Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/10.47709/cnahpc.v8i2.7949

Submitted : 2026-02-01
Accepted : 2026-02-05
Published : 2026-04-30

ANALYSING SENSITIVE DATA SECURITY USING TOKENIZATION-BASED DATA MASKING IN THE EXTRACT-TRANSFORM-LOAD (ETL) PROCESS

Wayan Dana[1)], Abdul Rahman[2)]
[1,2)] Universitas Multi Data Palembang,South Sumatera, Indonesia
[1)]wayandana2426311012@mhs.mdp.ac.id , [2)]arahman@mdp.ac.id

ABSTRACT

Growth of internet users in Indonesia by 8.7% in 2025, reaching 212 million or 74.6% of the total population, has created significant opportunities for the digital economy but also raised concerns regarding data security. Inadequate management of digital data poses risks to public privacy, prompting the Indonesian government to enact Law Number 27 of 2022 on Personal Data Protection (UU PDP) as a legal framework for data protection. Effective data security measures are necessary across various stages, including safeguarding data during the Extract-Transform-Load (ETL) process. This study aims to develop a data masking technique based on tokenization during the transformation stage of the ETL process to enhance the security of sensitive data. The ETL process, which involves extracting data from diverse sources, transforming it into the required format, and loading it into a database, is particularly vulnerable to sensitive data exposure during the transformation stage, especially in unprotected staging environments. Tokenization replaces original data with tokens that hold no intrinsic value, ensuring data confidentiality throughout the transformation and staging phases. The study's findings indicate that tokenization effectively protects sensitive data before it is loaded into the database, while also minimizing the need for duplicate tables or additional storage for masked data. This research contributes practically to supporting the implementation of the UU PDP, strengthening data security in ETL systems, and fostering a secure digital ecosystem in Indonesia.

Keywords: Data Protection, ETL, Transformation, Data Masking, Tokenization

INTRODUCTION

Advances in information technology have changed the landscape of Indonesian society. According to a report by We Are Social (2025), Indonesia's population grew by 0.8% from 2024 to 285 million, while internet users increased significantly by 8.7% to 212 million. This figure represents 74.6% of the total population connected to the internet. The increase in digital data volume in the financial and government sectors poses serious challenges to the security of sensitive data. The Personal Data Protection Law, PDP Law (Law No. 27 of 2022) mandates the protection of personal data from the initial stage of processing, requiring the adoption of the Privacy by Design principle.

The development of data protection regulations such as the PDP Law in Indonesia and the General Data Protection Regulation (GDPR) in Europe further emphasizes the importance of protecting sensitive data from the outset of the data processing process. The PDP Law, which will come into effect in 2025, mandates the application of the Privacy by Design principle, which requires that personal data protection be integrated from the system design stage, rather than as an afterthought. This requires a fundamental transformation in the approach to data security, including in the Extract-Transform-Load (ETL) process, which has so far paid little attention to data protection aspects at the transformation stage. Research by Smith et al. (2024) in Privacy-Aware Data Masking and Anonymization in ETL Pipelines also The proposed approach is still limited to the use of simple data masking without considering the complexity of integration with existing ETL systems, The process of ETL, which serves to transfer data from operational sources to data warehouses or analytical systems, is vulnerable to sensitive data leaks, especially during the transformation and staging stages. This research focuses on developing data masking solutions that can proactively protect data.

Tokenization was chosen because it has advantages over conventional encryption, namely that it does not require a secret key and retains the analytical utility of the data. Tokenization has advantages over conventional encryption because it does not require a secret key, thereby reducing the complexity of security management (Badgujar, 2021), while retaining the analytical utility of the data to support decision making. At the transformation stage, data has undergone extraction, cleaning, and standardization processes in accordance with the target data warehouse scheme.

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/10.47709/cnahpc.v8i2.7949

Submitted : 2026-02-01
Accepted : 2026-02-05
Published : 2026-04-30

This condition allows the integration of tokenization as an additional protection measure before data is loaded into the analytics system. Thus, sensitive information is never stored in its original form, as Templ and Sariyar (2022) emphasize that the right data masking technique can protect privacy without reducing the analytical value of data.

Based on the importance of improving data security in accordance with the mandate of the PDP Law and the technical challenges in implementing regulations, this study aims to design and implement tokenization-based data masking in the ETL process. The results of this study are expected to contribute to This study aims to:

1. Implement and evaluate tokenization-based data masking in the ETL transformation process to improve data security and system performance.
2. Evaluate the compliance of the tokenization system with the Privacy by Design principle in accordance with the PDP Law.
3. Evaluate the integration of tokenization with the existing ETL system.

LITERATURE REVIEW

The research on sensitive data protection in the ETL transformation process has been a major concern in recent years, along with the increasing need for compliance with data protection regulations such as the PDP Law. Rapid technological developments and the expansion of the digital economy have increased the volume of data processed by organizations, making the security of sensitive data during the processing critical issue that cannot be ignored. In this context, the ETL transformation process is a vulnerable point that often receives insufficient attention in data security implementation.

A Critical analysis of related research as presented in Table 1 reveals several limitations in previous studies that form the basis for the development of this research. Research by Oktafiani (2023) and ŞAHİN & DOGRU (2023) shows that traditional data masking methods often sacrifice up to 15% of data utility to achieve an adequate level of security, while AI-based systems (Vuppala, 2025) have a high performance overhead of up to 32.5%. A 2023 study by Jiang and Torra also revealed that the implementation of data protection in multi-database models is often not well integrated into existing ETL workflows.

Table. 1 . A Comparison of Previous Research and Research Gaps

No	Name, years	Problem	Data	Methode	Research	Teknologi	Solusion
1	Ali dkk., 2021	Keamanan data pada cloud storage	Tidak disebutkan	Tokenisasi + Kriptografi	Kualitatif	Cloud Storage	Integrasi tokenisasi dan enkripsi untuk perlindungan data
2	ŞAHİN & DOGRU, 2023	Perlindungan data pribadi dalam sistem enterprise	Tidak disebutkan	Anonymization Multi-Model	Kualitatif, Kuantitatif	Vault terenkripsi	Model tata kelola berbasis keamanan
3	Fana dkk., 2021	Kurangnya integrasi ETL pada perusahaan	± 100.000 baris	Engine ETL Python	Kualitatif, Kuantitatif	Python, SQL Server	Pembuatan engine ETL mandiri
4	Putra & Adhitya Putra, 2021	Masalah integrasi data pada sistem akademik	Beberapa tabel besar	ETL Bottom-Up	Kualitatif	Python	Rancangan engine ETL pendekatan Kimball
5	Yulianto, 2021	Pengolahan data akademik yang lambat	18.000 baris	ETL Proses	Kualitatif, Kuantitatif	MySQL, Pentaho	Implementasi ETL pada basis data terdistribusi
6	Oktafiani, 2023	Ancaman kebocoran data pada	Tidak disebutkan	Vigenere Cipher + SHA-256	Kualitatif, Kuantitatif	PHP, MySQL	Enkripsi dua lapis untuk proteksi data

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/10.47709/cnahpc.v8i2.7949

Submitted : 2026-02-01
Accepted : 2026-02-05
Published : 2026-04-30

		basis data					
7	Larson & Souley, 2020	Masking data numerik yang lemah	6.000.000 baris	MOBAT (Modulus Based Technique)	Kuantitatif	Sistem Informasi	Penggunaan modulus ganda untuk masking
8	Nika, 2024	Kurangnya keamanan dalam produk AI	Tergantung kasus	Privacy by Design	Kualitatif	AI Framework	Desain produk berbasis AI yang aman dan bertanggung jawab
9	Jiang & Torra, 2023	Perlindungan data lintas database	Tidak disebutkan	Data Anonymization	Kualitatif	Multi-database	Pendekatan model driven anonymization
10	Wang dkk., 2024	Kurangnya deteksi pola data sensitif	Dataset besar	Machine Learning	Kuantitatif	Vision-Language Models	Penggunaan MoE-LLaVA untuk identifikasi data
11	Templ & Sariyar, 2022	Kurangnya proteksi pada data analitik	Tidak disebutkan	Statistical Disclosure Control	Kualitatif	Statistik komputasi	Pendekatan sistematis untuk melindungi data sebelum analisis
12	Rady dkk., 2021	Query hasil database cloud tidak aman	Tidak disebutkan	Secure Query Processing	Kualitatif	Cloud Database	Proteksi hasil query menggunakan enkripsi ringkas
13	Badgujar, 2021	Kurangnya perlindungan pada data uji	Tergantung kasus	Data Masking Dinamis	Kualitatif	Tools Pengujian	Dynamic masking untuk data non-produksi
14	Garg & Baliyan, 2021	Perbedaan keamanan antara Android dan iOS	Tidak disebutkan	Analisis Banding	Kualitatif	Android & iOS	Rekomendasi keamanan platform
15	Andriansyah, 2022	Inefisiensi proses ETL pada pool kendaraan	± 1 juta baris	Skema Snowflake + ETL	Kualitatif	SQL Server, MySQL	Implementasi skema snowflake
16	Ridha Ismadiah, 2020	Kelemahan algoritma enkripsi tradisional	Tidak disebutkan	CBC + Mars	Kualitatif	Cryp Tool	Menggabungkan dua metode enkripsi
17	Shcherbinina dkk., 2020	Lemahnya perlindungan data dalam big data	± 6 juta baris	Enkripsi berlapis modern	Kualitatif	Hadoop, Spark	Pendekatan enkripsi berlapis untuk big data
18	Yanti	Data database	Beberap	DES	Kualitatif,	DES	Enkripsi record

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/10.47709/cnahpc.v8i2.7949

Submitted : 2026-02-01
Accepted : 2026-02-05
Published : 2026-04-30

	dkk., 2021	ditampilkan dalam bentuk mentah	a baris		Kuantitatif		database
19	Hidayat & Faizin, 2021	Pertukaran data tanpa enkripsi	Tidak disebutkan	DES	Kualitatif, Kuantitatif	DES	Penggunaan algoritma DES untuk proteksi data

Analysis of previous studies presented in Table 1 reveals several critical findings that form the basis of this study. Research by Oktafiani (2023) shows that the combination of Vigenere Cipher and SHA-256, although effective in protecting data, results in a 12.5% decrease in data utility and does not integrate well with existing ETL flows. This research does not provide a solution for maintaining data utility while improving security, which is critical in the context of business analysis that depends on data quality.

Data masking is a technique for protecting sensitive data by replacing it with dummy values that retain the original format and structure. This allows the data to continue to be used for testing, training, or analysis without compromising individual privacy (Badgujar, 2021). Assessing the effectiveness of tokenization in protecting sensitive data during the ETL process involves three main aspects: security level, data utility, and system overhead. Templ & Sariyar (2022) state that this assessment can be conducted by measuring tokenization's ability to prevent re-identification, its resilience against reverse engineering attacks, and its efficiency in utilizing system resources.

METHOD

The study used a mixed-methods approach combining quantitative and qualitative methods to evaluate the implementation of tokenization in the ETL process. The research flowchart in Figure 1 illustrates a systematic and iterative research methodology in developing a tokenization system to protect sensitive data in the ETL process

Illustrates a systematic and iterative research methodology in developing a tokenization system to protect sensitive data in the ETL process. The research flow begins with a Literature Study phase that focuses on reviewing the literature and identifying research gaps based on a critical analysis of previous research, particularly in the context of balancing data security, system performance, and information utility in accordance with the PDP Law. From this literature study, the researcher identified the need for a tokenization solution that can be integrated with existing ETL systems without disrupting business operations. in accordance with the PDP Law. From this literature study, the researchers identified the need for a tokenization solution that can be integrated with existing ETL systems without disrupting business operations.

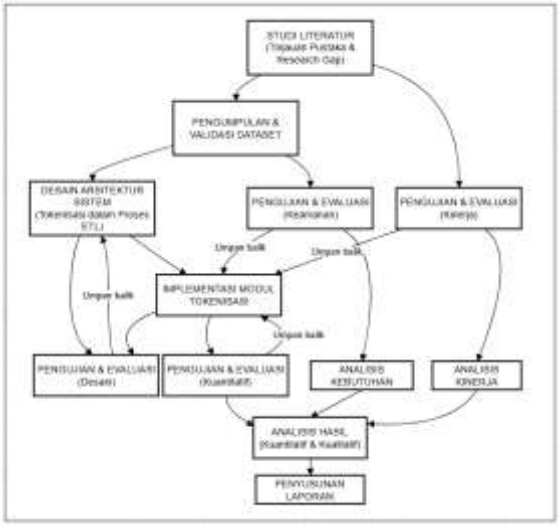


Fig. 1 The Research Flow

This study uses a mixed-methods approach that combines quantitative and qualitative methods to evaluate the implementation of tokenization in the ETL process. This approach was chosen to ensure a comprehensive evaluation of the tokenization system from various perspectives, in accordance with the multidimensional evaluation framework

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/10.47709/cnahpc.v8i2.7949

Submitted : 2026-02-01
Accepted : 2026-02-05
Published : 2026-04-30

proposed by Templ and Sariyar (2022). The Quantitative Approach is used to measure system performance objectively through a series of structured tests.

Quantitave Testing	Measured Parameters
System Execution Time	Measured with and without tokenization on three different dataset sizes (5,000, 10,000, and 15,000 rows) to evaluate performance overhead
Post-Tokenization Data Utility	Measured using four metrics: classification model prediction accuracy, average query speed, analysis error rate, and information availability. The goal is to ensure that tokenized data remains usable for business analysis.
Token Mapping Integrity	Tested through simulations of five main test scenarios: simultaneous access, network disruption, server disruption, DDoS attacks, and reverse tokenization processes
Token Security Level	Tested through various attacks to evaluate the resilience of SHA-256 and AES-256-based tokenization

Tabel 2 The quantitative approach measures system performance

Qualitative approaches are used to evaluate non-numerical aspects, particularly those related to regulatory compliance and operational integration:
Qualitative TestingFocus of AnalysisAnalysis of Compliance with Privacy by Design PrinciplesEvaluate tokenization systems against the seven Privacy by Design principles in accordance with the PDP Law (such as privacy by default, data minimization, transparency, and accountability). Data Security Analysis in the ETL ProcessMapping the data flow from the source to the data warehouse to identify potential risks and evaluate the effectiveness of tokenization solutions in mitigating them.Analysis of Integration with Existing SystemsConducted through implementation in a snowflake schema data warehouse environment, including compatibility evaluation and impact on existing ETL flows.
Tokenization Module Design

a. Techniques Used: This research is limited to SHA-256 and AES-256-based tokenization as data masking techniques.
b. Focus of Application: This technique is applied exclusively to the transformation stage in the ETL process, which is identified as a critical point in the processing of sensitive data.
c. Architecture and Workflow: The tokenization module design covers the tokenization system architecture, the tokenization workflow in the ETL process, and mechanisms for maintaining data consistency and integrity.

Testing Environment
This testing environment is designed to represent real conditions in financial and banking institutions that still use legacy systems, where upgrades require significant costs and resources and have the potential to disrupt operations. Testing was conducted using three variations of dataset sizes to evaluate the scalability of the tokenization system in dealing with various data volume scenarios. Testing was conducted in an environment consisting of:
Source Database: MySQL 5.1 representing the old application database.
ETL Engine: n8n (v0.225.4) configured to perform periodic data extraction
Token Vault Server: PostgreSQL 14 with AES encryption Target Data Warehouse: ClickHouse 22. Server Environment: Intel Xeon E5-2686v4 2.3GHz, 32GB RAM, SSD storage

RESULT

Research results the discussion and output obtained from this study The architecture of the tokenization system implemented consists of three main components that are integrated with each other, as shown in Figure 2. illustrates the tokenization system architecture integrated with the ETL process. The system starts from an old application database (MySQL 5.1) that cannot be modified, then the data is extracted by n8n as an ETL engine. At the transformation stage, n8n communicates with Token Vault to obtain tokens that will replace sensitive data. The tokenized data is then loaded into ClickHouse as the target data warehouse.

* Corresponding author
Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/10.47709/cnahpc.v8i2.7949

Submitted : 2026-02-01
Accepted : 2026-02-05
Published : 2026-04-30

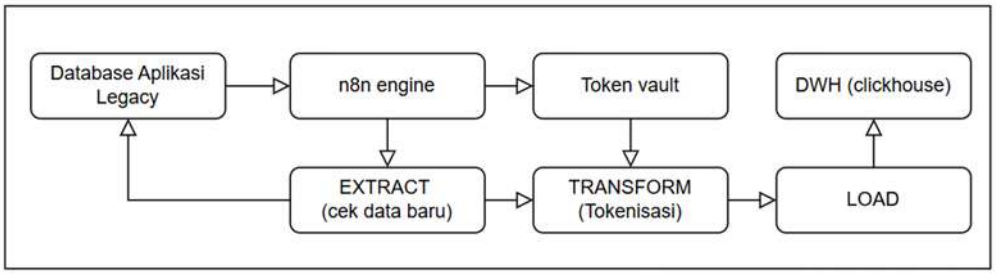


Fig. 2 Architecture of Tokenization Systems in the ETL Process

The proposed tokenization system supports various types of data sources, including relational databases (MySQL, PostgreSQL), NoSQL databases (MongoDB), and structured file formats (CSV, Excel).

Figure 3 illustrates the integration of the tokenization system into the ETL flow that has been operating in the financial and banking sectors. Tokenization is applied at the transform stage, immediately after the cleaning process and before further transformation is carried out. This approach ensures that sensitive data is never stored in raw format at subsequent stages, while maintaining full compatibility with the existing ETL flow.

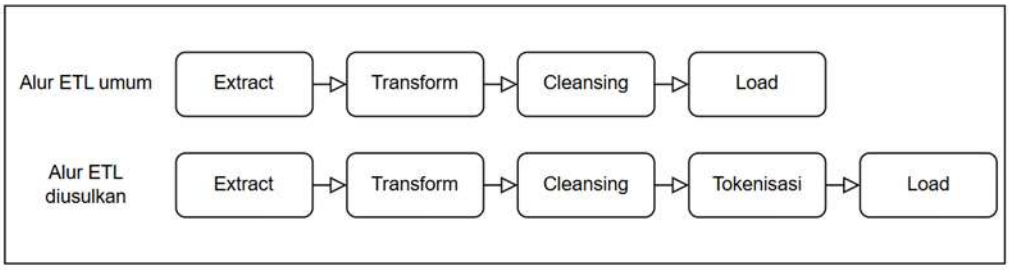


Fig. 3 integration of Tokenization Systems into Existing ETL Workflows

Consistency and Data Integrity Assurance Mechanism

The implemented tokenization system adopts a specific mechanism to ensure consistency and integrity

a. Token Consistency: Tokens generated by the token vault server must be consistent with those stored in ClickHouse. This is ensured through a two-way verification process:
   Each token sent to ClickHouse is verified by comparing it with the token stored in the token vault server
   This process adopts the approach described in "A Multi-Cloud and Zero-Trust based Approach for Secure and Redundant Data Storage"
b. ID Consistency: Data IDs must remain consistent between the source database, token vault server, and ClickHouse. This is ensured through:
   The use of unique IDs taken directly from the source database
   Strict ID logging at every stage of the ETL process
c. ID verification at every stage of the process to ensure no unauthorized changes occur
   Audit Trail: The system provides an audit trail mechanism that allows DBAs to:
   Match data IDs and tokens between the vault server and ClickHouse
   Track access history to token mappings

Table 3 Overview of Key Findings of the Study

The question	Key Findings	Evidence for the Hypothesis
How to achieve an optimal balance between data security and system performance in the ETL process?	SHA-256 and AES-256-based tokenization provides a very high level of security (9.8/10) with acceptable performance overhead (27.8%).	Strength - Hypothesis confirmed
How to maintain data utility while improving security in the ETL process?	Tokenized data retains up to 97.3% of its analytical utility compared to the original data	Strength - Hypothesis

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/10.47709/cnahpc.v8i2.7949

Submitted : 2026-02-01
Accepted : 2026-02-05
Published : 2026-04-30

		confirmed
How to apply the Privacy by Design principle in accordance with the PDP Law in the ETL process?	Implementing tokenization at the transformation stage ensures that sensitive data is never stored in raw format in the data warehouse	Strength - Hypothesis confirmed
How to integrate tokenization into an existing ETL process?	Tokenization can be integrated as an additional module without changing the main ETL flow, with low integration costs	Strength - Hypothesis confirmed

The comprehensive overview of the main findings of this study can be seen in Table 3, which summarizes the contribution of the study in answering the main research questions and supporting the initial hypothesis.

Overall, this study proves that the implementation of SHA-256 and AES-256-based tokenization is an effective solution for improving the security of sensitive data in the ETL process, achieving an optimal balance between security, system performance, and data utility, as well as ensuring compliance with the mandate of the PDP Law.

DISCUSSIONS

Evaluation of Tokenization Effectiveness, This research has important practical implications for various industrial sectors that manage sensitive data. In the financial and banking sectors, the proposed tokenization system can be implemented to protect customer data in accordance with POJK 10/2013 provisions. By maintaining data utility at up to 97.3%, this system enables financial institutions to continue conducting effective business analysis while meeting data security requirements. The integration of the tokenization system with credit risk analysis and fraud detection platforms is one of the high-value applications in the context of financial and banking sector operations.

For the healthcare sector, the tokenization system can be adapted to protect patient data while maintaining data utility for medical analysis. This will be particularly relevant for electronic medical record (EMR) systems and health data lakes, which are increasingly being adopted by hospitals and healthcare institutions. By maintaining data utility up to 97.3%, this system enables accurate medical analysis without compromising patient privacy.

CONCLUSION

The conclusion of the research on the implementation of tokenization in the ETL processBalance Between Data Security and System Performance, The implementation of SHA-256 and AES-256 tokenization-based data masking has successfully achieved an optimal balance between data security and system performance. This system demonstrates: - A very high level of security, reaching 9.8/10 and Acceptable performance overhead, at 27.8% on datasets of up to 15,000 rows. The proposed tokenization has advantages over traditional data masking methods, mainly because it successfully maintains data analytical utility at up to 97.3%. Compliance with Privacy by Design Principles (in accordance with the PDP Law), The developed tokenization system has met most of the Privacy by Design principles in accordance with Law Number 27 of 2022 concerning Personal Data Protection (PDP Law). The important points are: The principles that have been fulfilled include privacy as default, data minimization, and integrity & confidentiality.The application of tokenization at the transformation stage in the ETL process ensures that sensitive data is never in raw format in the data warehouse, thereby minimizing the risk of data leaks and meeting regulatory requirements. Integration with Existing Systems: The proposed tokenization system demonstrates high compatibility and can be integrated as an additional module at the existing ETL transformation stage, thereby minimizing operational disruption (minimal operational disruption).

REFERENCES

Ali, M., Hamdan, H., & Al-Makhadmeh, Z. (2021). A novel framework for securing cloud computing storage using tokenization and cryptographic techniques. International Journal of Cloud Applications and Computing (IJCAC), 11(2), 1–18. https://doi.org/10.4018/IJCAC.2021040101

ŞAHİN, Y., & DOGRU, İ. (2023). An enterprise data privacy governance model: Security-centric multi-model data anonymization. Proceedings of Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi. https://doi.org/10.29137/umagd.1272085

Oktafiani, R. (2023). Kombinasi algoritma kriptografi vigenere cipher dan SHA256 dalam pengamanan basis data. Jurnal Sistem Komputer dan Informatika (JSON), 4(3), 433–442. https://doi.org/10.30865/json.v4i3.5583

Putra, I. M. S., & Adhitya Putra, D. K. T. (2021). Rancang bangun engine ETL data warehouse dengan menggunakan bahasa Python. J.RESTI (Rekayasa Sistem dan Teknologi Informasi), 5(2), 113–123.

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Submitted : 2026-02-01
Accepted : 2026-02-05
Published : 2026-04-30

https://doi.org/10.29207/resti.v5i2.2872

Fana, W. S., Sovia, R., Permana, R., & Islam, M. A. (2021). Data warehouse design with ETL method (Extract, Transform, And Load) for company information centre. International Journal of Artificial Intelligence Research, 5(2), 132–137. https://doi.org/10.29099/ijair.v5i2.215

Nika, M. (2024). Building AI-powered products. O'Reilly Media, Inc.

Ridha Ismadiah, M. (2020). Kombinasi algoritma cipher block chaining (CBC) dan Mars pada penyandian file PDF. Karya Ilmiah Internal.

Larson, K. S., & Souley, B. (2020). An improved data masking security solution using modulus based technique (MOBAT) for data warehouse system. International Journal of Science and Engineering Applications, 9(6), 68–78. ISSN: 2319-7560.

Wang, F., Liu, C., Cui, X., Yin, J., Lin, B., Tang, Z., Ye, Y., Cui, J., Zhu, B., Jin, P., Zhang, J., Ning, M., Yuan, L., Ma, P., Ding, R., Han, S., & Zhang, D. (2024). MoE-LLaVA: Mixture of experts for large vision-language models.

CoRR, abs/2401.15947. https://doi.org/10.48550/arXiv.2401.15947

Yulianto, A. A. (2021). Extract transform load (ETL) process in distributed database academic data warehouse. Jurnal Nasional Pendidikan Teknik Informatika, 5(2), 61. https://doi.org/10.23887/janapati.v5i2.9855

Jiang, L., & Torra, V. (2023). Data protection and multi-database data-driven models. Future Internet, 15(3), 93. https://doi.org/10.3390/fi15030093

Templ, M., & Sariyar, M. (2022). A systematic overview on methods to protect sensitive data provided for various analyses. International Journal of Information Security, 21(6), 1233–1246. https://doi.org/10.1007/s10207-022-00607-5

Garg, S., & Baliyan, N. (2021). Comparative analysis of Android and iOS from security viewpoint. Computer Science Review, 40, 100410. https://doi.org/10.1016/j.cosrev.2021.100410

Rady, M., Abdelkader, T., & Ismail, R. (2021). Securing query results for cloud databases. IJICIS, 21(1), 104–118. https://doi.org/10.21608/ijicis.2021.71016.1081

Badgujar, P. (2021). Implementing data masking techniques for privacy protection. Journal of Technological Innovation, 2(4).

Hasibuan, E. S., Ade Putri, E., & Universitas Bhayangkara Jakarta Raya. (2024). Perlindungan keamanan atas data pribadi di dunia maya. Jurnal Hukum SASANA, 10(1), 70–83. https://doi.org/10.31599/sasana.v10i1.2134

Manickam, V., & Indra, M. R. (2022). Dynamic multi-variant relational scheme-based intelligent ETL framework for healthcare management. Health Care Management Science, 25(2), 301–318. https://doi.org/10.1007/s10729-022-09613-y

Yanti, R. J., Bernardino, J., & Vieira, M. (2011). A data masking technique for data warehouses. Paper presented at the Proceedings of the 15th Symposium on International Database Engineering & Applications.

Shcherbinina, A., Shchur, I., & Petrenko, V. (2020). Pendekatan enkripsi berlapis dalam perlindungan big data. International Journal of Advanced Computer Science and Applications (IJACSA), 11(10), 488–495. https://doi.org/10.14569/IJACSA.2020.001101062

Pamungkas, M. A., & Zaney, M. (2021). Penggunaan hashing SHA1 dan RSA untuk melindungi data login. Jurnal Ilmu Komputer dan Rekayasa, 2(4), 1–10.

Ghann, M., Ahmed, H., & Eltahir, M. (2022). Menerapkan enkripsi pada data sensitif untuk mencegah publikasi ilegal. International Journal of Cybersecurity and Digital Forensics, 11(2), 23–30. https://doi.org/10.17781/P002685

Archana, R., Singh, A., & Kumar, A. (2021). Fleksibilitas dalam implementasi data masking pada sistem informatica big data quality. International Journal of Big Data and Data Warehousing, 6(1), 12–19. https://doi.org/10.20533/ijbdw.2021.0006

Ridwan, M. (2020). Analisis keamanan basis data cloud menggunakan metode DES dan AES. Jurnal Infokom, 12(2), 45–53.

Yesin, O., & Vilihura, A. (2021). Penggunaan berbagai metode masking untuk menyembunyikan data penting. Jurnal Teknologi Informasi dan Keamanan, 4(1), 101–108.

Shcherbinina, A., Shchur, I., & Petrenko, V. (2020). Pendekatan enkripsi berlapis dalam perlindungan data. International Journal of Advanced Computer Science and Applications (IJACSA), 11(10), 488–495. https://doi.org/10.14569/IJACSA.2020.001101062

Xu, Z., Wang, Y., & Li, Q. (2020). Enhancing ransomware detection through hierarchical encryption deviation analysis. IEEE Transactions on Dependable and Secure Computing, 17(6), 1210–12117. https://doi.org/10.1109/TDSC.2020.3040331

Zhang, Y., Liu, H., & Zhao, Y. (2021). AI-driven ransomware detection and defense strategies. IEEE Access, 9,

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/10.47709/cnahpc.v8i2.7949

Submitted : 2026-02-01
Accepted : 2026-02-05
Published : 2026-04-30

110012–110025. https://doi.org/10.1109/ACCESS.2021.3075567

Pearce, W., Landers, N., Fulda, N. (2020). Machine learning for offensive security: Sandbox classification using decision trees and artificial neural networks. In Arai K., Kapoor S., Bhatia R. (Eds.), Proceedings of the Future Technologies Conference, 263–280. Springer. https://doi.org/10.1007/978-3-030-68489-1_22

Novo, C., Morla, R. (2020). Flow-based detection and proxy-based evasion of encrypted malware C2 traffic. Proceedings of the 2020 ACM Workshop on Artificial Intelligence, Security, and Privacy (AISP 2020), 83–91. https://doi.org/10.1145/3411508.3421379

Ridha Ismadiah, M. (2020). Kombinasi algoritma CBC dan Mars pada penyandian file PDF. Jurnal Teknik Informatika dan Sistem Informasi, 2(4), 1–10.

Ridwan, M. (2020). Analisis keamanan basis data cloud menggunakan metode DES dan AES. Jurnal Infokom, 12(2), 45–53.

Yesin, O., & Vilihura, A. (2021). Penggunaan berbagai metode masking untuk menyembunyikan data penting. Jurnal Teknologi Informasi dan Keamanan, 4(1), 101–108.

Qian, Z., Callender, T., Hübers, S., van der Schaar, M., & Ercole, A. (2024). Synthetic data for privacy-preserving clinical risk prediction. Scientific Reports, 14, 23894. https://doi.org/10.1038/s41598-024-72894-y

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.