International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

A Survey of Optimized Merkle Tree Structures for Query Processing in Blockchain-Based HTAP Systems

Mr. Sagar R. Mali [1], Dr. Dheeraj Hebri [2], Dr. Amolkumar N. Jadhav[3]

[1]Research Student, Computer Science and Engineering, Srinivas University, Manglore
[2]Associate Professor, Computer Science and Engineering, Srinivas University, Manglore
[3]Associate Professor, Computer Science and Engineering, Annasaheb Dange College of Engineering Technology, Ashata

Abstract
Systems that provide real-time analytics on live transactional data are known as Hybrid Transactional/Analytical Processing (HTAP) systems. Immutability, provenance tracking, and tamper-proof storage are just a few benefits of incorporating blockchain technology into HTAP systems as data volumes increase and system topologies become more decentralized. High latency, restricted support for range and authenticated queries, and a dearth of adaptive optimization strategies are some of the major issues with query processing in blockchain-based HTAP systems. Merkle tree architectures, optimization algorithms, and hybrid indexing strategies are the main topics of this review, which examines the most recent approaches to blockchain-based query processing. We specifically examine the efficacy of recent developments such query authentication methods, Verkle Trees, and Multi-State Merkle Patricia Tries. We also draw attention to the growing importance of bio-inspired metaheuristics in resolving performance bottlenecks, such as the recently suggested Giant Armadillo Optimization (GAO). Future directions are suggested to close the gaps in the current systems' scalability, flexibility, and multi-query handling. The foundation for creating safe, effective, and intelligent query processing techniques for blockchain-integrated HTAP platforms is provided by this thorough review.

Keywords: Blockchain, HTAP, Query Optimization, Merkle Tree, Bio-Inspired Algorithms, Armadillo Optimization

1. Introduction

Hybrid Transactional and Analytical Processing (HTAP) systems have become more popular as a result of the increasing demand for real-time insights from transactional data. These systems serve mixed workloads on a single platform by combining the capabilities of Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP). Although HTAP systems provide a single view of historical and operational data, they have inherent problems with query optimization, performance isolation, and data consistency, particularly when dealing with high-volume, diverse workloads. Blockchain has become a supplementary layer in HTAP systems to address issues with data integrity, auditability, and trust in distributed situations. Improved data security, provenance tracing, and tamper-proof logging are provided by its decentralized design and immutable ledger features. However, because of its append-only structure, lack of indexing, and high transaction confirmation latency, integrating blockchain adds a great deal of complexity to query processing.

Blockchain systems have made extensive use of conventional indexing and authentication techniques, including Merkle trees, to facilitate secure queries and data verification. However, when it comes to handling more complex query types like range queries and k-nearest neighbor (kNN) searches, these approaches frequently suffer with scalability and flexibility. Furthermore, in order to guarantee constant performance, the dynamic nature of HTAP workloads necessitates the use of clever and flexible optimization techniques. Bio-inspired optimization algorithms, including the Giant Armadillo Optimization (GAO), have emerged as promising ways to improve query efficiency in recent years. In order to minimize computing cost, enable adaptive indexing, and dynamically optimize query pathways, these techniques imitate natural behaviors. In blockchain-based HTAP systems, combining these techniques with conventional Merkle tree structures creates new opportunities for scalable, secure, and intelligent query processing. An extensive overview of the state of query processing in blockchain-integrated HTAP systems is given in this review paper. In order to get around the drawbacks of traditional methods, it investigates a variety of cryptographic structures, optimization strategies, and authentication procedures. Particular focus is placed on hybrid models that improve range query and authentication capabilities by utilizing both algorithmic and structural optimizations, such as Harmonic Analysis in conjunction with GAO. This review establishes the foundation for next-generation secure and optimal data processing frameworks by examining the body of existing literature, pointing out gaps, and suggesting future paths.

2. Review Methodology

The purpose of this review of the literature was to methodically investigate and assess the body of knowledge regarding query processing in blockchain-based Hybrid Transactional and Analytical Processing (HTAP) systems. Authenticated indexing structures, optimization algorithms, and hybrid data processing models were the main areas of focus.

2.1. Scope and Objectives

The primary objective of this review is to analyze current techniques and models that address:

- Secure and effective query processing in blockchain settings
- Authenticated data indexing using Merkle trees and its variations
- Application of bio-inspired optimization algorithms (e.g., GAO) to enhance query performance
- HTAP system integration issues and fixes in decentralized environments

The review particularly emphasizes methods that support range queries, multi-query processing, kNN-based authentication, and adaptability to dynamic workloads.

2.2. Literature Sources

To ensure the review captures the most relevant and credible works, we used the following digital libraries:

- **IEEE Xplore**
- **ACM Digital Library**
- **ScienceDirect (Elsevier)**
- **SpringerLink**
- **arXiv** and **Google Scholar** for preprints and emerging methods.

2.3. Search Strategy

The following keywords and their combinations were used for querying the databases:

- "HTAP systems"
- "blockchain query processing"
- "Merkle tree optimization"
- "range query in blockchain"
- "authenticated data structures"
- "Giant Armadillo Optimization"
- "bio-inspired optimization in databases"

The search was restricted to articles published between **2015 and 2024**, with a focus on peer-reviewed journals, high-impact conferences, and leading technical reports.

2.4. Inclusion and Exclusion Criteria

Inclusion Criteria:

- Papers focusing on blockchain-based data processing or HTAP systems
- Studies proposing or evaluating authenticated query mechanisms
- Research involving Merkle trees, spatio-temporal queries, or optimization techniques

Exclusion Criteria:

- Non-peer-reviewed blog posts or non-academic reports
- Articles focusing purely on cryptocurrency or unrelated blockchain finance applications
- Studies without technical contributions (e.g., opinion papers)

2.5. Analysis and Categorization

Selected papers were analyzed and categorized based on:

- The type of query supported (e.g., range, kNN, cross-chain)
- The indexing or authentication method used (e.g., Merkle tree, Verkle tree, Patricia Trie)
- The optimization techniques applied (e.g., heuristic, AI-based, evolutionary)
- Identified limitations and gaps

The findings are synthesized in the following sections to highlight methodological trends, strengths, and areas needing improvement.

3. Thematic Literature Review

The literature is arranged in this area according to major themes that are pertinent to blockchain-integrated HTAP systems. These themes include query processing strategies, data authentication frameworks, performance optimization methodologies, and the constraints of hybrid systems. Key findings, current and foundational contributions, and gaps that inspire more research are all highlighted in each area.

3.1 Blockchain Query Processing Techniques

Blockchain data's decentralized storage format and immutable, append-only nature create special difficulties for query processing. Blockchain systems do not lend themselves well to traditional relational query paradigms; hence new frameworks are required for effective data access.

In order to enable SQL-like queries on Ethereum, Han et al. [4] suggested integrating relational logic into smart contracts. Although their work made it easier to query transactional data, it was unable to support activities that required a lot of analytics and did not scale well in high-throughput HTAP settings. With the introduction of vChain+ by Wang et al. [6], blockchain capabilities were expanded to accommodate boolean range queries with verifiability guarantees. Nevertheless, it was unable to manage intricate processes like aggregations and joins, which are essential in analytical settings.

Härer [7] investigated cross-chain querying, which aims to increase interoperability by bringing dissimilar blockchain networks together. However, neither HTAP-based workload support nor deep integration with storage-level optimization was present in the system.

Gap Identified: Despite advancements, most solutions remain limited to either transactional or analytical queries but not both. There is a need for hybrid systems that unify query expressiveness, integrity verification, and real-time responsiveness.

3.2 Merkle Trees in Data Authentication

The de facto standard for enabling authentication and data integrity in blockchain systems is Merkle trees. Their hash-based hierarchical structure allows for rapid validation of data inclusion and concise proofs.

Merkle tree topologies were used by Xu et al. [1] in Authenticated kNN and Range Queries, with excellent accuracy and effective proof verification. This idea was expanded to Multi-State Merkle Patricia Tries (MSMPT) by Mardiansyah et al. [2], allowing for quick, lightweight multi-query processing in blockchain environments. Despite their efficiency, these systems frequently can't adjust to changing query patterns.

The Verkle AR-tree*, which Chen and Liang [3] presented, expanded authentication capabilities to multidimensional spatiotemporal data. Although it was difficult to handle and computationally demanding under varied query loads, this was especially helpful in location-aware systems.

Gap Identified: Current Merkle-based authentication schemes offer solid integrity but suffer from limited scalability and poor adaptability to diverse or evolving queries. There's limited research on integrating Merkle trees with adaptive or hybrid optimization strategies to improve overall system responsiveness.

3.3 Optimization Techniques for Query Speed and Accuracy

A number of studies have looked into algorithmic optimization strategies to lower latency and increase accuracy because of the computational expense of requesting and validating blockchain data.

Bao et al. [8] developed a dual-stage verification mechanism for spatial blockchain applications, which decreased query time for geospatial data, while Wang et al. [6] used cryptographic indexing in vChain+ for efficient boolean range queries. However, these strategies did not generalize effectively and were closely related to particular topics.

Furthermore, Qu et al. [21] focused on spatiotemporal blockchain indexing, but they mostly dealt with static data models in their work. Not enough research has been done on optimization for streaming or near-real-time data, especially in an HTAP setting.

Gap Identified: Most optimization strategies are rigid, domain-specific, or lack general applicability. There is a need for adaptable optimization frameworks that can dynamically reconfigure query paths and indexes based on system workload and data distribution.

3.4 Bio-Inspired Algorithms: From GA to GAO

Bio-inspired algorithms are extremely relevant in the context of blockchain query processing because they provide robust and adaptive solutions for high-dimensional, non-linear optimization issues.

The use of early techniques, including Genetic Algorithms (GA), for indexing and search space reduction was constrained by their complexity and convergence durations. Alsayyed et al. [22] recently presented Giant Armadillo Opti-mization (GAO), a novel metaheuristic that draws inspiration

from the animal's adaptive and defensive actions. When it comes to addressing intricate optimization problems, GAO exhibits high efficiency, quick convergence, and superior solution quality.
Nevertheless, GAO has not yet been thoroughly included into HTAP and blockchain systems. By integrating GAO with harmonic analysis and Merkle tree architectures, the proposed Harmonic Giant Armadillo Optimization (HGAO) model seeks to close this gap and increase query execution flexibility and range query efficiency.

Gap Identified: There is a lack of comprehensive frameworks that integrate bio-inspired optimization (e.g., GAO) with authenticated data structures for real-time, multi-query processing in blockchain systems. This area presents a significant opportunity for innovative hybrid models.

3.5 Limitations in Current HTAP + Blockchain Solutions

Existing HTAP systems still encounter difficulties when integrating blockchain technology, despite the encouraging developments. HyPer and SAP HANA are examples of traditional HTAP platforms that are not made to handle the immutability limitations or decentralized nature of blockchain systems.
Polypheny-DB, a polystore that enables HTAP workloads across heterogeneous databases, was introduced by Vogt et al. [14]. Despite addressing federated querying and polyglot persistence, the solution lacked native blockchain integration and cryptographic data verification techniques.
Similarly, HyBench, a benchmarking tool for HTAP system evaluation, was proposed by Zhang et al. [13]. However, the difficulties brought about by blockchain, like proof creation, chain reorganization, and smart contract querying, are not reflected in these tools.

4. Comparison Table

Author(s)	Method	Strengths	Weaknesses	Year
Xu et al. [1]	Authenticated kNN / Range Query (AKQ/ARQ)	High correctness, efficient authentication	Limited to basic authenticated queries; lacks cross-chain support	2023
Mardiansyah et al. [2]	Multi-State Merkle Patricia Trie (MSMPT)	Fast multi-query processing, suitable for lightweight chains	Lacks AI-based adaptability for element categorization	2023
Chen & Liang [3]	Verkle AR*-Tree	Improves spatio-temporal query performance; cube indexing	Not suited for all queries; requires flexible multi-dimensional indexes	2022
Han et al. [4]	Ethereum-based SQL Query Support	Reduces smart contract access cost; uses relational logic	Scalability not improved; lacks embedded DB optimization	2023
Almi'Ani et al. [5]	Graph-Based Profiling	High trustworthiness accuracy us-	Performance degrades with large datasets;	2023

		ing historical data	lacks adaptability	
Wang et al. [6]	vChain+	Ensures query integrity; efficient key management	No support for complex queries like joins or aggregates	2022
Härer [7]	Cross-Chain Query Language	Boosts interoperability across blockchains	Doesn't support advanced storage integration or multi-chain analytics	2023
Bao et al. [8]	Spatial Blockchain (Dual-stage Verification)	Secure transactions; efficient spatial query support	Doesn't integrate GIS fully; lacks broader applicability	2023
Vogt et al. [14]	Polypheny-DB	Supports polyglot persistence and HTAP workloads	No native blockchain compatibility	2020
Zhang et al. [13]	HyBench	New HTAP performance benchmark	Doesn't incorporate blockchain-specific performance metrics	2024

5. Research Gaps and Future Directions

HTAP systems and blockchain-based query processing have advanced significantly, but there are still a number of important issues that need more research. The literature now in publication offers solid basic work in authentication, indexing, and optimization. The integration of adaptive indexing, real-time benchmarking, and intelligent optimization approaches in decentralized hybrid systems is still lacking, though.

5.1 Integration of Merkle Trees with Intelligent Optimizers

Merkle trees are well known for making it possible for blockchain systems to verify data effectively and impenetrably. But the majority of implementations lack flexibility and are static. They do not dynamically modify their structure in response to query burden or data distribution, nor do they communicate with clever optimizers. Even though Xu et al. [1] and Mardiansyah et al. [2] used Merkle variants to create authenticated query systems, these methods do not optimize the tree design through learning-based or metaheuristic methods. Promising approaches for creating adaptive Merkle trees that increase query latency and proof production speed include harmonic-based techniques and bio-inspired algorithms like Giant Armadillo Optimization (GAO) [22].

Future Direction: Hybrid Merkle structures that incorporate intelligent optimizers (such as GAO, Particle Swarm, and Reinforcement Learning) are desperately needed in order to provide multi-query flexibility and dynamic reconfiguration.

5.2 Real-Time Query Performance Benchmarks in HTAP

Blockchain-specific needs like cryptographic verification, immutability, and append-only storage are not taken into account by existing HTAP benchmarks like HyBench [13], which mainly concentrate

on latency and throughput under conventional workloads.Furthermore, equitable comparison and system development are hampered by the absence of real-time benchmark suites for assessing authenticated or range-based queries in blockchain-integrated HTAP systems. Query verifiability, proof size, and Merkle proof validation latency—all crucial components in blockchain contexts—are frequently not measured by current performance metrics.

Future Direction: There is a great demand for thorough benchmarking frameworks that use blockchain to assess consistency assurance, verification cost, and end-to-end query performance in real-time HTAP systems.

5.3 Hybrid Algorithms for Adaptive Indexing

Under different workloads, static indexing methods frequently deteriorate. Although they offer considerable flexibility, Patricia attempts [2] and Verkle trees [3] are not entirely adaptable to mixed or changing query patterns.Hybrid indexing algorithms that blend adaptive behavior and cryptographic assurance are few. Workload-aware tuning should enable such systems to modify their indexing structure in real time. Moreover, native support for integrating blockchain indexing with adaptive processes is absent from current polystores and HTAP systems [14].

Future Direction: Future systems ought to use hybrid index architectures, which combine adaptive algorithms such as GAO or harmonic-based models with Merkle trees. Depending on demand patterns, they should dynamically reorganize data pathways to save query time and storage costs.

Summary of Identified Gaps:

Gap	Impact	Proposed Research
Static Merkle tree design	High latency in dynamic queries	Integrate with GAO/harmonic models
No blockchain-aware HTAP benchmarks	Poor system comparability	Design real-time benchmarking tools
Lack of adaptive indexing	Inefficient query handling	Develop hybrid, workload-aware indexing

6. Conclusion

This paper focuses on significant developments in HTAP systems and blockchain query processing, particularly through improved data authentication structures like Merkle trees and their variations. The efficiency and query integrity of current approaches are enhanced, but they frequently lack flexibility and support for sophisticated, authenticated queries. Hybrid models, such as the suggested HGAO_M-Tree, which combines harmonic analysis, Merkle trees, and bio-inspired optimization, provide a possible path to fill these gaps. Future studies should concentrate on real-time benchmarking and adaptive indexing to satisfy the changing needs of safe, decentralized data systems.

References

1. H. Xu, B. Xiao, X. Liu, L. Wang, S. Jiang, W. Xue, J. Wang, and K. Li, "Empowering authenticated and efficient queries for STK transaction-based blockchains," IEEE Transactions on Computers, vol. 72, no. 8, pp. 2209–2223, 2023.
2. V. Mardiansyah, A. Muis, and R. F. Sari, "Multi-state merkle patricia trie (MSMPT): High-performance data structures for multi-query processing based on lightweight blockchain," IEEE Access, 2023.
3. H. Chen and D. Liang, "Adaptive spatio-temporal query strategies in blockchain," ISPRS International Journal of Geo-Information, vol. 11, no. 7, p. 409, 2022.
4. J. Han, Y. Seo, S. Lee, S. Kim, and Y. Son, "Design and Implementation of Enabling SQL–Query Processing for Ethereum-Based Blockchain Systems," Electronics, vol. 12, no. 20, p. 4317, 2023.
5. K. Almi'Ani, Y. C. Lee, T. Alrawashdeh, and A. Pasdar, "Graph-based profiling of blockchain oracles," IEEE Access, vol. 11, pp. 24995–25007, 2023.
6. H. Wang, C. Xu, C. Zhang, J. Xu, Z. Peng, and J. Pei, "vChain+: Optimizing verifiable blockchain boolean range queries," in Proc. IEEE 38th Int. Conf. Data Eng. (ICDE), 2022, pp. 1927–1940.
7. F. Härer, "A Cross-Chain Query Language for Application-Level Interoperability Between Open and Permissionless Blockchains," arXiv preprint arXiv:2307.00951, 2023.
8. Y. Bao, Z. Gui, Z. Sun, A. An, and Z. Huang, "Spatial Blockchain: Enhancing Spatial Queries and Applications through Integrating Blockchain and Spatial Database Technologies," Electronics, vol. 12, no. 20, p. 4287, 2023.
9. P. Ayyalasomayajula and M. Ramkumar, "Optimization of Merkle Tree Structures: A Focus on Subtree Implementation," in Proc. Int. Conf. Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2023, pp. 59–67.
10. D. Przytarski, C. Stach, C. Gritti, and B. Mitschang, "Query processing in blockchain systems: Current state and future challenges," Future Internet, vol. 14, no. 1, p. 1, 2021.
11. Y. Zhang, L. Liu, D. Li, F. Liu, and X. Lu, "DHT-based range query processing for web service discovery," in Proc. IEEE Int. Conf. Web Services, 2009, pp. 477–484.
12. H. Song, W. Zhou, F. Li, X. Peng, and H. Cui, "Rethink query optimization in HTAP databases," Proc. ACM Manage. Data, vol. 1, no. 4, pp. 1–27, 2023.
13. C. Zhang, G. Li, and T. Lv, "HyBench: A New Benchmark for HTAP Databases," Proc. VLDB Endowment, vol. 17, no. 5, pp. 939–951, 2024.
14. M. Vogt et al., "Polypheny-DB: towards bridging the gap between polystores and HTAP systems," in Heterogeneous Data Management, Polystores, and Analytics for Healthcare: VLDB Workshops, Poly 2020 and DMAH 2020, 2021, pp. 25–36.
15. S. Huang, G. Wang, Y. Yan, and X. Fang, "Blockchain-based data management for digital twin of product," J. Manuf. Syst., vol. 54, pp. 361–371, 2020.
16. L. Zhu, Y. Wu, K. Gai, and K. K. R. Choo, "Controllable and trustworthy blockchain-based cloud data management," Future Generation Computer Systems, vol. 91, pp. 527–535, 2019.
17. R. Chopade and V. Pachghare, "Data tamper detection from NoSQL Database in forensic environment," J. Cyber Security and Mobility, pp. 421–450, 2021.
18. Y. Yang, R. Dunlap, M. Rexroad, and B. F. Cooper, "Performance of full text search in structured and unstructured peer-to-peer systems," in Proc. IEEE INFOCOM, 2006, pp. 1–12.

19. M. Ramkumar, "Executing large-scale processes in a blockchain," J. Capital Markets Studies, vol. 2, no. 2, pp. 106–120, 2018.
20. S. Wan, Y. Zhao, T. Wang, Z. Gu, Q. H. Abbasi, and K. K. R. Choo, "Multi-dimensional data indexing and range query processing via Voronoi diagram for internet of things," Future Generation Computer Systems, vol. 91, pp. 382–391, 2019.
21. Q. Qu, I. Nurgaliev, M. Muzammal, C. S. Jensen, and J. Fan, "On spatio-temporal blockchain query processing," Future Generation Computer Systems, vol. 98, pp. 208–218, 2019.
22. O. Alsayyed, T. Hamadneh, H. Al-Tarawneh, M. Alqudah, S. Gochhait, I. Leonova, O. P. Malik, and M. Dehghani, "Giant Armadillo optimization: A new bio-inspired metaheuristic algorithm for solving optimization problems," Biomimetics, vol. 8, no. 8, p. 619, 2023.
23. E. Damsleth and E. Spjøtvoll, "Estimation of trigonometric components in time series," J. Am. Stat. Assoc., vol. 77, no. 378, pp. 381–387, 1982.
24. G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE, Catania, Sicily, Italy, 2003, pp. 986–996.
25. H. Li, R. Lu, L. Zhou, B. Yang, and X. Shen, "An efficient merkle-tree-based authentication scheme for smart grid," IEEE Systems Journal, vol. 8, no. 2, pp. 655–663, 2013.