A Hybrid Approach to Boolean Keyword Search: Combining B+ Tree Indexing with Distributed Query Processing

Mr. Sangeetham Lokesh[1], Mrs. T. Lavanya[2]
[1]*PG Scholar, Vemu Institute of Technology*
[2]*Assistant Professor, Dept. of CSE, Vemu institute of Technology*

***Abstract*—Cloud data warehouses (CDWs) offer scalable storage and computational efficiency, but ensuring secure keyword searchability remains a challenge. This project proposes a Boolean keyword searchable encryption (BKSE) framework incorporating Partial Homomorphic Encryption (PHE), B+Tree indexing, and blockchain verification to enhance security, accuracy, and performance. The system facilitates Boolean search queries on encrypted data, ensuring secure, verifiable, and efficient keyword retrieval. Extension concepts introduce HMAC authentication to secure network packets and a multi-VM routing mechanism for faster query responses. These enhancements significantly improve security, integrity, and query processing efficiency in cloud-based data warehousing systems.**

***Index Terms*—Searchable Encryption, Boolean Keyword Search, Cloud Data Warehouse, Data Security**

I. INTRODUCTION

Cloud-based data storage has revolutionized business intelligence and big data analytics. However, traditional searchable encryption techniques struggle with Boolean query execution and secure keyword searches. This research addresses these challenges by integrating Partial Homomorphic Encryption (PHE) with optimized indexing structures like B+Tree and inverted index mapping. Blockchain-based authentication ensures verifiability and tamper-proof search results, eliminating the need for third-party verification. The system is designed to efficiently handle large-scale encrypted databases, supporting secure search queries while maintaining data confidentiality. Extensions such as HMAC-based authentication for secure packet transmission and multi-VM query routing further enhance the framework's reliability. This paper presents an advanced model for scalable and privacy-preserving cloud data warehouse (CDW) search operations.

II. LITERATURE SURVEY

1. BPVSE: Publicly Verifiable Searchable Encryption for Cloud-Assisted Electronic Health Records (Chen et al., 2023)
- The study explores publicly verifiable searchable encryption (PVSE) for electronic health records (EHRs).
- Implements bilinear pairing-based cryptography to ensure secure keyword searches and integrity verification.
- Utilizes Bloom filters to enhance search efficiency while maintaining privacy.

2. A Practical Framework for Secure Document Retrieval in Encrypted Cloud File Systems (Fu et al., 2022)
- Introduces a hybrid cryptographic approach combining homomorphic encryption (HE) and attribute-based encryption (ABE).
- Focuses on efficient query processing in encrypted file systems, ensuring fine-grained access control.

3. CASE-SSE: Context-Aware Semantically Extensible Searchable Symmetric Encryption (Chen et al., 2023)
- Proposes context-aware symmetric searchable encryption (SSE) to enhance semantic keyword search.
- Utilizes deep learning models to refine query processing and contextual matching in cloud-based environments.

4. Device-Oriented Keyword-Searchable Encryption Scheme for Cloud-Assisted Industrial IoT (Zhou et al., 2022)

- Develops a keyword-searchable encryption framework tailored for Industrial IoT applications.
- Integrates lightweight cryptographic mechanisms and federated learning for enhanced security and efficiency.

5. DSAS: A Secure Data Sharing and Authorized Searchable Framework for e-Healthcare System (Xue, 2022)

- Introduces role-based access control and proxy reencryption mechanisms for healthcare data sharing.
- Leverages multi-keyword searchability with privacy-preserving attributes.

6. Dual Traceable Distributed Attribute-Based Searchable Encryption (Yang et al., 2023)

- Implements attribute-based searchable encryption (ABSE) with traceability mechanisms to support distributed data ownership transfer.
- Enhances data accountability through auditable access logs.

7. Efficient and Privacy-Preserving Search Over Edge–Cloud Collaborative Entity in IoT (Zhang et al., 2023)

- Proposes a privacy-preserving searchable encryption scheme for edge–cloud architectures.
- Utilizes Secure Multi-Party Computation (SMPC) to prevent unauthorized data access.

8. EMK-ABSE: Efficient Multikeyword Attribute-Based Searchable Encryption (Liu et al., 2022)

- Optimizes attribute-based encryption for multi-keyword searchability.
- Focuses on cloud-edge coordination to reduce computation overhead.

9. Enabling Verifiable and Dynamic Ranked Search Over Outsourced Data (Liu et al., 2022)

- Implements a ranking-based searchable encryption model.
- Ensures data integrity verification using blockchain-based consensus mechanisms.

10. Forward Secure Public Key Encryption with Keyword Search for Outsourced Cloud Storage (Zeng et al., 2022)

- Proposes a forward-secure keyword search encryption scheme to mitigate keyword guessing attacks.
- Utilizes public key encryption (PKE) with post-compromise security resilience.

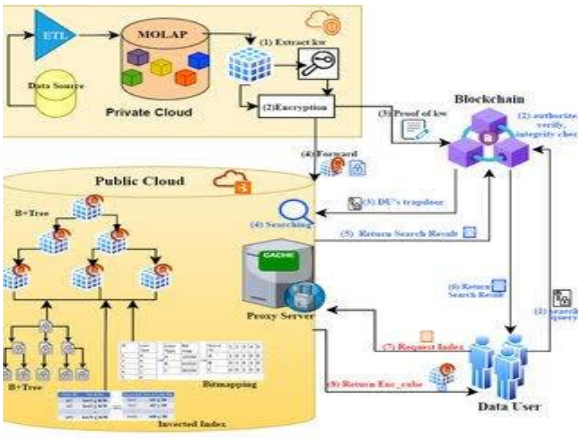
PROBLEM STATEMENT:

Conventional cloud-based search encryption systems are limited in supporting Boolean queries and fail to provide verifiable search results. This study proposes a scalable and efficient framework that ensures privacy, integrity, and high-speed keyword searches on encrypted cloud databases.

PROPOSED METHOD:

The proposed system leverages Partial Homomorphic Encryption (PHE) for secure query execution, enabling Boolean keyword searches on encrypted cloud data. A combination of B+Tree and inverted index structures ensures optimized search performance. Blockchain-based smart contracts provide authentication, verifying search results and preventing tampering. Extensions such as HMAC authentication enhance network security, while multi-VM routing improves query processing speed. This framework ensures a scalable, secure, and verifiable cloud data warehouse solution.

ARCHITECTURE:



III. METHODOLOGY

1. Data Encryption and Storage Module

Objective: To ensure the secure encryption and storage of sensitive data before being uploaded to the cloud.

Process:

The user encrypts the dataset using PHE, allowing computations to be performed without decryption.

The encrypted data is uploaded to the cloud data warehouse (CDW) and indexed securely.

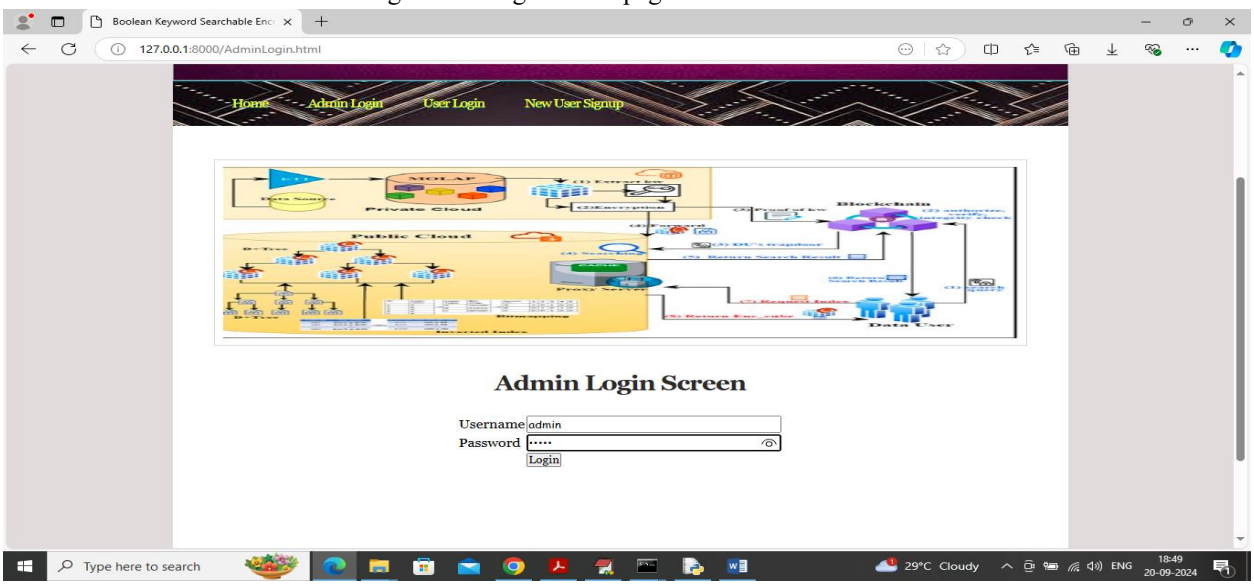
Data remains encrypted throughout its lifecycle to prevent unauthorized access.

Outcome: Enhances data security while enabling efficient keyword searches without exposing plaintext information.

2. Indexing and Query Optimization Module
Objective: To optimize search efficiency using structured indexing mechanisms.
Process:
When new encrypted data is added, B+Tree and inverted index structures update automatically.
The system maps encrypted keyword tokens to relevant data entries.
Upon receiving a search query, the index retrieves relevant encrypted records efficiently.
Outcome: Minimizes query processing latency and enhances search performance in large-scale cloud environments.
3. Secure Boolean Keyword Search Module
Objective: To enable Boolean search operations over encrypted cloud data.
Process:
The user submits an encrypted Boolean keyword query.
The system processes the query while maintaining encryption integrity.
The relevant encrypted search results are retrieved and sent back to the user.
Outcome: Allows users to perform secure and efficient Boolean searches without exposing query terms or results.
4. Blockchain-Based Authentication Module
Objective: To ensure verifiable search result integrity and prevent data tampering.
Process:
Each query request and its corresponding search result hash are stored on the blockchain.
Users can verify that search results have not been tampered with.
Blockchain-based validation ensures trust and transparency in search operations.
Outcome: Provides verifiable search results, enhancing trust and security in cloud data retrieval System Implementation
5. Extension Modules
Extension 1: HMAC Authentication for Secure Data Packets
Objective: Secure network communication and prevent packet tampering.
Process: Uses cryptographic signatures to validate data integrity in cloud transmissions.
Extension 2: Multi-VM Load Balancing for Faster Query Processing
Objective: Optimize query execution speed by distributing loads dynamically.
Process: Implements intelligent routing mechanisms for parallel query handling.
Outcome: The extensions further enhance security, efficiency, and scalability, making the system adaptable to enterprise-level data warehousing needs.

IV. RESULTS

In above screen click on 'Admin Login' link to get below page.



V. CONCLUSION

In this paper, we have presented a flexible, verifiable, and secure searchable encryption scheme with support for Boolean expression over encrypted data cubes within a cloud-based data warehouse. Our scheme enjoys both security and search performance based on the integration of partial homomorphic encryption, inverted index, and B+Tree. In addition, we leveraged blockchain technology to streamline the automation of search permission verification, user authentication, and search result validation processes. These tasks are executed in a manner that ensures scalability and immutability. Notably, we have utilized various search function types to suit different data types applicable for searching over multidimensional data, such as inverted indexes, B+Trees, and bitmapping functions. Another key advantage of our proposed B+Tree indexing scheme is to reduce the search space. Our experiments have demonstrated that our scheme can significantly save time and resources. The system can also provide reasonable system throughput for supporting multiple concurrent OLAP query requests. For future works, we will investigate the technique to achieve fully forward security in supporting the keyword update

REFERENCES

[1] H. Yin, W. Zhang, H. Deng, Z. Qin, and K. Li, ‘‘An attribute based searchable encryption scheme for cloud-assisted IIoT,’’ IEEE Internet Things J., vol. 10, no. 12, pp. 11014–11023, Jun. 2023, doi: 10.1109/JIOT.2023.3242964.

[2] X. Liu, H. Dong, N. Kumari, and J. Kar, ‘‘A pairing-free certificateless searchable public key encryption scheme for industrial Internet of Things,’’ IEEE Access, vol. 11, pp. 58754–58764, 2023, doi: 10.1109/ACCESS.2023.3285114.

[3] S. Guo, H. Geng, L. Su, S. He, and X. Zhang, ‘‘A rankable Boolean searchable encryption scheme supporting dynamic updates in a cloud environment,’’ IEEE Access, vol. 11, pp. 63475–63486, 2023, doi: 10.1109/ACCESS.2023.3284904.

[4] Y. Zheng, R. Lu, J. Shao, F. Yin, and H. Zhu, ‘‘Achieving practical symmetric searchable encryption with search pattern privacy over cloud,’’ IEEE Trans. Services Comput., vol. 15, no. 3, pp. 1358–1370, May 2022, doi: 10.1109/TSC.2020.2992303.

[5] Y. Wang, S.-F. Sun, J. Wang, J. K. Liu, and X. Chen, ‘‘Achieving searchable encryption scheme with search pattern hidden,’’ IEEE Trans. Services Comput., vol. 15, no.

[6] J. Li, X. Lin, Y. Zhang, and J. Han, ‘‘KSF-OABE: Outsourced attribute-based encryption with keyword search function for cloud storage,’’ IEEE Trans. Services Comput., vol. 10, no. 5, pp. 715–725, Sep. 2017, doi: 10.1109/TSC.2016.2542813.

[7] Q. Zhang, S. Wang, D. Zhang, J. Sun, and Y. Zhang, ‘‘Authorized data secure access scheme with specified time and relevance ranked keyword search for industrial cloud platforms,’’ IEEE Syst. J., vol. 16, no. 2, pp. 2879–2890, Jun. 2022, doi: 10.1109/JSYST.2021. 3093623.

[8] B. Chen, T. Xiang, D. He, H. Li, and K. R. Choo, ‘‘BPVSE: Publicly verifiable searchable encryption for cloud-assisted electronic health records,’’ IEEE Trans. Inf. Forensics Security, vol. 18, pp. 3171–3184, 2023, doi: 10.1109/TIFS.2023.3275750.

[9] J. Fu, N. Wang, B. Cui, and B. K. Bhargava, ‘‘A practical framework for secure document retrieval in encrypted cloud file systems,’’ IEEE Trans. Parallel Distrib. Syst., vol. 33, no. 5, pp. 1246–1261, May 2022, doi: 10.1109/TPDS.2021.3107752.

[10] L. Chen, Y. Xue, Y. Mu, L. Zeng, F. Rezaeibagha, and R. H. Deng, ‘‘CASE-SSE: Context-aware semantically extensible searchable symmetric encryption for encrypted cloud data,’’ IEEE Trans. Services Comput., vol. 16, no. 2, pp. 1011–1022, Mar. 2023, doi: 10.1109/TSC.2022. 3162266.