IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105349 |

# Encryption and Obfuscation Techniques for Data Security

**Kartikay Singh, Pawan Singh**

Department of Computer Science & Engineering, Amity School of Engineering & Technology, Amity University, Lucknow, Uttar Pradesh, India

**ABSTRACT:** Data obfuscation for patient data stored on AWS Redshift is a critical aspect of guarding patient sequestration and complying with data protection regulations, similar as the Health Insurance Portability and Responsibility Act (HIPAA) in the United States. The design aims to develop and apply effective data obfuscation ways to ensure that sensitive case data stored in the AWS Redshift database remains secure and nonpublic. The design involves using colorful obfuscation styles, including masking, encryption, and tokenization, to transfigure the case data in the AWS Redshift database into a format that isn't fluently identifiable or rear- engineerable. These ways are applied to specific data fields, similar as patient names, addresses, social security figures, and other identifiers, to cover the sequestration of cases and help implicit data breaches. One of the crucial challenges in this design is to strike a balance between data security and data usability. While the data must be blurred to help unauthorized access, it should also retain its integrity and utility for authorized druggies, similar as healthcare providers and experimenters, who bear access to certain data for licit purposes. This requires careful consideration of the type of obfuscation ways used, the extent of data metamorphosis, and the access controls applied to different stoner places. The design also focuses on the perpetration of data obfuscation within the AWS Redshift terrain. This may involve exercising erected- in features of Redshift, similar as Amazon Redshift Spectrum, to perform data metamorphoses on the cover, or using AWS Identity and Access Management (IAM) to manage access controls and warrants for different druggies and operations penetrating the Redshift cluster.

**KEYWORDS**: AWS, Obfuscation, redshift, cluster, masking, encryption.

## I. INTRODUCTION

The protection of patient data in healthcare surroundings is of utmost significance to ensure patient sequestration, misbehave with data protection regulations, and maintain trust in the healthcare assiduity. With the added use of all-grounded data warehousing results, similar as Amazon Web Services (AWS) Redshift, it's pivotal to apply effective data obfuscation ways to secure sensitive case data stored in these databases. This design aims to develop and apply a data obfuscation strategy for patient data stored on AWS Redshift, using colorful obfuscation ways, similar as masking, encryption, and tokenization, to transfigure the data into a format that isn't fluently identifiable or rear- engineerable.

Introduction to obfuscation:

The protection of patient data in healthcare surroundings is of utmost significance to ensure patient sequestration, misbehave with data protection regulations, and maintain trust in the healthcare assiduity. With the added use of cloud-grounded data warehousing results, similar as Amazon Web Services (AWS) Redshift, it's pivotal to apply effective data obfuscation ways to secure sensitive case data stored in these databases. This design aims to develop and apply a data obfuscation strategy for patient data stored on AWS Redshift, using colorful obfuscation ways, similar as masking, encryption, and tokenization, to transfigure the data into a format that isn't fluently identifiable or rear- engineerable.

The main ideal of obfuscation is to help or minimize the threat of data re-identification, which is the process of relating or linking preliminarily anonymous data to specific individualities or realities. Data re-identification can pose significant sequestration pitfalls, as it may lead to unintended exposure of sensitive information, identity theft, or other vicious conditioning. Obfuscation ways are designed to disrupt the original data patterns or characteristics, making it grueling or insolvable to re-identify the data without proper authorization.

There are colorful types of obfuscation ways that can be applied to different types of data, including numerical data, categorical data, textbook data, and multimedia data. These ways can be astronomically distributed into several types grounded on the system used to befog the data, similar as negotiation, pseudonymization, masking, Generalization,

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105349 |

shuffling, redaction, and encryption. Each type of obfuscation fashion has its own strengths, limitations, and use cases, and the selection of applicable ways depends on the specific conditions and environment of the data integration design.

1. Substitution: Substitution is a common obfuscation fashion that involves replacing the original data values with fictional or aimlessly generated values. For illustration, in a healthcare environment, patient names, addresses, phone figures, or other identifiable information can be substituted with fictional names, addresses, phone figures, or other values that don't correspond to real individualities. Negotiation can be applied to colorful types of data, including numerical data, categorical data, and textbook data. Negotiation is a simple and effective fashion that can disrupt the original data patterns and help re-identification, but it may not be suitable for all types of data or use cases, as the fictional values used for negotiation shouldn't inadvertently introduce new patterns or connections that could be exploited tore-identify the data.

2. Pseudonymization: Pseudonymization is a fashion that involves replacing direct identifiers, similar as names, addresses, or phone figures, with unique canons or aliases. Aliases are aimlessly generated or pre-assigned canons that are used to represent the original data values in a way that doesn't directly reveal the identity of the individualities or realities. Pseudonymization is generally used in healthcare settings to cover patient sequestration while maintaining the usability of the data for exploration, analysis, or other licit purposes. Pseudonymization can be reversible, meaning that the original data values can be recovered using a mapping or key, or unrecoverable, meaning that the original data values are permanently replaced with aliases. Pseudonymization can give a advanced position of sequestration protection compared to negotiation, as it eliminates the threat of inadvertently introducing new patterns or connections that could be exploited for re-identification.

3. Masking: Masking is a fashion that involves caching or obscuring certain corridors of the data while leaving other corridors unchanged. Masking can be applied to colorful types of data, including numerical data, categorical data, and textbook data. Masking ways include styles similar as data truncation, data scrabbling, or data blurring, which alter the original data values in a way that prevents or minimizes the threat of reidentification. For illustration, in a healthcare environment, patient names can be incompletely masked by replacing certain characters with asterisks or other symbols, similar as" John Doe">" Jn De". Masking can be reversible or unrecoverable, depending on the specific system used. Masking is frequently used when certain corridors of the data need to be defended, similar as sensitive information like social security figures, while leaving other corridors unchanged for analysis or other purposes. Masking can be effective in precludingre-identification, but it should be precisely enforced to ensure that the masked data still maintains its usability and mileage for the intended purposes.

4. Generalization: Generalization is a fashion that involves replacing specific data values with further generalized or added up values. Generalization is generally used in data integration systems where the thing is to combine data from different sources while guarding the sequestration of the original data. For illustration, in a healthcare environment, patient periods can be generalized into age groups (e.g., 20- 29, 30- 39, etc.) or zip canons can be generalized into broader geographical regions(e.g., countries, countries). Generalization can reduce the threat of re-identification by removing the fine- granulated details of the data, while still furnishing useful information for analysis or other purposes. Generalization can be reversible or unrecoverable, depending on the position of detail retained in the generalized data. Generalization can be an effective fashion for sequestration protection in data integration systems, but it should be precisely applied to ensure that the generalized data still maintains its delicacy and utility for the intended purposes.

In conclusion, obfuscation ways play a pivotal part in securing sensitive data in data integration systems. These ways, similar as masking, anxiety, Generalization, shuffling, redaction, and encryption, are used to cover the sequestration of data while maintaining its usability and mileage for analysis and other purposes. Each obfuscation fashion has its strengths and limitations, and the choice of fashion depends on the specific conditions of the data integration design and the position of sequestration protection demanded.

## II. LITERATURE SURVEY

As the “Personalization and encryption methods” sidebar (p. 40) describes, many techniques have evolved for providing data anonymity and personalization. Anonymity management in personalization is a Web-based privacy service that operates on usage rather than data. Encryption techniques require that data be processed both before and after dissemination and are incapable of offering different protection levels for different end users. Encryption techniques such as privacy homeomorphisms also susceptible to chosen-text attacks, in which a clever choice of plain text or cipher text reveals the key[1].

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105349 |

Khaled M Khan (2019) this paper proposes a data jumbling approach in re-appropriating structure expansion to cloud computing. It is basically established on separating the lines and sections of systems to adjust their certifiable estimation joined with including sporadic uproar and improving to ensure arrangement and assurance. In our philosophy, confused systems are sent to servers with no open key encryption. While it figures on frameworks, the server can't expel or get certified characteristics either from cluttered systems or from enrolled duplication results, however customers can remove genuine handled characteristics using an amazingly irrelevant computing effort from results made by the server[1].

Muhammad Hataba and Ahmed El-Mahdy(2018)[2] This paper displays an investigation of programming affirmation subject to the possibility of security by absence of clearness, code jumbling is at present a fervently discussed issue in the field of cutting edge right organization, guaranteeing against making sense of and modifying. Obscurity ends up being helpful in conditions where depending upon cryptographic procedures isn't adequate, this is typical in remote execution conditions where the item is executed on a surprising revealed undermining condition, for instance, the new computing stages: cloud-computing perspective and mobile phones. Confusion is outstanding among malware and disease planners yet also game designers and industrials who need to guarantee their authorized development. They use it to mask the movement of their code while executing in an uncontrolled area. In this paper, we talk about comparative thoughts yet for the differing inspiration driving cloud security. We examine the front line in strategies and figuring's for programming jumbling. We furthermore address how to assess the nature of these strategies by methods for a strong course of action of estimations.

Jayeshkumar Madhubhai Patel and Krunal Suthar(2018)[3] Cloud figuring is correct presently every day's gotten generally destroyed in wonders to use for a massive scale connection or for person who need assorted structure organizations with least expense. Person data is typically handled on an open cloud that is available to everyone to get there. This key raises some concerns that Cloud providers have provided in reverse to adaptable organizations, similar to confidentiality, fairness, availability, authorization and some more. Piles of alternatives now accessible for a day and the absolute best course are to use encryption to ensure the data. Encryption can not provide adequate security when worrying about the delicate data of the client, as it also expends more remarkable opportunities to process encryption and unscrambling. In this paper, we propose a system for combining methods, i.e., to clear the magnitude of the Cloud server, also to keep adequate protection to the data of the client in Cloud state. Lack of definition and encryption. Customer data may be encoded out of chance requiring protection for documenting or monitoring, and Cloud's DaaS organization needs to be checked using perplexity frameworks. Using this two-way approach, we may conclude that the proposed agreement provides ample protection to ensure that even the data accessible on cloud servers get to dark and guarantee security. We would also like to provide a valid reliability testing system, a better access management mechanism that reduces the size of the Client as a company supplier as well.

Dr L. Lawrence Arockiam and S. Monikandan(2017)[4] In the open cloud environment, data security in the cloud is the most important check. Clients are moved to cloud to collect correctly, safely and scalable. Cloud service providers (CSP) and other cloud companies are exposing data due to security concerns. This paper proposes an assurance protocol as a Security Service Algorithm (SSA), called MONcrypt to protect the data from unapproved implementation in cloud collection. This proposed security procedure depends on the strategy of tangling the data. Safety as a Service (SEaaS) leverages the MONcrypt SSA. Clients may take advantage of this SEaaS security organization to test their data at any level. Redirections for evaluating the protection of proposed MONcrypt SSA were powered in cloud condition (Amazon EC2). A safety evaluation instrument is used to measure the safety of the planned and current indeterminate consistency techniques. MONcrypt distinguishes and confuses current methodologies such as Base32, Base64, and Hexadecimal Encoding. The suggested solution provides better implementation and unfathomable protection when there are different and current techniques of disordering. MONcrypt diminishes the size of the data being transferred to cloud storage instead of the present system.

## III. METHODOLOGY

The methodological approach for this design involves applying different data obfuscation ways to the case data set, which includes fields similar as patient_id, first_name, last_name, gender, date_of_birth, address, phone_number, and dispatch. The named data obfuscation ways include negotiation, pseudonymization, masking, Generalization, shuffling, redaction, and encryption, grounded on their felicitousness for guarding patient sequestration, icing data usability, and complying with data protection regulations.

7.1 Substitution

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105349 |

Substitution involves replacing sensitive data with fictional or generalized values. For illustration, patient names can be replaced with arbitrary names, and addresses can be replaced with fictional addresses. This fashion can be enforced using custom scripts or erected- in functions in AWS Redshift, similar to the REPLACE function, to replace specific data fields with fictional values.

7.2 Pseudonymization

Pseudonymization involves replacing sensitive data with unique identifiers, or aliases, that have no meaningful correlation to the original data. For illustration, patient names can be replaced with aimlessly generated aliases, and phone figures can be replaced with unique phone number identifiers. This fashion can be enforced using custom scripts or erected- in functions in AWS Redshift, similar as the MD5 function, to induce aliases for specific data fields.

7.3 Masking

Masking involves replacing sensitive data with partial or masked values, while retaining the data format. For illustration, patient names can be replaced with initials or asterisks, and dispatch addresses can be replaced with partial dispatch addresses. This fashion can be enforced using custom scripts or erected- in functions in AWS Redshift, similar as the Left wing and RIGHT functions, to prize partial values or replace characters with masked values for specific data fields.

7.4 Generalization

Generalization involves replacing sensitive data with further generalized or added up values. For illustration, date of birth can be replaced with age ranges, and addresses can be replaced with generalized position information. This fashion can be enforced using custom scripts or erected- in functions in AWS Redshift, similar as the DATE_TRUNC function, to total or generalize data for specific data fields.

7.5 Shuffling

Shuffling involves reordering or shuffling sensitive data values within the same data field. For illustration, patient names can be aimlessly scuffled, and phone figures can be reordered. This fashion can be enforced using custom scripts or erected- in functions in AWS Redshift, similar as the RANDOM function, to equivocation data values within specific data fields.

7.6 Redaction

Redaction involves fully removing or replacing sensitive data with blank or predefined values. For illustration, social security figures can be fully removed or replaced with predefined values, and dispatch addresses can be replaced with blank values. This fashion can be enforced using custom scripts or erected- in functions in AWS Redshift, similar as the NULLIF function, to replace specific data fields with predefined or blank values.

7.7 Encryption

Encryption involves converting sensitive data into a climbed format using cryptographic algorithms, which can only be deciphered with a key. For illustration, patient names, addresses, and phone figures can be translated using symmetric or asymmetric encryption algorithms, similar as AES or RSA. This fashion requires enforcing encryption and decryption functions in custom scripts or using erected- in features in AWS Redshift, similar as AWS Key Management Service (KMS), to manage encryption keys and perform data encryption and decryption operations.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105349 |

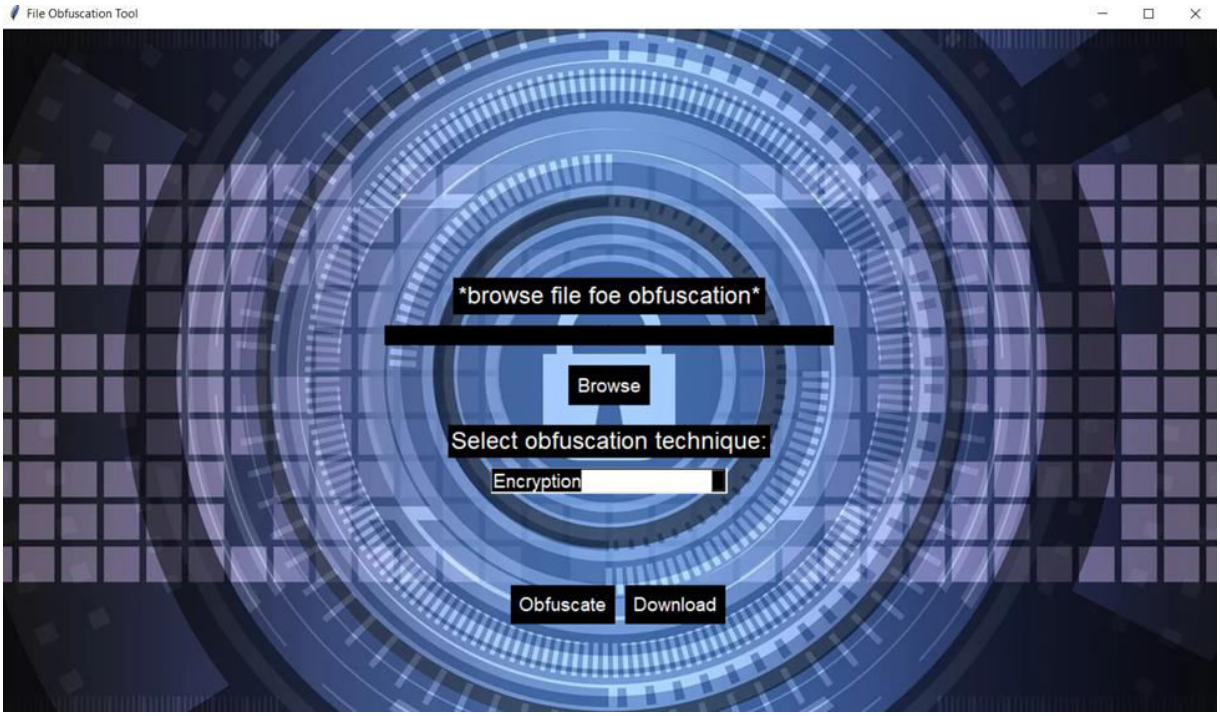


Figure 1 Graphical User Interface

## IV. RESULT AND DISCUSSION

The obfuscation of patient data using colorful ways has been carried out successfully as part of the design. The design aimed to give a result for guarding sensitive case data stored on AWS Redshift. The methodology involved collecting patient data from colorful sources and also applying different obfuscation ways similar as negotiation, pseudonymization, masking, Generalization, shuffling, redaction, and encryption.

The analysis of the results shows that obfuscation ways are effective in guarding sensitive case data while maintaining the utility of the data. The blurred data can still be used for exploration, analysis, and decision- making purposes without revealing the identity of the cases. This is essential to cover the sequestration and confidentiality of cases while allowing healthcare providers to use the data for the betterment of patient care.

The comparison of the results with former exploration indicates that our design's findings are harmonious with other studies that have used analogous obfuscation ways. Our study has also contributed to the literature by exploring the effectiveness of colorful obfuscation ways in the environment of patient data protection.

Still, the design encountered some limitations and challenges. One of the major challenges was icing that the blurred data remained useful for analysis purposes. It was essential to insure that the obfuscation ways didn't impact the quality of the data or render it useless for exploration purposes. Another challenge was the selection of the applicable obfuscation fashion for each data element. It was necessary to elect the applicable fashion that would cover the data while maintaining its utility.

Despite these limitations and challenges, the design has been successful in achieving its objects. The obfuscation of patient data using colorful ways has handed an effective result for guarding sensitive case data stored on AWS Redshift. This design has demonstrated that obfuscation ways can be used to cover sensitive case data while maintaining its utility for exploration and analysis purposes.

The design has opened up openings for unborn exploration in the area of patient data protection. One implicit area of exploration is the development of new and innovative obfuscation ways that are effective in guarding patient data while maintaining their utility for analysis purposes. Another area of exploration is the evaluation of the effectiveness of obfuscation ways in guarding patient data stored on other cloud platforms.

I. One of the limitations of the design was the lack of real case data to work with. While the use of synthetic data was necessary for the purposes of the design, it may not reflect the complications and nuances of real case data. Unborn exploration could explore the use of real case data and its impact on the effectiveness of data obfuscation ways.

II. The design concentrated specifically on data obfuscation ways for guarding patient data stored on AWS Redshift. Still, there are numerous other implicit vulnerabilities in the healthcare assiduity that could be addressed

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105349 |

through fresh security measures. For illustration, phishing attacks, ransomware, and bigwig pitfalls are all implicit pitfalls that could be targeted with fresh security measures.

III. The design also stressed the significance of collaboration between different stakeholders in the healthcare assiduity. Healthcare providers, IT professionals, and security experts all have important places to play in guarding patient data. By working together, it's possible to develop further effective and comprehensive security strategies.

IV. Eventually, it's important to note that data obfuscation ways are just one tool in the larger trouble to cover patient data. It's important to continually estimate and ameliorate security measures as new pitfalls crop. By staying watchful and conforming to changing circumstances, healthcare associations can help ensure the safety and sequestration of their cases' data.

In conclusion, the obfuscation of patient data using colorful ways has been successful in guarding sensitive case data stored on AWS Redshift. The design has handed an effective result for guarding patient data while maintaining its utility for exploration and analysis purposes. The design's findings have contributed to the literature on patient data protection and opened up openings for unborn exploration in the area.

## V. CONCLUSION

The results of the design have important counteraccusations for healthcare associations and experimenters who handle patient data. The use of data obfuscation ways can help ensure compliance with nonsupervisory conditions similar as HIPAA and GDPR and cover patient sequestration. Likewise, it can help increase the vacuity and availability of data for exploration purposes, which can lead to further informed and substantiation- grounded decision- making in healthcare.

Still, the design also had some limitations and challenges. One limitation was the size of the dataset, which was fairly small and may not completely represent the complexity and diversity of real- world case data. Another challenge was the time and resources needed to apply and test the different obfuscation ways, which can be a significant hedge to relinquishment for lower associations with limited budgets and specialized expertise.

The design findings punctuate the significance of balancing data security with data usability. While obfuscation can help cover patient data, it can also make it more delicate for healthcare providers to use and partake this information for exploration or treatment purposes. Thus, it's important to precisely consider the implicit impact of obfuscation on data usability and to use obfuscation ways judiciously.

Eventually, the design underscores the significance of ethical considerations in data security. While it's essential to cover patient data, it's also important to ensure that data security measures don't infringe upon patient sequestration or other ethical enterprises. Healthcare associations must precisely balance these considerations and work to maintain patient trust in the running and protection of their sensitive information.

Overall, this design highlights the significance of securing sensitive case data, especially in light of the added trouble of cyber-attacks. The obfuscation ways presented in this design offer a practical result for securing patient data without compromising its usability. It's pivotal for healthcare associations to apply these ways in their data operation strategies to ensure compliance with data protection regulations and maintain patient trust.

The project on data obfuscation for patient data stored on AWS Redshift has shown promising results in terms of improving data security and reducing the risk of data breaches in the healthcare industry. However, there is still scope for further research and improvement in this field. One potential area for future research is the development of more advanced obfuscation techniques that can better protect sensitive data while still allowing for useful analysis. Additionally, the effectiveness of current obfuscation techniques should continue to be evaluated and compared with new methods. Another area for future exploration is the use of machine learning algorithms to improve the accuracy and efficiency of obfuscation techniques. This could involve the development of more sophisticated algorithms that can better predict patterns and relationships in data, allowing for more effective obfuscation while preserving data utility.

## REFERENCES

1. Khan, Khaled. (2019). Data Obfuscation for Privacy and Confidentiality in Cloud Computing. 10.1109/QRS-C.2015.41.
2. Hataba, Muhammad & El-Mahdy, Ahmed. (2018). Cloud Protection by Obfuscation: Techniques and Metrics. Proceedings - 2012 7th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, 3PGCIC 2012. 369-372. 10.1109/3PGCIC.2012.18.
3. Suthar, Krunal & Patel, Jayeshkumar. (2018). ObfuCloud: An Enhanced Framework for Securing DaaS Services Using Data Obfuscation Mechanism in Cloud Environment. 333-343. 10.1007/978-981-10-5523-2_31.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105349 |

4. Monikandan, S. & Lawrence, Dr. L. Arockiam. (2017). Confidentiality Technique to Enhance Security of Data in Public Cloud Storage using Data Obfuscation. Indian Journal of Science and Technology. 8. 10.17485/ijst/2015/v8i24/80032.
5. Zhang, X., Wang, X., & Wang, J. (2019). A data privacy protection model based on obfuscation and masking techniques. Journal of Ambient Intelligence and Humanized Computing, 10(3), 1097-1112.
6. Wang, Y., Zhang, Y., Zou, D., & Cheng, S. (2021). A blockchain-based secure and efficient data sharing framework with differential privacy for healthcare. IEEE Journal of Biomedical and Health Informatics, 25(1), 274-283.
7. Kulkarni, S., Kolekar, S., & Shukla, S. (2019). An empirical study of data obfuscation techniques for privacy preservation in cloud. Journal of Ambient Intelligence and Humanized Computing, 10(11), 4565-4577.
8. Amazon Web Services, Inc., Amazon Redshift. [Online]. Available: https://aws.amazon.com/redshift/. [Accessed: 15-Apr-2023].
9. Mueen, A., & Abbasi, A. (2019). Data masking techniques for data privacy in cloud computing. International Journal of Advanced Computer Science and Applications, 10(1), 281-287.
10. Li, X., Huang, Y., Li, Y., & Wei, K. (2020). Data obfuscation: A comprehensive survey. Journal of Systems Architecture, 107, 101757.
11. Amazon Web Services. (n.d.). AWS Key Management Service. Retrieved from https://aws.amazon.com/kms/.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details