Journal of Electronic Research and Application, 2025, Volume 9, Issue 6
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Digital Twin Spatiotemporal Indexing Engine: "Edge Cloud" Layered Storage and Fast Backtracking of Real-Time Streaming Data

Huanjing Huang*

Shenzhen 518000, Guangdong, China

**Author to whom correspondence should be addressed.*

Abstract: This article examines the design and implementation of a digital twin spatiotemporal indexing engine. It outlines the core theoretical foundations, including spatiotemporal mapping mechanisms, and discusses key enabling technologies such as hybrid spatiotemporal indexing structures, edge-cloud collaborative storage architectures, and protocol conversion middleware. The study further evaluates system performance through an experimental platform, comparing a layered storage architecture with traditional storage models. The results demonstrate clear advantages in terms of efficiency, scalability, and responsiveness. Finally, the paper explores practical application scenarios and outlines future development directions for next-generation spatiotemporal indexing engines in digital twin systems.

Keywords: Digital twin; Spatiotemporal indexing engine; Tiered storage

Online publication: December 16, 2025

1. Introduction

With the rapid development of the digital economy, China's "14th Five Year Plan for the Development of the Digital Economy" released in 2021 emphasizes the importance of innovative application of digital technology. As an emerging technology, the digital twin's spatiotemporal indexing engine is crucial for processing real-time streaming data. This engine plays a vital role in various applications, from the spatiotemporal mapping relationship and related mechanisms between digital twins and physical entities, to the core structure and algorithms of spatiotemporal indexing engines, to edge and cloud storage technologies, as well as various key technologies in data processing. All these aspects revolve around how to efficiently process real-time streaming data. This not only aligns with the policy guidance for the innovative application of digital technology but also provides a strong research basis for the application of the digital twin spatiotemporal indexing engine in industrial Internet and cross-industry applications. The continuous evolution of this technology promises to enhance data processing capabilities, improve operational efficiencies, and drive innovation across multiple sectors, making it a cornerstone for the future development of the digital economy.

2. Theoretical architecture of digital twin spatiotemporal indexing engine

2.1. Core theoretical framework of digital twin

There is a close spatiotemporal mapping relationship between digital twins and physical entities. By extracting and analyzing the temporal and spatial features of physical entities, a multidimensional spatiotemporal feature parameter model is established. This model can accurately describe the state and behavior of physical entities at different times and spatial locations. Meanwhile, in order to achieve virtual real synchronization, it is necessary to establish an effective mechanism. This mechanism is based on real-time monitoring of changes in the state of physical entities, transmitting relevant data to the digital twin in a timely manner, and updating and adjusting it based on multidimensional spatiotemporal feature parameter models to ensure that the digital twin can reflect the actual situation of the physical entity in real time and accurately [1]. This spatiotemporal mapping relationship and related modeling methods and synchronization mechanisms are important components of the core theoretical framework of digital twins, providing a theoretical basis for the construction of digital twin spatiotemporal indexing engines.

2.2. Operating mechanism of spatiotemporal indexing engine

The hybrid index structure that integrates spatiotemporal topology is the core of spatiotemporal indexing engines. It comprehensively considers the characteristics of spatial location and time series, effectively organizing and storing spatiotemporal data. This structure can quickly locate and retrieve the required data, improving data access efficiency. The dynamic segmentation algorithm based on spatiotemporal windows further optimized the data processing process. It divides data reasonably based on the dynamic changes in time and space, enabling more accurate indexing of data in different spatiotemporal ranges. The trajectory prediction index construction method predicts the motion trajectory of objects and establishes corresponding indexes through the analysis and learning of historical data. This helps to obtain potential data in advance, improving the system's response speed and accuracy [2].

3. Edge cloud tiered storage architecture

3.1. Edge storage optimization strategy

Research on lightweight spatiotemporal data compression algorithms and dynamic slicing techniques for temporal databases for edge storage is essential. Lightweight spatiotemporal data compression algorithms can effectively reduce data storage space and improve storage efficiency while ensuring data accuracy [3]. These algorithms are designed to handle the unique challenges of spatiotemporal data, which often involves large volumes and high velocity. The dynamic slicing technology of time-series databases can perform reasonable slicing based on the temporal characteristics of data, further optimizing the data storage structure. This technology ensures that data is segmented in a way that enhances query performance and reduces redundancy.

Meanwhile, a cache eviction mechanism based on timeliness is proposed. This mechanism considers the timeliness of data, timely eliminates expired or no longer important data, releases storage space, ensures that edge storage always maintains efficient operation, and can quickly store and process real-time streaming data. This approach not only optimizes storage usage but also enhances the overall performance of edge computing systems, providing strong support for subsequent data backtracking and analysis. By integrating these techniques, edge storage solutions can better manage the complexities of real-time data processing, ensuring that critical information is readily available and efficiently managed.

3.2. Principles of cloud tiered storage technology

Cloud tiered storage technology is based on certain principles to achieve efficient data management. For building a distributed spatiotemporal index directory service, it organizes and manages index information through specific algorithms and architectures to ensure fast data localization and retrieval. In terms of the hierarchical storage strategy for hot and cold data based on spatiotemporal semantics, data is divided into different levels according to factors such as access frequency, temporal characteristics, and spatial correlation. Hot data is usually data that has been frequently accessed recently and is closely related to the current spatiotemporal scene, stored in high-performance storage media to ensure fast response. Cold data refers to data with lower access frequency, which can be stored in low-cost storage devices. At the same time, the data lifecycle management model manages the entire process of data from generation to final deletion or archiving, and reasonably plans the storage location and migration strategy of data to optimize storage resource utilization and data access efficiency [4].

4. Key technologies for real-time streaming data processing

4.1. Data collection and preprocessing technology

4.1.1. Multi source heterogeneous data fusion

Developing protocol conversion middleware for IoT devices is the key to achieving multi-source heterogeneous data fusion. Through this middleware, IoT device data of different protocols can be converted to have a unified spatiotemporal benchmark and semantic standard. This process not only solves the problem of data format and semantic differences caused by different device sources but also provides a standardized data foundation for subsequent data analysis and processing. For example, different sensors may adopt different timestamp formats and spatial coordinate systems. The protocol conversion middleware can unify these data into a standard spatio-temporal framework to ensure the accuracy and consistency of data in the fusion process. This is crucial for supporting the “edge cloud” hierarchical storage and rapid backtracking of real-time stream data by the digital twin spatio-temporal index engine [5].

Additionally, the middleware enhances the interoperability of IoT devices, enabling seamless communication and data exchange between devices from different manufacturers and with varying capabilities. By standardizing the data at the protocol level, it ensures that the data can be effectively utilized in a wide range of applications, from real-time monitoring and control to advanced analytics and machine learning, thereby maximizing the value of IoT data in smart systems.

4.1.2. Incremental data collection

Incremental data collection is crucial in real-time streaming data processing. Building an event-triggered collection mechanism is one of the key factors. By setting specific event conditions and triggering data collection when these conditions are met, valuable data can be obtained more accurately, avoiding the collection of a large amount of useless data and improving the efficiency and pertinence of data collection [6]. This approach not only enhances the precision of data capture but also optimizes resource utilization by focusing on relevant events rather than continuous, indiscriminate data ingestion.

At the same time, research on data temporal integrity verification and real-time outlier repair methods is also indispensable. In the process of real-time streaming data collection, various factors may cause timing issues or abnormal values in the data, which can affect subsequent data processing and analysis. Through effective verification and repair methods, ensure the temporal integrity of data, promptly correct outliers, and provide a high-quality data foundation for subsequent processing. These methods are essential for maintaining the reliability and accuracy of real-time data streams, ensuring that the data used for decision-making and analysis is both timely and trustworthy. By integrating these techniques, organizations can significantly enhance their ability to manage and analyze real-time data, leading to more informed and effective decision-making processes.

4.2. Optimization of streaming data transmission

4.2.1. Edge computing node deployment strategy

In order to achieve efficient real-time streaming data processing, the deployment strategy of edge computing nodes is crucial. In this paper, an adaptive edge computing node location model based on space-time density and a dynamic resource scheduling algorithm are proposed. This model comprehensively considers spatiotemporal density factors to more accurately determine node positions, enabling it to better adapt to the spatiotemporal characteristics of real-time streaming data. By analyzing the distribution density of data in different spatiotemporal regions, planning the node layout reasonably, and improving data processing efficiency. Meanwhile, the resource dynamic scheduling algorithm can flexibly allocate node resources based on the traffic and processing requirements of real-time streaming data, avoiding resource idle or overload. The application of this strategy helps to optimize the deployment of edge computing nodes, improve the processing capacity of the whole system for real-time streaming data, and provide strong support for the "edge cloud" hierarchical storage and fast backtracking of real-time streaming data [7].

4.2.2. Data transmission quality assurance

It is crucial to design a multi-path parallel transmission protocol to ensure the quality of data transmission. This protocol can fully utilize multiple paths in the network, effectively improve transmission efficiency, and avoid data transmission interruptions caused by single path failures [8]. In the meantime, building a dynamic QoS control mechanism that takes into account the network state is also crucial. By monitoring network status in real-time, such as bandwidth, latency, packet loss rate, etc., dynamically adjusting transmission strategies, and allocating reasonable network resources for real-time streaming data. For important data or in case of network congestion, priority should be given to ensuring the transmission quality of critical data, ensuring that the data can arrive at the destination accurately and timely, thus meeting the strict requirements of real-time streaming data processing for transmission quality.

5. System implementation and verification

5.1. Platform architecture design

5.1.1. Implementation of microservice componentization

The spatiotemporal index service module adopts advanced algorithms to construct spatiotemporal index structures for efficient processing and retrieval of spatiotemporal data [9]. Its architecture design focuses on the accuracy and speed of indexing, and through optimizing algorithms and data structures, achieves efficient management of large-scale spatiotemporal data. The stream processing engine module focuses on real-time stream data processing and has high-performance stream data processing capabilities. It adopts a distributed architecture that can process multiple data streams in parallel, ensuring real-time and accuracy of the data. The storage controller module is responsible for managing data storage, designing a reasonable storage strategy, and achieving “edge cloud” hierarchical storage. It can automatically allocate data to appropriate storage layers based on its importance and access frequency, improving storage efficiency and data availability. These core modules collaborate with each other to form the platform architecture of the digital twin spatiotemporal indexing engine, providing efficient solutions for the processing and storage of real-time streaming data.

5.1.2. Distributed coordination mechanism

It is crucial to study blockchain based metadata management solutions and cross layer data consistency assurance methods in the design of distributed coordination mechanisms for platform architecture. Blockchain technology provides immutable and traceable features for metadata management, ensuring the authenticity and integrity of data [10]. Through mechanisms such as smart contracts, efficient management and updating of metadata can be achieved while ensuring consistency of data at different levels. In a distributed environment, nodes need to work collaboratively, which relies on effective distributed coordination mechanisms. The blockchain based solution can better cope with the dynamic joining and exiting of nodes, as well as the concurrent access of data, providing strong support for the stable operation of the entire system.

5.2. Experimental platform construction

5.2.1. Construction of testing environment

It is crucial to establish a hybrid cloud experimental platform and build a testing environment for the system validation of the digital twin spatiotemporal indexing engine. It is necessary to configure a heterogeneous device networking environment with typical scenario characteristics, considering the performance, communication protocol, and data format differences of different devices, to ensure the rationality and compatibility of the networking. On this basis, build a hybrid cloud platform that integrates the resource advantages of public clouds and the security features of private clouds, and allocates computing, storage, and network resources reasonably. For the construction of the testing environment, it is necessary to simulate real-time streaming data input in real scenarios, set different data traffic, frequency, and types, and consider the interaction between edge and cloud environments to ensure that the performance and functionality of the system can be accurately reflected in layered storage and fast backtesting, providing a reliable experimental basis for the validation of the digital twin spatiotemporal indexing engine.

5.2.2. Generation of benchmark dataset

During system implementation and validation, the construction of the experimental platform and the generation of benchmark datasets constitute critical stages. The experimental platform must be designed to accommodate the operational requirements of the digital twin spatiotemporal indexing engine, including hardware configurations capable of supporting real-time streaming data processing and a software architecture that enables coordinated operation among system modules. Benchmark dataset generation should be guided by the characteristic patterns of spatiotemporal data flows in application domains such as manufacturing and transportation. This involves developing simulation models grounded in actual business processes and data generation mechanisms, extracting representative data features, and designing a library of test cases that reflect real-world scenarios. The resulting benchmark dataset provides a rigorous and reliable basis for subsequent system verification.

5.3. Performance evaluation analysis

5.3.1. Storage efficiency verification

In terms of storage efficiency verification, emphasis is placed on comparing tiered storage with traditional architectures. For storage space utilization, tiered storage can allocate based on the real-time and importance of data through a reasonable “edge cloud” tiered strategy. Edge storage of data with high real-time requirements reduces unnecessary storage redundancy in the cloud, thereby improving overall storage space utilization. However, traditional architectures may lack this hierarchical mechanism, resulting in scattered data storage and low utilization. In terms of data organization density, hierarchical storage effectively classifies and integrates real-time streaming data, allowing similar data to be closely arranged and improving data organization density. The data organization of traditional architecture is relatively loose, and querying and backtracking require traversing more data, which affects efficiency. By comparing these indicators, the advantages of tiered storage in terms of storage efficiency are clearly demonstrated.

5.3.2. Query response test

The evaluation of the system focuses on assessing both the efficiency of index queries and the accuracy of complex spatiotemporal relationship retrieval. Under varying spatiotemporal conditions, large-scale simulated datasets are used to conduct index query operations while recording metrics such as query latency. For datasets with narrow spatiotemporal ranges, the index is expected to rapidly locate relevant records; as the queried spatiotemporal extent increases, a moderate rise in query time is acceptable, provided it remains within a reasonable threshold that demonstrates the robustness of the indexing mechanism. For validating complex spatiotemporal relationship queries, test cases incorporating relationships such as intersection and containment are constructed. The returned results are then compared against expected outcomes to assess correctness. Accurate correspondence between expected and actual results indicates that the engine can reliably interpret and process intricate spatiotemporal queries, thereby ensuring dependable data retrieval capabilities for digital twin applications.

6. Conclusion

This study delves into the “edge cloud” hierarchical storage and fast backtracking issues of the digital twin spatiotemporal indexing engine in processing real-time streaming data. Through research, it reveals the remarkable technical advantages of the engine in the industrial Internet, including efficient data storage and fast retrieval capabilities, which provides strong support for real-time monitoring and decision-making in industrial production. At the same time, an innovative idea of introducing reinforcement learning to optimize the index structure has been proposed, which is expected to further improve the performance of the engine. In terms of application, it is not limited to the industrial field, but also explores the expansion direction of cross industry application scenarios, such as smart cities, intelligent transportation and other fields. In the future, further in-depth research is needed on the application of reinforcement learning in index optimization, strengthening cross industry cooperation and practice, and promoting the application and development of digital twin spatiotemporal index engines in more fields.

Disclosure statement

The author declares no conflict of interest.

References

[1] Friederich J, Francis D, Lazarova-Molnar S, et al., 2022, A Framework for Data-Driven Digital Twins of Smart Manufacturing Systems. Computers in Industry, 2022(136): 103586.

[2] Bocewicz G, Wójcik R, Sitek P, et al., 2022, Towards Digital Twin-Driven Performance Evaluation Methodology of FMS. Applied Computer Science, 18(3): 5.

[3] Tong X, Liu Q, Pi S, et al., 2020, Real-Time Machining Data Application and Service Based on IMT Digital Twin. Journal of Intelligent Manufacturing, 31(5): 1113–1132.

[4] Fan Y, Yang J, Chen J, et al., 2021, A Digital-Twin Visualized Architecture for Flexible Manufacturing System. Journal of Manufacturing Systems, 2021(60): 176–201.

[5] Zhuang C, Miao T, Liu J, et al., 2021, The Connotation of Digital Twin, and the Construction and Application Method of Shop-Floor Digital Twin. Robotics and Computer-Integrated Manufacturing, 2021(68): 102075.

[6] Luo D, Guan Z, He C, et al., 2022, Data-Driven Cloud Simulation Architecture for Automated Flexible Production Lines: Application in Real Smart Factories. International Journal of Production Research, 60(12): 3751–3773.

[7] Resman M, Protner J, Simic M, et al., 2021, A Five-Step Approach to Planning Data-Driven Digital Twins for Discrete Manufacturing Systems. Applied Sciences, 11(8): 3639.

[8] Li X, Wang L, Zhu C, et al., 2021, Framework for Manufacturing-Tasks Semantic Modelling and Manufacturing-Resource Recommendation for Digital Twin Shop-Floor. Journal of Manufacturing Systems, 2021, 58: 281-292.

[9] Kumbhar M, Ng A, Bandaru S, 2023, A Digital Twin Based Framework for Detection, Diagnosis, and Improvement of Throughput Bottlenecks. Journal of Manufacturing Systems, 2023(66): 92–106.

[10] Magalhães L, Magalhães L, Ramos J, et al., 2022, Conceiving a Digital Twin for a Flexible Manufacturing System. Applied Sciences, 12(19): 9864.