secure and highly efficient blockchain PBFT consensus algorithm for microgrid power trading. *Sci Rep.* 2024;14(1):8300.
- [13] Pu X, Jiang R, Song Z, Liang Z, Yang L. A medical big data access control model based on smart contracts and risk in the blockchain environment. *Front Public Health.* 2024;12:1358184.
- [14] Tan B, Chen Y, Zhou Y, Li S, Dong Z. A Cross-Institution Information-Sharing Scheme Based on a Consortium Blockchain. *Electronics.* 2023;12(21):4512.
- [15] Davies J. Enhanced scalability and privacy for blockchain data using Merklized transactions. *Front Blockchain.* 2023;6:1222614.
- [16] Kosasih EE, Margaroli F, Gelli S, Aziz A, Wildgoose N, Brintrup A. Towards knowledge graph reasoning for supply chain risk management using graph neural networks. *Int J Prod Res.* 2024;62(15):5596-5612.
- [17] Yang S, Ogawa Y, Ikeuchi K, Shibasaki R, Okuma Y. Post-hazard supply chain disruption: Predicting firm-level sales using graph neural network. *Int J Disaster Risk Reduct.* 2024;110:104664.
- [18] Abushaega MM, Moshebah OY, Hamzi A, Alghamdi SY. Multi-objective sustainability optimization in modern supply chain networks: A hybrid approach with federated learning and graph neural networks. *Alex Eng J.* 2025;115:585-602.
- [19] Kuzlu M, Pipattanasomporn M, Gurses L, Rahman S. Performance Analysis of a Hyperledger Fabric Blockchain Framework: Throughput, Latency and Scalability. In: *Proceedings of the 2019 IEEE International Conference on Blockchain*; 2019 Jul 14-17; Atlanta, USA. Piscataway: IEEE; 2019.
- [20] Dreyer J, Fischer M, Tönjes R. Performance analysis of hyperledger fabric 2.0 blockchain platform. In: *Proceedings of the Workshop on Cloud Continuum Services for Smart IoT Systems*; 2020; Virtual Event, Japan. New York: ACM; 2020.
- [21] Wasi AT, Islam M, Akib AR. SupplyGraph: A benchmark dataset for supply chain planning using graph neural networks. *arXiv preprint.* 2024;arXiv:2401.15299.
- [22] Zhou Z, Onireti O, Zhang L, Imran MA. Slotted ALOHA Based Practical Byzantine Fault Tolerance (PBFT) Blockchain Networks: Performance Analysis and Optimization. *Sensors.* 2024;24(23):7688.
- [23] Gasco-Hernandez M, Gil-Garcia JR, Luna-Reyes LF. Unpacking the role of technology, leadership, governance and collaborative capacities in inter-agency collaborations. *Gov Inf Q.* 2022;39(3):101710.
- [24] Rahman MS, Omar AA, Bhuiyan MZA, Basu A, Kiyomoto S, Wang G. Accountable Cross-Border Data Sharing Using Blockchain Under Relaxed Trust Assumption. *IEEE Trans Eng Manag.* 2020;67(4):1476-1486.
- [25] Hasan MR, Alazab A, Joy SB, Uddin MN, Uddin MA, Khraisat A, Gondal I, Urmi WF, Talukder MA. Smart Contract-Based Access Control Framework for Internet of Things Devices. *Computers.* 2023;12(11):240.
- [26] Kwok AOJ, Treiblmaier H. Blockchain technology as a driver of economic development in small economies: A dynamic capabilities framework. *J Decis Syst.* 2024;33(3):413-438.
- [27] Siddiqui S, Hameed S, Shah SA, Arshad J, Ahmed Y, Draheim D. A smart-contract-based adaptive security governance architecture for smart city service interoperations. *Sustain Cities Soc.* 2024;113:105717.
- [28] Putrama IM, Martinek P. Heterogeneous data integration: Challenges and opportunities. *Data in Brief.* 2024;56:110853.
- [29] Khan MM, Khan FS, Nadeem M, Khan TH, Haider S, Daas D. Scalability and Efficiency Analysis of Hyperledger Fabric and Private Ethereum in Smart Contract Execution. *Computers.* 2025;14(4):132.
- [30] Thirasak K, Chainarong D, Chuaphanngam T, Fugkeaw S. SSX-EHRs: secure and scalable cross-domain EHRs sharing with blockchain sharding and dynamic proxy re-encryption. *EURASIP J Inf Secur.* 2025;2025(1):15.