Grenze International Journal of Engineering and Technology, January Issue

Har-GAO: Harmonic Giant Armadillo Optimization based Merkle Tree for Hybrid Query Processing in Block Chain based HTAP System

Sagar R. Mali[1], Dheeraj Hebri[2] and Amolkumar N. Jadhav[3]
[1-2]Srinivas University/Computer Science and Engineering, Mangalore, India
Email: sagarmali6@gmail.com, dheeraj.h7@gmail.com
[3]Annasaheb Dange College of Engineering Technology/ Computer Science and Engineering, Ashata, India
Email: pramolkumar451@gmail.com

***Abstract*—Modern data-intensive applications demand the integration of analytical and transactional processing to support real-time business operations. Hybrid Transactional and Analytical Processing (HTAP) systems have emerged as a response, combining Online Transactional Processing (OLTP) and Online Analytical Processing (OLAP) engines to efficiently handle interleaved workloads. However, most current distributed HTAP solutions rely on asynchronous data replication between specialized stores, limiting their utility for real-time operational analytics and lacking support for polyglot persistence. To address diverse and dynamically changing workloads, there is a need for a flexible HTAP-aware polystore system that combines transactional guarantees with support for heterogeneous data sources and query interfaces. Additionally, ensuring data integrity and provenance has become crucial. Blockchains, with their immutable, append-only structure, offer resistance to tampering and denial-of-service attacks, enabling historical data queries and provenance analysis. However, they often require external storage and local query processing. Distributed Hash Tables (DHTs) have gained popularity in decentralized service discovery, yet their basic exact-match functionality falls short for complex range queries. Layered DHT-based range query schemes attempt to overcome these limitations but face scalability and performance challenges. To further ensure data integrity and enable efficient verification, Merkle hash trees provide cryptographic commitments to dynamic data structures and are increasingly applied in distributed systems.**

***Index Terms*— HTAP, Polystore Systems, Blockchain, Distributed Hash Tables (DHTs), Merkle Hash Trees.**

I. INTRODUCTION

These days, data-intensive systems frequently use enormous volumes of data for various real-time business operations, which calls for combining transactional and analytical processing methods. As a result, a lot of recent research and development work has gone into creating hybrid transactional and analytical processing (HTAP) systems, which are supposed to segregate the performance of interleaved workloads and enable quick analysis of new data.

Grenze ID: 01.GIJET.12.1.6

An online transactional processing (OLTP) engine that facilitates high throughput transaction processing and an online analytical processing (OLAP) engine that facilitates low latency analytics are often found in a functional HTAP database.

Two engines with specialized data stores and asynchronously replicated data from one copy to the other are commonly used in a popular category of distributed HTAP databases to efficiently manage mixed workloads [12] [13]. Regularly copying transactional data to the polystore renders it unusable for the majority of operational analytical queries that need the most recent data. Moreover, federated access for transactional workloads is not provided by this method. HTAP systems have become more and more popular recently. These systems are effective for combined transactional and analytical workloads [15] [16]. Although these systems typically feature support for data manipulation, they lack common properties of a polystore system, such as polyglot persistence and support for a wide range of workloads, because they are primarily built for concrete applications and their typical workloads. Handling varied, heterogeneous, and dynamically changing workloads requires a comprehensive and flexible polystore system that simultaneously provides the transaction capabilities of an HTAP system. Since there isn't a ready-made solution for every use case, such a "HTAP-aware polystore" needs to be flexible to allow for the inclusion of support for a broad range of query languages, interfaces, and underlying data stores [14].

Because traditional database systems lack the data integrity checks that guarantee recorded data remains in the same condition as when it was first inserted, data are not secured against this attack vector. Hardening these databases against data manipulation is hence the focus of recent attention [17]. Blockchain technology provides an answer to these problems. First of all, it is protected against data change because it is tamper-resistant and immutable. Second, it is immune to denial-of-service attacks due to its decentralized nature. In this regard, blockchains offer a multitude of interesting additional use cases for inquiries due to their internal data management. A data log is produced where many data revisions coexist since data on a blockchain are only appended. This allows the data history to be queried for provenance analyses because there is no native data log to query.

The availability of this blockchain data history forces applications to store data outside of a blockchain, and often they also need to perform additional query processing, mostly locally [10]. On the past ten years, there has been a growing trend of hosting several Web services on decentralized networks that are structured using the distributed hash table (DTH) model. Due to their scalability, self-organization, and adaptability to changes in the underlying topology, DHTs are a viable foundation for achieving effective Web service discovery [18]. DHTs only offer exact-match queries as their primary function, which may be sufficient for certain straightforward applications.

However, many more complex services cannot be accommodated by the straightforward exact-match interface. Range inquiries, which are the queries seen above, are crucial for Web service discovery [11]. Several DHT-based methods have been proposed recently for processing range queries. Layered schemes are an important class of DHT-based range query techniques that are constructed on top of existing DHTs without requiring changes to the underlying infrastructure. Among their many methodological benefits is error isolation and design simplicity. However, current layered techniques are inefficient because they do not adapt the behavior of underlying DHTs to range requests.

In most proposed layered methods, the query delay depends on the size of the network and the properties of the query. Because of this, current approaches are unable to ensure that the results will be returned in a finite time that is solely dependent on the network's size [11].

Calculating a cryptographic commitment to a dynamic database of entries is made simpler by a data structure known as a Merkle hash tree. In such a tree, each leaf node represents the cryptographic hash of a record, whereas each non-leaf node represents the cryptographic hash of its two "child nodes" [19] [20]. This commitment to the entire database is symbolized by the "root node." One way to conceptualize the leaves of a Merkle hash tree [25] is as distinct data-blocks that require independent "processing." The primary advantage of using a Merkle tree to provide a succinct, dynamic commitment to each block is efficiency. Because Merkle trees can efficiently and securely verify the contents of enormous data structures, they have found use in a variety of real-world applications [21] [9].

II. Related Works

In recent years, there has been a growing interest in transaction-based blockchain systems that handle queries efficiently and securely. To improve interoperability among blockchains, facilitate range and KNN searches, and strengthen data authentication, several models have been put forth.The Authenticated kNN/Range Query

(AKQ/ARQ) architecture was presented by Xu et al. [1], which enhanced the accuracy and authentication efficiency of blockchain-based inquiries. Nevertheless, the model's practical application in multi-blockchain setups is limited due to its lack of support for cross-chain inquiries. The Multi-State Merkle Patricia Trie (MSMPT) was proposed by Mardiansyah et al. [2] to enable multi-query processing in lightweight blockchains, resulting in quicker query performance and lower storage overhead. However, the methodology does not incorporate artificial intelligence or optimization methods for effective data classification. Chen and Liang [3] created the Verkle AR*-tree to facilitate spatio-temporal queries, improving spatial indexing and authentication for more intricate queries.

Although this approach enhanced query efficiency, it was hampered by its inability to accommodate a wide variety of query forms and by the absence of strategies like block pruning. By supporting SQL-style queries on Ethereum without the requirement for specialized data structures, Han et al. [4] solved the need for query accessibility while cutting expenses. Nevertheless, the approach's performance and scalability on sizable smart contract datasets are constrained.

Regarding data trustworthiness, Almi'Ani et al. [5] presented a graph-based profiling technique that used historical source data to obtain high accuracy in assessing blockchain oracle reliability. In large-scale scenarios, it had significant processing delays notwithstanding its efficacy. vChain+, a blockchain-based query model with integrity verification and public key management, was presented by Wang et al. [6]. Nevertheless, the system's lack of support for sophisticated functions like join and aggregation searches limits its usefulness in intricate query contexts.

Data views across heterogeneous blockchains are made possible by cross-chain query technologies like Härer's Cross-Chain Query Language (CCQL) [7]. Although this method allows for unified querying, it is not optimized for large-scale applications or distributed storage. A spatial blockchain architecture with two-level verification methods was created by Bao et al. [8] for real estate transactions. The approach does not facilitate interaction with GIS technology for more comprehensive smart-city applications, despite its effectiveness in preserving transaction integrity.

These current approaches demonstrate the advancements in blockchain query frameworks, but they also draw attention to enduring drawbacks, like poor scalability, a narrow range of query types, and a lack of intelligent optimization. The suggested HGAO_M-Tree framework combines Merkle Tree topologies, harmonic analysis, and giant armadillo optimization (GAO) to fill in these gaps. This hybrid architecture offers a solid solution for next-generation blockchain query systems by supporting dynamic KNN and range queries, increasing query efficiency through intelligent block pruning, and enhancing scalability and fault tolerance.

III. LITERATURE SURVEY

This section highlights recent research discoveries and advancements while reviewing the literature on several traditional query processing methods for transaction-based blockchains.

A. Main Challenge

The major complications handled by the traditional query processing schemes for transaction based blockchain is given as follows,

- The developed approach in [2], utilized high performance data structure for multi query processing based on a light weight block chain and attained improved block chain information retrieval. Nevertheless, this scheme failed to consider various optimization techniques to improve the operational efficiency.
- The developed Verkle AR*-tree based Adaptive Spatio-Temporal Query Strategies in [3] accomplished better versatility. Moreover, it failed to consider block pruning which reduces the processing time.
- In [4], a new scheme was created with lower management expenses that allowed Structured Query Language (SQL) query operations to extract common transaction information and smart contract information within the blockchain system. Despite the fact that the system's scalability was not enhanced.
- A novel graph-based profiling method was devised in [5], to determine the trustworthiness of blockchain oracles. This model accomplished highest accuracy. But, it failed to focus on reducing delays in processing while analyzing large datasets.
- Blockchain makes it possible to create smart contracts and process transactions. It gives end users the ability to carry out contracts without the need for middlemen. However, there are certain problems with information retrieval, such as inconsistent results, limited query possibilities, and lengthy processing times. To get over the aforementioned restrictions, it is crucial to create an optimization-based query processing model utilizing blockchain technology.

TABLE I

Authors	Methods	Advantages	Disadvantages
Xu, H., et al.[1]	Authenticated kNN/Range Query (AKQ/ARQ)	The devised approach had improved correctness and feasibility and demonstrated query and authentication efficiency.	Cross-chain queries and other more effective authenticated query processing methods were not added to this methodology.
Mardiansyah, V., et al. [2]	Multi-State Merkle Patricia Trie (MSMPT)	This concept offered lightweight blockchain multi-query services that were quick and effective.	AI methods for regulating the category for every linked list element were not supported by the developed model.
Chen, H. and Liang, D., [3]	Verkle AR*-tree	This method successfully enhanced the use of the spatial index from point to the cube and enhanced the spatiotemporal query's performance.	The developed method was not well-suited for all types of queries, and required highly flexible multi-dimensional indexing.
Han, J., et al. [4]	Ethereum-based blockchain (quorum)	This technique reduced the maintenance expenses and allowed smart contracts to use range queries without the need for user-defined data structures.	This architecture failed to use an embedded relational database to save overhead and improve system performance, scalability, and efficiency.
Almi'Ani, et al. [5]	Graph-Based Profiling	The method's effectiveness was greatly enhanced by the historical data sources' information regarding the reliability and correctness of the data, and it was highly accurate in determining the reliability of the sources.	The impact of the adopted application with varying source distributions on the method's performance was not examined by the created scheme.
Wang, H., et al. [6]	vChain+	The developed approach achieved efficient query performance with integrity assurance & enabled practical public key management.	This scheme neglected to support more complex queries such as aggregation queries and join queries.
Härer, F., [7]	Cross-Chain Query Language	This approach had increased application-level interoperability, and leverage multiple blockchains to establish a unified view on data while relying on verifiable transactions.	This model neglected to focus on augmenting data storage distributed on multiple blockchains, and failed to provide advanced integration methods towards enabling blockchains as decentralized application platforms.
Bao, Y., et al. [8]	Spatial Blockchain	The utilized dual-stage verification mechanism ensured database security & allowed efficient real estate transactions.	This model did not incorporate blockchain and geographic information system (GIS) technologies, illuminating a pathway to improve the applicability of the system.

III. PROPOSED METHODOLOGY

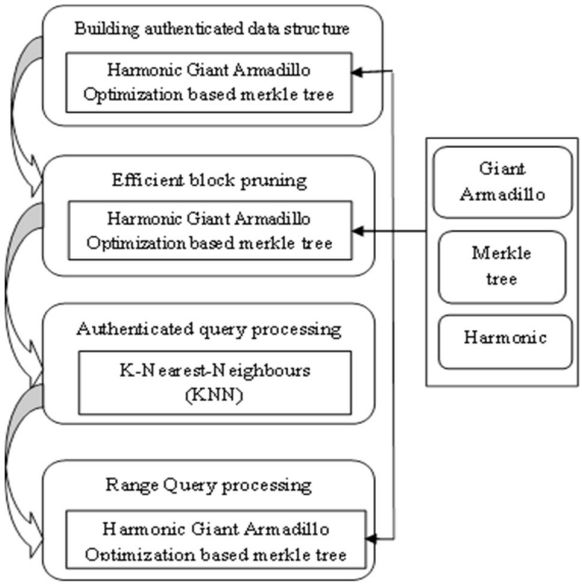

Figure 1. Block diagram HGAO_M-Tree for query processing for transaction based blockchain

The purpose of a blockchain is to create dependable and consistent agreements in a decentralized, untrusted setting. The blockchain also makes it possible to create smart contracts and perform transactions. It gives end users the ability to carry out contracts without the need for middlemen. However, conventional models had limited throughput, large computational costs, and delay, especially for single or multi-query processing, and therefore were unable to retrieve information effectively.To bridge these complications, this research proposes a Harmonic Giant Armadillo optimized merkle tree (HGAO_M-Tree) for query processing using transaction based blockchain.

Here, the proposed hybrid query processing model using transaction based blockchain consists of four steps. In first step, authenticated data structure will be developed using HGAO_M-Tree, which will be devised by the combination of Harmonic analysis [23], Giant Armadillo Optimization (GAO) [22] and Merkle tree (M-Tree) [25] . After that, efficient block pruning will be performed based on hybrid HGAO_M-Tree. In third step, authenticated query processing will be done using k-Nearest-Neighbours (KNN) [24]. In fourth step, range query processing will be done using HGAO_M-Tree. Moreover, the proposed query processing model for transaction based blockchain will be implemented in PYTHON tool. In addition, the metrics, such as Memory size, Query time, Storage capacity will be employed to validate the proposed model.

IV. Main Objectives

The major goal of this research is given as follows,

- Decentralized query processing is made possible by an effective blockchain, which also lowers the possibility of single points of failure and increases system resilience.
- To perform range query processing, Harmonic Giant Armadillo Optimization based merkle tree (HGAO_M-Tree) is used for enabling faster response times. Here, the HGAO_M-Tree is devised by combining GOA and harmonic analysis with M-Tree.
- KNN is used to perform authenticated query processing. In order to preserve data privacy, the KNN finds the nearest data points in a dataset using distance measures.

V. Discussion

Significant progress has been made in blockchain-based query systems, according to the studied literature, especially in terms of maintaining query integrity, improving performance, and accommodating specific query forms. The majority of current solutions, however, have significant drawbacks, including poor support for complex or dynamic queries (such as joins and aggregation), a lack of cross-chain interoperability, a lack of intelligent query execution optimization, and difficulties with scalability in large or dispersed datasets.

In order to overcome these drawbacks, the suggested HGAO_M-Tree architecture combines three essential elements: Merkle Tree structures for data authentication, Giant Armadillo Optimization (GAO) for intelligent query block trimming, and Harmonic Analysis for data segmentation. This synergistic design offers several important advantages over standard methods:

- Enhanced Query Efficiency: HGAO_M-Tree integrates a global optimization method via GAO, in contrast to MSMPT [2] and Verkle AR*-tree [3], which mainly concentrate on index structure enhancements. Particularly with multidimensional blockchain datasets, this enables a smaller search space and quicker retrieval in both kNN and range searches.
- Intelligent Query Pruning Support: While the majority of modern systems, like vChain+ [6], verify data integrity, they do not optimize the data retrieval path. Our approach, on the other hand, uses a pruning technique inspired by metaheuristics that dynamically removes unnecessary nodes, lowering latency and increasing system performance.
- Fault Tolerance and Scalability: While cross-chain query technologies such as CCQL [7] and spatial blockchain architectures [8] provide better interoperability, they are devoid of inherent scalability and fault tolerance features. This is addressed by HGAO_M-Tree, which uses adaptive pruning logic and the distributed nature of Merkle Trees to maintain steady performance even as the size of the blockchain grows.
- Greater Query Support: Although SQLQuery [4] and AKQ/ARQ [1] provide authenticated and SQL-style inquiries, they are not very flexible with regard to different kinds of searches. The suggested model's extensible tree-based index and optimal traversal approach enable it to accommodate a broad variety of queries, including aggregation and spatial-temporal searches.
- Security and Trustworthiness: By integrating hash-based proofs using the Merkle Tree, our architecture guarantees verifiability and data trust, building on the ideas of trust-based oracle verification [5].

Furthermore, by facilitating anomaly identification during the query phase, the use of harmonic signal-based segmentation enhances resilience to data tampering.

The HGAO_M-Tree model presents additional difficulties in spite of these advantages. One such difficulty is the GAO algorithm's computational cost during setup, which could necessitate meticulous parameter adjustment to guarantee responsiveness in real-time settings. Additionally, future research may need to focus on creating a universal query interface or middleware layer in order to integrate with other blockchain systems.

The suggested framework is a workable solution for high-performance, secure, and intelligent query processing in blockchain applications—particularly those involving transaction-heavy or cross-domain datasets like supply chain systems, smart cities, and healthcare—because it performs better overall than traditional models in terms of query flexibility, speed, and security.

VI. Conclusions

The HGAO_M-Tree, a novel hybrid query processing model for blockchain-based HTAP systems, combining Harmonic Giant Armadillo Optimization and Merkle trees. The approach enhances query performance, data integrity, and security, addressing limitations in real-time analytics and complex query handling. By leveraging blockchain's immutable structure and using k-Nearest Neighbors (KNN) for authenticated queries, the model improves efficiency and ensures robust data privacy. Performance metrics validate its effectiveness, making it a promising solution for secure, scalable, and real-time blockchain applications.

References

[1] In IEEE Transactions on Computers, vol. 72, no. 8, pp. 2209-2223, 2023, Xu, H., Xiao, B., Liu, X., Wang, L., Jiang, S., Xue, W., Wang, J., and Li, K., "Empowering authenticated and efficient queries for STK transaction-based blockchains."

[2] "Multi-state merkle patricia trie (msmpt): High-performance data structures for multi-query processing based on lightweight blockchain," by Mardiansyah, V., Muis, A., and Sari, R.F., IEEE Access, 2023.

[3] The article "Adaptive spatio-temporal query strategies in blockchain" by Chen, H. and Liang, D. was published in the ISPRS International Journal of Geo-Information in 2022.

[4] In "Design and Implementation of Enabling SQL–Query Processing for Ethereum-Based Blockchain Systems," Han, J., Seo, Y., Lee, S., Kim, S., and Son, Y., Electronics, vol. 12, no. 20, pp. 4317, 2023.

[5] [5] “Graph-based profiling of blockchain oracles,” IEEE Access, vol. 11, pp. 24995-25007, 2023; Almi’Ani, K., Lee, Y.C., Alrawashdeh, T., and Pasdar, A.

[6] "vChain+: Optimizing verifiable blockchain boolean range queries," Wang, H., Xu, C., Zhang, C., Xu, J., Peng, Z., and Pei, J., in IEEE's 38th International Conference on Data Engineering (ICDE), 2022, pp. 1927-1940.

[7] A Cross-Chain Query Language for Application-Level Interoperability Between Permissionless and Open Blockchains, Härer, F., arXiv preprint arXiv:2307.00951, 2023.

[8] In "Spatial Blockchain: Enhancing Spatial Queries and Applications through Integrating Blockchain and Spatial Database Technologies," Bao, Y., Gui, Z., Sun, Z., An, Z., and Huang, Z., Electronics, vol. 12, no. 20, pp. 4287, 2023.

[9] "Optimization of Merkle Tree Structures: A Focus on Subtree Implementation," by Ayyalasomayajula, P. and Ramkumar, M., in the 2023 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), pp. 59–67, IEEE, 2023.

[10] "Query processing in blockchain systems: Current state and future challenges," Future Internet, vol. 14, no. 1, pp. 1, 2021, Przytarski, D., Stach, C., Gritti, C., and Mitschang, B.

[11] "DHT-based range query processing for web service discovery," Zhang, Y., Liu, L., Li, D., Liu, F., and Lu, X., In 2009 IEEE International Conference on Web Services, pp. 477-484, IEEE, 2009.

[12] "Rethink query optimization in htap databases," Proceedings of the ACM on Management of Data, vol. 1, no. 4, pp. 1-27, 2023; Song, H., Zhou, W., Li, F., Peng, X., and Cui, H.

[13] "HyBench: A New Benchmark for HTAP Databases," Proceedings of the VLDB Endowment, vol. 17, no. 5, pp. 939-951, 2024, Zhang, C., Li, G., and Lv, T.

[14] In Heterogeneous Data Management, Polystores, and Analytics for Healthcare: VLDB Workshops, Poly 2020 and DMAH 2020, Virtual Event, August 31 and September 4, 2020, Revised Selected Papers 6, pp. 25-36, Springer International Publishing, 2021, Vogt, M., Hansen, N., Schönholz, J., Lengweiler, D., Geissmann, I., Philipp, S., Stiemer, A., and Schuldt, H., "Polypheny-DB: towards bridging the gap between polystores and HTAP systems."

[15] "Blockchain-based data management for digital twin of product," Huang, S., Wang, G., Yan, Y., and Fang, X., Journal of Manufacturing Systems, vol. 54, pp. 361-371, 2020.

[16] "Controllable and trustworthy blockchain-based cloud data management," Future Generation Computer Systems, vol. 91, pp. 527-535, 2019, Zhu, L., Wu, Y., Gai, K., and Choo, K.K.R.

[17] "Data tamper detection from NoSQL Database in forensic environment," by R. Chopade and V. Pachghare, Journal of Cyber Security and Mobility, pp. 421-450, 2021.

[18] Yang, Y., In Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications, pp. 1-12, IEEE, 2006; Dunlap, R., Rexroad, M., and Cooper, B.F., "Performance of full text search in structured and unstructured peer-to-peer systems."
[19] Ramkumar, M., "Using a blockchain to execute large-scale processes," Journal of Capital Markets Studies, vol. 2, no. 2, pp. 106-120, 2018.
[20] In "Multi-dimensional data indexing and range query processing via Voronoi diagram for internet of things," Wan, S., Zhao, Y., Wang, T., Gu, Z., Abbasi, Q.H., and Choo, K.K.R., Future Generation Computer Systems, vol. 91, pp. 382-391, 2019.
[21] "On spatio-temporal blockchain query processing," Future generation computer systems, vol. 98, pp. 208-218, 2019, Qu, Q., Nurgaliev, I., Muzammal, M., Jensen, C.S., and Fan, J.
[22] In "Giant Armadillo optimization: A new bio-inspired metaheuristic algorithm for solving optimization problems," Alsayyed, O., Hamadneh, T., Al-Tarawneh, H., Alqudah, M., Gochhait, S., Leonova, I., Malik, O.P., and Dehghani, M., Biomimetics, vol. 8, no. 8, pp. 619, 2023.
[23] "Estimation of trigonometric components in time series," by Damsleth and Spjøtvoll, E., Journal of the American Statistical Association, vol. 77, no. 378, pp. 381-387, 1982.
[24] In On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, Proceedings, pp. 986-996, Springer Berlin Heidelberg, 2003, Guo, G., Wang, H., Bell, D., Bi, Y., and Greer, K., "KNN model-based approach in classification."
[25] "An efficient merkle-tree-based authentication scheme for smart grid," IEEE Systems Journal, vol. 8, no. 2, pp. 655-663, 2013, Li, H., Lu, R., Zhou, L., Yang, B., and Shen, X.