World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

PCI-DSS–Aligned Analytics Pipelines: Tokenization, Vaulting, and PII Minimization for Payment Data

Ravi Kumar Vallemoni *

Senior Data Architect, FinTech Domain, USA.

World Journal of Advanced Research and Reviews, 2022, 16(01), 1258–1269

Publication history: Received on 02 September 2022; revised on 24 October 2022; accepted on 28 October 2022

Article DOI: https://doi.org/10.30574/wjarr.2022.16.1.1015

Abstract

The swift increase in the number of digital payments has increased the necessity of effective security services that might secure delicate payment card data. The compliance with the Payment Card Industry Data Safety Standard (PCI-DSS) provides a systematic methodology to safeguard the data of the cardholders (CHD) and reduce the risk of the data breach. The modern analytics pipelines usually handle large quantities of payment data in order to extract actionable insights; nonetheless, it brings about enormous privacy and compliance pressures. The current paper details an analytics framework, which is aligned to PCI-DSS that focuses on tokenization, vaulting, and minimization of Personally Identifiable Information (PII). The analysis remains useful, whereas tokenization substitutes sensitive card data with its non-sensitive counterparts, which substantially shrink the analysis surface. The creation of sensitive tokens or limited amounts of PII in a controlled environment with minimally sensitive data or information, binding secured to vaulting allows only access within a rigorously regulated setting. The minimization techniques of PII minimize exposure because only the necessary data is collected, and the sensitive data are made anonymous as much as possible. The proposed study introduces an end-to-end analytics pipeline, which combines these approaches and emphasizes operational efficiency, as well as compliance with regulations. This paper determines the performance and compliance advantages of this framework by doing a comparative study with the standard data processing pipelines. The results of the experiments indicate that the compliance risk is reduced significantly and the data security is improved without any weakening of analytical tools. The framework also facilitates auditability, real-time observations and it can be combined with the existing enterprise analytics platforms. With the compliance to the PCI-DSS standards and the use of advanced security measures on data, the offered pipeline has become a scalable, secure, and privacy-conscious service to support the recent payment analytics. The results point to the balance that is very critical between data utility and regulation compliance, and give recommendations to organizations that want to implement effective and secure payment analytics.

Keywords: PCI-DSS; Tokenization; Vaulting; PII Minimization; Payment Data; Analytics Pipelines; Data Security; Compliance

1. Introduction

The recent growth of a digital economy is the reason why the number of electronic payment systems, including online payment systems, mobile payment systems, contactless cards, and digital wallets, has increased dramatically. [1-3] this expansive growth of payment options has changed how people and companies transact financial transactions, as well as providing convenience and efficiency, but has also caused new security concerns. As more payments are done online, confidential financial data especially payment cards have become a favorite target of hackers, especially those who are after them. The consequences related to data breach of cardholder information may have extensive effects (such as significant financial losses by both consumers and organizations) and cannot exist without the loss of trust and reputation as well as infringement penalties. To address these risks, the Payment Card Industry Data Security Standard

* Corresponding author: Ravi Kumar Vallemoni.

(PCI-DSS) was instituted to form a global framework to protect cardholder data at any point during its existence over time, including gathering and processing to storage and delivery. Some of the commonly enforced security mechanisms specified by PCI-DSS are encryption to secure data in storage and transmission, stringent access policies that only authorized members have access to sensitive data, and ongoing monitoring and logging to identify and deal with suspicious activities. Adherence to PCI-DSS does not only assist the organizations to minimize chances of information breaches but also shows a desire to secure payment environments, making customers become more confident and strengthening their compliance with regulations. The dynamism of payment technologies, as well as the necessity to keep sensitive information confidential, reveals the role of the requirement to design the safe, efficient, and compliant payment processing and analytics systems in the current digital economy.

1.1. Needs of PCI-DSS–Aligned Analytics Pipelines

With both organizations and their payment processes becoming increasingly dependent on data-driven insights to optimize their payment operations, there has been a high demand on analytics pipelines that are scalable with the requirements of PCI-DSS. These kinds of pipes make sure that payment sensitive information is reviewed in a secure way, their requirements are fulfilled, and business intelligence procedures are operational. The next sub-sections define the most important requirements that lead to the implementation of PCI-DSS-congruent analytics pipelines.

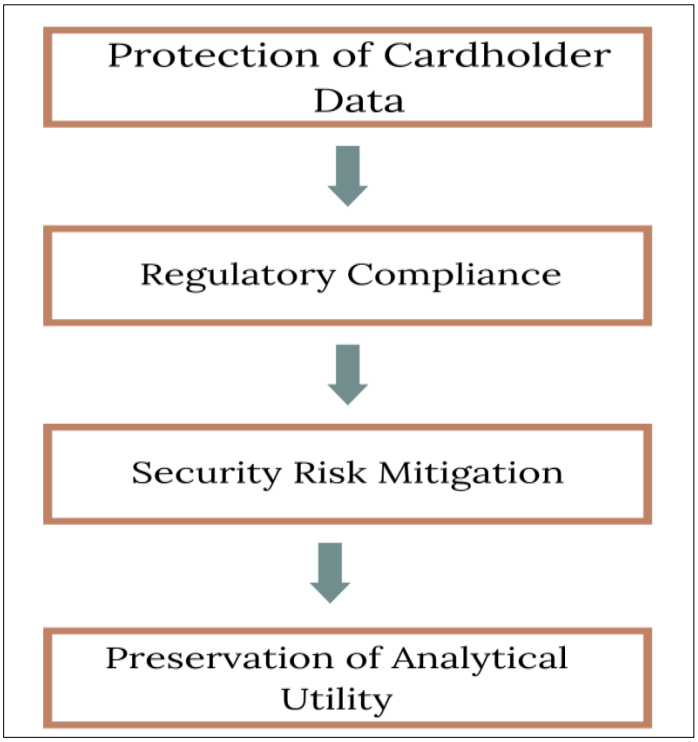

Figure 1 Needs of PCI-DSS–Aligned Analytics Pipelines

1.1.1. Protection of Cardholder Data

The strong security of cardholder data (CHD) throughout the whole analytics processing is one of the primary needs. The conventional CHD analytics pipeline usually takes place atop the raw data, putting organizations at a high risk of unauthorized access or breaches. An agreed pipeline fully inclusive of PCI-DSS is capable of mitigating such risks through the methods of tokenizing, encryption, and vaulting, which means that sensitive information is never revealed without significant business need in an operational system or analytic system. Not only helps to protect CHD assists in mitigating possible financial and reputational harm to organizations in the occurrence of a data breach.

1.1.2. Regulatory Compliance

Organizations involved in handling payment card data are legally and contractually required to comply with regulatory services like the PCI-DSS. The requirements of analytics pipelines regarding data storage, access management, and monitoring should be high to prevent fines and ensure certification. The aligned pipeline based on the PCI-DSS minimalizes compliance costs by cutting down the limits of the systems at risk of processing sensitive information, an

easier audit, and proactive security offerings. This will make sure that organizations could utilize analytics without feeling the threat of non-compliance and regulatory breaches.

1.1.3. Security Risk Mitigation

The increase in the threat terrain, programs such as cyberattacks and insider threats, have led to the need to have pipelines that reduce exposure to sensitive data. Secure analytics that include data tokenization and vaulting helps decrease the level of attack surface and improve the general level of data security. These will not allow attackers to see significant information despite breaking into the analytic systems in question, which has the potential to mitigate potential breaches, and transactions of the businesses will be secured.

1.1.4. Preservation of Analytical Utility

However, as the data security is kept, it is also interesting that analytics pipelines have a high operational utility as well. Pipelines aligned to PCI-DSS should be optimized on security and analytical capacity that provides organizations an opportunity to analyze transactions, detect fraud, and obtain a business intelligence without having access to the raw CHD. Technologies like tokenization, PII minimization and controlled de-tokenization can help generate meaningful insights and still ensure privacy and compliance.

1.2. Tokenization, Vaulting, and PII Minimization for Payment Data

Security Analytics of secured payments, tokenization, vaulting, and minimization of personally identifiable information (PII) are three complementary measures that help to safeguard sensitive data without alleviating the usability and usefulness of analytics. [4,5] The process of tokenization using non-sensitive tokens instead of sensitive cardholder data (CHD) consists of substituting cardholder data with non-sensitive tokens that cannot be used effectively in any context outside of the secure setting. This is done to make sure that in the event that tokens are stolen or tampered by unqualified personnel, the card information is still intact. The generation of tokens is done deterministically or randomly based on the security requirements and the identity of tokens to CHD is stored only in a secure vault. With tokenization, the analytics operations are no longer tied to the raw CHD, and the amount of sensitive data revealed to the world is minimized, as well as the systems that can be subjected to the PCI-DSS compliance. Vaulting is the safe data store of containing sensitive information and mapping of tokens to CHD. Vaults are based on high levels of encryption and use of multi-factor authentication, role-based access controls, and auditing logs, which make sure that authorized personnel only access or detokenize data. Vaulting eliminates insider threats, external intrusion, and unauthorized usage by isolating raw CHD of operational and analysis systems. The vault serves as a restricted space where sensitive information could be safely stored and retrieved according to the regulatory and organizational policies. PII minimization is used to supplement these methods by reducing the amount of personally identifiable data that is received and used or stored to a minimum required to be necessary to operational and analytical purposes. Techniques like data masking, anonymization, and rigid data retention controls minimize regulatory and security risks and hence meet the privacy requirements without distorting the value of data in analysis. Together, these three strategies combined, tokenization, vaulting, and PII minimization allow companies to safely handle payment information, conduct valuable analytics, and be regulation-compliant, thus creating a robust, privacy-acquisitive system of payment analytics pipelines.

2. Literature survey

2.1. PCI-DSS Compliance Framework

Payment Card Industry Data Security Standard (PCI-DSS) is an internationally accepted standard that is set to safeguard cardholder data (CHD) and secure payment handling. [6-9] The PCI-DSS creates an elaborate list of the security controls that customer service organizations that accept payment cards are required to enact to avert malicious activities of unauthorized access, fraud, and information breaches. Some of the key controls include data encryption, monitoring, access control and logging controls. The information is coded and it becomes unreadable to unauthorized individuals in a case of interception or compromise. The access controls allow the handling of sensitive information to the personnel who are authorized to handle the sensitive data based on their utilization by the organization in operation thus reducing the threat of insider attacks or unintentional leakage. Monitoring and logging are also critical since keeping comprehensive audit trails would allow organizations to identify any same activities, meet the requirements of security regulations, and conduct forensic investigations in the events of security breaches within organizations. The submission to PCI-DSS enhances the security posture of an organization besides creating confidence with the customers, regulatory and business partners by the way it indicates compliance with the industry best practices in ensuring that sensitive payment information is handled professionally.

2.2. Tokenization

One such method of data security is tokenization, the process of representing sensitive and cardholder data with non-sensitive unique tokens. The given approach has been well known in the literature due to its capability of minimizing the risk of data breach to a considerable extent, as well as its scope in terms of narrowing the scope of compliance with PCI-DSS. In contrast to encryption where CHD is changed to a format that cannot be read, and the original data remains unaltered; tokenization replaces the sensitive data permanently with a reference token which does not have any exploitable content beyond the secure setting. It has been shown in literature that tokenization can be especially effective in the situation where analytics or operational systems must handle payment data without exposing CHD. Through tokenization, organizations are able to analyze the details of transactions, business intelligence and fraudulent activities in tokenized data sets without handling real card information and as a result, reduce compliance load and make the impact of breaches less significant. Comparative literature regularly points out that tokenization has adequate performance features and data obfuscation properties, because processing tokens, in most cases, needs less computation resources than an encryption-based implementation.

2.3. Vaulting

Complementary technique to tokenization is called vaulting, in which the sensitive tokens or minimal personally identifiable information (PII) are stored in a controlled, isolated environment. Strong security measures are usually integrated in the vault like multi factor authentication, role-based access control, and defensive auditing logs to avoid unauthorized access. Literature highlights that in addition to protecting original sensitive data, vaulting is also used as a repository of token-to-CHD mapping that is necessary to de- tokenize when business operations with legitimate purposes need the access to the original card data. Vaulting can secure meaningful CHD by removing sensitive data of analytics and operational systems bypass. Researchers have noted that vaulting used in combination with tokenization forms a stacked defense model, which minimizes the attack surface besides easing the compliance with PCI-DSS by confining systems that explicitly process sensitive data. Besides, vaulting can be used to help enforce regulatory and contractual requirements through providing secure storage and controlled retrieval of vital payment and identity details.

2.4. PII Minimization

Minimization of person identifiable information (PII) is a security and privacy approach, aimed at obtaining only as much data as people need to conduct a particular business or analysis operation and anonymizing or masking sensitive fields. In the literature, it has been continuously emphasized that the minimum use of PII can not only ensure exposure to the potential breaches is limited, but it can also decrease regulatory and compliance costs. Such techniques as pseudonymization, masking, and aggregation are extensive to keep the analytical utility without having information which can be identified. To illustrate a case in point, in transaction analytics, it is feasible to substitute cardholder names with unique IDs or to generalize the demographic data and thereby obtain useful information without infringement on them. Research indicates that the PII reduction directly influences the risk reduction because the smaller the data used in operating systems, the smaller are the attack surface and further commitments to the regulations on each data protection policy, including GDPR and PCI-DSS. It is also in accordance with the principle of data economy, which requires organizations to reevaluate their data-gathering methods and only store the information that can be needed in the course of operations, analysis, and regulation.

2.5. Existing Analytics Pipelines

Conventional payment analytics pipelines do not usually include tokenization or vaulting services which makes them more exposed to the risk of breaches of the data security scope and the risks of data security breaches. According to literature reviews, these pipelines are often filled with CHD and PII, which are directly processed by operating and analysis systems and put the sensitive information at risk of insider attacks, cyberattacks, and unintentional information disclosure. Comparison between these traditional systems and the tokenized and vaulted systems illustrates a number of security and operational deficiencies. Un-tokenized pipelines will be prone to extensive encryption in numerous environments, raising the complexity of the system, its computation costs, and compliance issues. On the other hand adding tokenization and vaulting to analytics pipelines lets organizations have high-grade analytic capability but significantly lowers the total number of systems to both store and process CHD as well. Research is also finding this integration has the effect of instilling increased security levels, easing the auditing process, and reducing the cost of PCI-DSS compliance. Moreover, when using tokenized pipelines, the safe use of cloud-based analytics and machine learning applications is possible, which otherwise are limited when working with raw cardholder information. In general, the

literature presents the significance of implementing current data protection measures in the payment analytics field to balance operational approach, compliance adherence, and high level of security.

3. Methodology

3.1. System Architecture

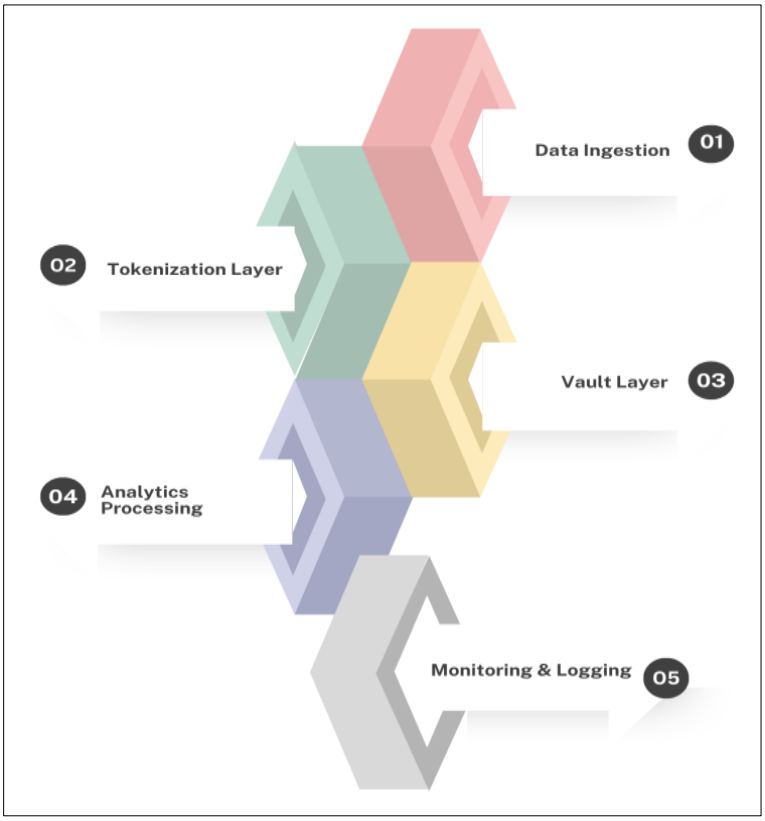

Figure 2 System Architecture

3.1.1. Data Ingestion

The data ingestion layer constitutes the first component of the analytics pipeline, and is where payment and [10-12] transaction data is received through a variety of sources (point-of-sale systems, e-commerce systems, and payment gateways). This layer makes sure that data is put up in a structured and regular format with primary validation done to mark aberration or missing records. Proper ingestion is also a necessity, as it provides the basis of safe and correct downstream processing and the maximum risk is the corruption or inconsistency of data passing the analytics pipeline.

3.1.2. Tokenization Layer

At the tokenization layer, sensitive cardholder data (CHD) is substituted with unique non-sensitive tokens that cannot be utilized in a non-secure environment. The process enables downstream analytics to work using data without revealing the raw CHD, therefore, mitigating the chances of data breach significantly. The use of tokenization also reduces the number of requirements of PCI-DSS since sensitive data is limited to the controlled portions of the pipeline and the safe execution of analytics is possible without the use of actual card numbers.

3.1.3. Vault Layer

The mapping of tokens to the original sensitive data as well as limited personally identifiable information (PII) required by the operations are stored in the vault layer in a manner that is secure. Multi-factor authentication, role-based access and encryption are all used to ensure access to the vault is strictly controlled where only authorized personnel access or de-tokenizes sensitive information. The vault layer may create a strong protection against breaches of sensitive data by isolating analytical data and may help address regulatory demands.

3.1.4. Analytics Processing

After being tokenized/anonymized, the sensitive data is sent into the analytics processing layer. In this case, business intelligence, fraud detection and transactional analytics are executed without raw CHD or any additional PII being displayed. The non-identifying and encrypted datasets enable the analyst and automated systems to take insights, as well as preserving privacy, showing a trade-off between analytics and data security. The processing layer may comprise aggregation, statistical analysis and machine learning and all these are to work on non-sensitive representations of payment data.

3.1.5. Monitoring and Logging

Monitoring and logging layer constantly monitors the access to the data, events of tokens, and changes made to the datasets. Comprehensive audit paths are kept to monitor peculiarities of activities, assist with investigations, and meet the demands set by PCI-DSS and other regulation standards. Also, accountability and transparency can be monitored, and administrators can check whether sensitive data are being processed, as well as monitor the fact that the analytics pipeline is running safely and reliably.

3.2. Tokenization Process

A basic security measure in the current payment analytics pipelines is the tokenization, created to substitute the sensitive cardholder data (CHD) with an unidentifying and non-sensitive identifier called a token. [13-15] This can be mathematically represented as;

3.2.1. Token=f (Card Number,Tokenization Key)

As the input to this operation the card number of the original card is taken, and the tokenization key is one of the cryptography parameters that maintain uniqueness and non-predictability of the token created. The concretization obtained is a token that is in itself meaningless and cannot be reverse-engineered to obtain the original card number without knowing the secure token vault in which the mapping is stored. Such a one-way mapping has the benefit that even intercepted or exported tokens in a less secure environment do not give any usable information on the original CHD. The generation of unique tokens is a very critical facet of tokenization. All cards are given cards with unique numbers this prevents the occurrence of a correlation among my transactions made using the same card. This specialness is critical to data integrity during analytics processes where data about transactions aggregated or made anonymous is processed to find patterns, detect fraud, or create business insights. Decoupling transaction identifiers and real card number will allow organizations to do meaningful analytics without exposing sensitive data to more risk. Also, tokens cannot be undone out of secure vault. In comparison to encryption in which data may be decrypted in case of key loss; tokenization benefits are that the initial data of the card is never presented in plaintext within the functional or analytical environments. De-tokenization under tough access control can only be implemented in the vault that stores the mapping between tokens and card numbers in a secure environment. The given design shows a considerable decrease in the area of PCI-DSS compliance, restricts the attack surface, and improves the situation with data security. On the whole, tokenization process has the advantage of balancing operational efficiency, analytical capability and effective security overall, and is a backbone of securing payment data and of the contemporary analytics pipelines.

3.3. Vaulting Mechanism

The vaulting is the essential part of the secure payment data management because it is the central place where the sensitive cardholder information (CHD) is stored, as well as the mapping between tokens and their respective initial data. In the proposed system, AES-256 encryption that is one of the strongest symmetric encryption standards today is used in its vaulting in order to make sure that all the data stored cannot be read by a third party. AES-256 is highly cryptographically secure and therefore the attacker will not be able to decrypt the sensitive data even when the vault storage is accessed. This encryption is not only being applied in the CHD only, any minimum personally identifiable information (PII) required to run the operations is also encrypted and hence all sensitive content in the vault is secured. The vault also uses multi-factor authentication (MFA) in addition to strong encryption to protect access. MFA also mandates the use of two or more authentication methods by the user e.g. password or hardware device, or a biometric determine before access can be granted. This type of security is layered to a large extent, which minimizes the chances of accessing applications without authorization despite the cases of compromised credentials. Role-based controls further narrow access policies such that only authorized staff with clearly stated operation needs can detokenize or look into the original CHD. All the vault operations, retrieving data, tokenization mapping and changing the system are recorded into an extensive audit trail to monitor compliance and security. AES-256 encryption combined with multi-factor authentication as well as stringent access control policies makes the vault a highly-secured environment, separating sensitive data with analytics pipelines and other operational systems. Centralizing sensitive data and

providing strict access furthermore, in addition to external breaches prevention, vaulting averts insider threats. Moreover, the compliance with the PCI-DSS and other regulatory standards is also facilitated by this mechanism as it restricts the number of systems that have direct access to CHD, thus, decreasing the operational risk and compliance costs. In general, the vaulting creates a safe and regulated bankruptcy in which delicate payment information can be safely kept and never claimed except in the circumstances of a valid justification.

3.4. PII Minimization Techniques

Figure 3 PII Minimization Techniques

3.4.1. Data Masking

Data masking is the practice of covering sensitive values of personally identifiable information (PII) in order to keep the original values hidden to unauthorized users or systems. To consider the example of payment analytics, only four final digits of the card number can be shown, and the remaining ones can be substituted with asterisks or random symbols. This method will enable the business operations and analytics to continue processing with partially masked data without negatively affecting operation performance and uncovering complete sensitive data. In environments where the original data is needed, masking is reversible; however, it is necessary to keep the original data hidden so that the systems and operators (analysts) are never exposed to the full PII.

3.4.2. Anonymization

Anonymization has been defined as the act of eliminating or altering identifiers irreversibly to ensure that, a person cannot be re-identified using the data. In contrast to masking, anonymization cannot be undone and this implies that the original PII cannot be reconstructed once the procedure has been implemented. Most methods of data analysis, including the pseudonymization, generalization, or aggregation method are normally employed to anonymize datasets used in data analysis. It is also noted in literature that anonymization can be particularly useful in decreasing regulatory exposure because anonymized data is sometimes not subject to strict privacy laws such as GDPR or PCI-DSS and still can be useful in statistical or trend analysis.

3.4.3. Data Retention Policies

The introduction of stringent data retention measures is necessary in order to reduce the risk of sensitive PII storage. Retention time should be organized in a way specified in business, regulatory and compliance obligation such that, sensitive data should be retained as long as it is required to support operations and analytics in their operational activities. Once the retention period is over, all the systems including those in the backups and logs must be purged with PII. Reducing the long-term storage of sensitive data reduces the risk of exposure to the possible breach of information, decreases the compliance risk, and imposes a postulate of data economy, according to which organizations store and attain only information that their business processes require to unquestionably exist.

3.5. Compliance and Security Checks

Regulatory checks and security scan is the essential part of any payment analytics pipeline which is defensible and provides a guarantee to stay within the framework of a certain regulatory standards like off-the-casb bushes like PCI-DSS, at the same time being lossless in data protection. [16-18] the entire process relies on continuous monitoring, which allows continuous visibility of all operations of the pipeline such as the ingestion of data, its tokenization, access to the vault, and analytics actions. With constant monitoring of events in the system, administrators will be able to identify any abnormal system behavior, attempts of access by unauthorized users, or any unusual trend of data which can possibly reflect a security invasion. This proactive nature gives organizations the ability to react fast to the threats and reduce the risk before they escalate as well as maintaining all the processes within the specified compliance ranges. Along with real time monitoring audit logs are also important in ensuring traceability and accountability in the pipeline. All activities containing sensitive cardholder data (CHD) or personally identifiable information (PII) such as token generation, de-tokenization requests, vault access and data transformations are carefully captured. These logs offer a full record of system events that may be reviewed internally, report to the regulatory bodies, and used in forensic analysis in case of security intrusion. Detailed audit trails equally support periodic compliance evaluations as well as demonstrating compliance with the PCI-DSS guidelines support, including access control implementation, data retention approaches and monitoring guidelines. Moreover, the compliance and security checks will be incorporated with automated alerting and reporting systems to make sure that the policies or regulation violations has been pointed out instantly to the accountable authorities. This integration improves human error, improves operational control and promotes security awareness culture. Having a resilient, safe analytics environment protecting sensitive payment data, reducing risks, and ensuring regulatory compliance can be achieved by ensuring smooth monitoring and rigorous audit logging. Finally, the provided measures would not only protect the pipeline against cyber threats but also support the confidence of its customers, partners, and regulators as an active business approach to data protection and privacy.

4. Results

4.1. Experimental Setup

The experimental will be structured to test the efficiency of proposed secure payment analytics pipeline in regards to the PCI-DSS regulations, system workability, and analytical usefulness. In experiment, a test data of 1 million simulated payment records was created. Such records resemble real situations of transactions, such as cardholder details, transaction value, merchant identifier, and so on, and other appropriate metadata. The variability in data could be controlled with the help of simulation and the sensitive cardholder data (CHD) could be safely tested without being exposed to actual payment data. The data was designed with the characteristics of the messaging that would comprise the commonity of transactions and volume of transactions that are commonly found in actual payment networks to allow significant evaluation of the process of tokenization, vaulting, and analytics processes under a setting that is similar to actual production networks. Three main indicators were used in order to measure the pipeline quantitatively. Firstly, PCI-DSS scope reduction reflects the degree of reduction in sensitive data disclosure because of digitalization and vaulting. This measure can evaluate the effectiveness of the proposed architecture in decreasing the number of systems and processes processing raw CHD to decrease compliance and risk of breaches. Second, performance latency measures latency of security controls like tokenization, vault access, and encryption. Latency measurement is necessary so that security efforts may not slow down the processing of transactions or analytics processes, especially when transactions are high. Third, Accuracy of analytics verifies whether anonymization, the tokenization of data, and the minimization of PII have an impact on the integrity and utility of the derived analytics. The importance is ensuring that the security of processing payment data will not reduce the capabilities to conduct significant trend analysis, detect fraud, or business intelligence operations. The experimental setup was put up such that it simulated a complete end-to-end pipeline, and the data ingestion process, tokenization, vaulting, anonymization and analytics processing. With a systematic assessment of these metrics with the large-scale test data, the experiment gives detailed insights into the security trade-off with regulatory compliance and operating efficiency. This system will allow a repetitive, representative, and controlled evaluation of the intended secure analytics infrastructure within a realistic payment setup.

4.2. Performance Evaluation

Table 1 Performance Evaluation

Metric	Traditional Pipeline (%)	PCI-DSS Pipeline (%)
Latency	100%	108.3%
Compliance Scope	100%	30%
Data Breach Risk	100%	20%
Analytics Accuracy	99%	98.5%

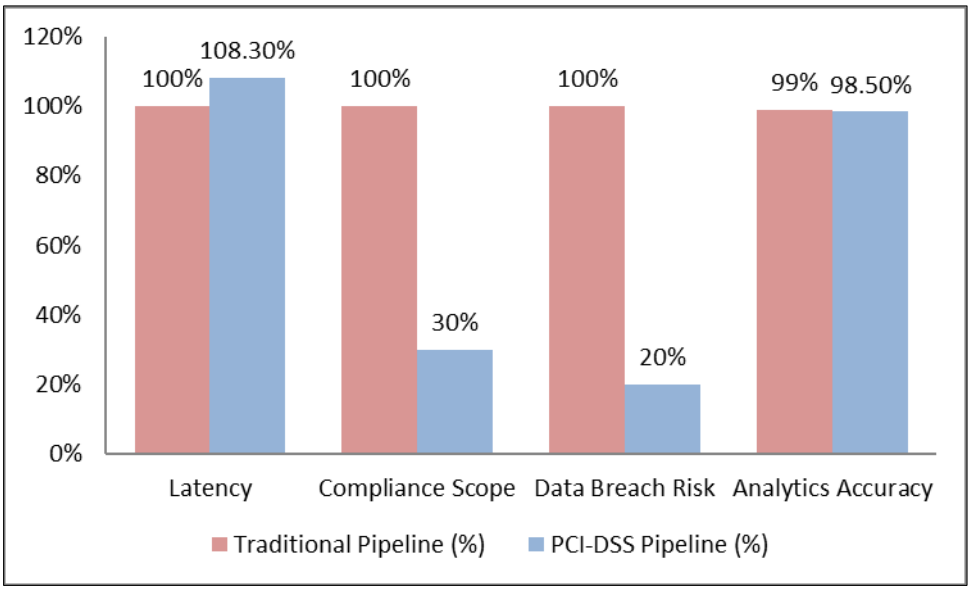

Figure 4 Graph representing Performance Evaluation

4.2.1. Latency

The latency is the time that the data spends on walking the analytics pipeline including ingestion and eventual processing. The traditional pipeline base is set to 100% and the PCI-DSS compliant pipeline recorded slightly higher base to 108.3% considering the extra security procedures, which include tokenization and vaulting. This is a moderate overhead that represents the trade-off between improved security and processing speed portraying that effects in a system-wide compliance implementation have minimal effects on system performance and does not delay functionality of analytics processes.

4.2.2. Compliance Scope

Compliance scope is defined as the number of systems and processes that deal with sensitive cardholder data and that thus are subject to provisions of PCI-DSS. The scope in the traditional pipeline is at 100% so that almost all systems that process CHD can be subjected to compliance audit. With the inclusion of tokenization and vaulting, the PCI-DSS pipeline plays down the number of scopes significantly to 30 percent, meaning that the majority of sensitive information is either isolated, and only a few systems would need their compliance addressed directly. Such reduction eases the audit process, reduces regulation, and decreases the risk of working with CHD.

4.2.3. Data Breach Risk

Risk of data breach implies probability of unauthorized access, or exposure of sensitive information. It is assumed that the traditional pipeline, When the entire CHD is open to several systems, is placed at 100% risk. Conversely, the PCI-DSS pipeline lowers the risk to 20 percent by ensuring that sensitive data is secured by tokenization and encryption as well as limited access to vaults. This proves the fact that the separation of raw data between operational and analysis

operations is a great way to reduce the risk of breach to a minimum and improve the overall system security and safety of the organization and its customers.

4.2.4. Analytics Accuracy

Measure The accuracy of analytics can be defined as the precision and accuracy of conclusions drawn based on processed information. The traditional pipeline has an accuracy of 99 and the PCI-DSS pipeline has a high accuracy level of 98.5 regardless of the tokenization and anonymization. This minor loss is an indication of the fact that data transformation does not have an extraordinarily great impact on the utility of the analysis. It has shown that quality analytics may co-exist with secure processing enabling companies to create actionable insights without breaching sensitive data.

4.3. Security Benefits

The suggested analytics pipeline has several security advantages as it entails the usage of tokenization, vaulting, and minimization of PII, which jointly reinforce the security of sensitive payment information. The concept of tokenization is in the center of attention, where instead of having raw cardholder data (CHD), they are represented by unique non-sensitive tokens that are not oriented in any meaningful way beyond the secure infrastructure. This process is so ascertained that in case a hacker were to access the analytics systems, the decoded tokens would not expose the actual card numbers and any other sensitive data. Ensuring the decoupling of transactional data and the underlying CHD, tokenization is a crucial protection against the direct disclosure as well as minimizes the possible harm in the event of breaches considerably, therefore it is a crucial line of defense of any payment data pipelines. Vaulting is a supplement to tokenization and it offers a safe place to store sensitive information and the mapping of tokens to CHD. The factor of entry into the vault is highly regulated by a multi-factor authentication system, customized roles, and detailed audit records. All of these allow only authorized administrators to access or detokenize information so that insider threats and unauthorized access can be avoided. The vault also serves a centralized locality of safeguarding so that sensitive data is not linked with operational and analytic systems. Vaulting helps in upholding compliance with PCI-DSS standards and mitigate the full attack surface by reducing the amount of control in raw CHD. Also, PII minimization is another way of enhancing security because it ensures that the amount of personally identifiable information that is collected, stored, and processed are kept to a minimum, which is required by face of necessity in conducting business and analytics. Using anonymization, masking the data, and hard retention is a set of techniques that makes regulation liability a minimized set of the tools protecting the data analysis power of datasets. Limiting the disclosure of PII will reduce the chances of breaches of compliance and the impact of the breach of the information. All these three mechanisms combined tokenization, Vaulting, and PII minimization build a multi-layered security system that covers the sensitive data, keeps them in compliance with the regulatory requirements, and minimizes the organizational risk. The combination of the practices shows that solid analytics on the payment data can be done without significant security and privacy loss, which offer strong operation efficiency and comfort to the stakeholders.

5. Discussion

The experimental results of the suggested secure analytics pipeline indicate that there is an essential trade-off between the performance, security, and compliance. Although tokenization, vaulting and a reduction of the PII overhead already results in a small performance overhead (witnessed by the slight overhead of latency in comparison to the old pipeline) the overhead is insignificant in comparison to the large security and compliance benefits. Tokenization means that cardholder data (CHD) is not disclosed to any uncontrolled environments, and these factors significantly decrease the possible effects of data breaches. This protection is further enforced by the Vaulting which offers a safe storage of CHD and token to data mappings, as well as a multi-factor authentication, access control measures, and audit logging. The results of these measures are to form a status of a layered security framework that isolates sensitive data in non-analytic and non-operational systems, which greatly reduces the range of PCI-DSS compliance and makes the process of adherence to regulations a lot easier. Besides, PII minimization measures, such as data masking, anonymization, and rigid retention policies, also reduce the risks associated with both regulation and security as only the necessary personally identifiable information is gathered and stored. This strategy minimizes the organizational risk in the privacy regulations and ensures the data still has utility in the form of analysis. Critically, in these changes, the accuracy of analytics is not low, and performance changes are insignificant when compared to other pipelines making traditional analytical results. It means that companies are able to deploy effective security systems without a crucial reduction in the quality or reliability of information gathered using transactional and payment data. The discussion further emphasizes the fact that the strategic benefits of the small increment in computational overhead are more significant: a decrease in compliance overhead and reduced risk of breach as well as and an increase in customer and stakeholder trust. Through the application of this framework, organizations will be able to express active data protection environments, remain in harmony with the regulating requirements, and be efficient in their operations. Finally, the

results assert that safe, yet valid, and efficient payment analytics is achievable, in which the trade-offs in performance are marginalized by the significant gains in security posing, regulatory compliance, and general risk mitigation in an organization.

6. Conclusion

This paper has introduced a set of concepts to work out a complete strategy of designing an unsafe, PCI-DSS-conform payment analytics pipeline and incorporating tokenization, vaulting, and PII interjection. The suggested methodology modifies cardholder data-related critical issues related to the analytics systems management, such as the risks of data breaches, the regulatory compliance pressure, and privacy concerns. The solution to the CHD exposure within operational and analytical systems involves replacing raw card data with non-sensitive tokens, which assists in eliminating the exposure of CHD information. The tokens are uniquely created and cannot be recreated when outside the system that is controlled by the vault, so even in case of a breach in the system, sensitive data remains intact. This strategy wills greatly lower the attack surface, curtail the exposure of unauthorized data, and decrease the extent of systems under the PCI-DSS compliance and the auditing and regulatory control will be simplified.

The tokenization is reinforced by the vaulting mechanism that will offer a highly secured location where sensitive data and token correspondence can be stored. The vault will secure access of data by authorized employees using AES-256 encryption, multi-factor authentication and role-based access controls to allow only the authorized personnel to access or detokenize data. Detailed audit log and monitoring facilities strengthen accountability and traceability which enables organizations to ensure tight adherence to the standards of the PCI-DSS, besides aiding forensic investigation and operational controls. Vaulting acts as a vital component of a layered data protection approach by ensuring that the analytics processes are additional layers of insider and external attacks by separating sensitive information.

Moreover, the PII minimization is also significant to minimize security and regulatory risks. The information gathered, stored and used is restricted to the information necessary, by use of data masking, anonymization, and retention policies and reducing the possibility of being abused or accidentally revealed. Regardless of these defensive mechanisms, experimental test has proved that there are no significant changes to analytical utility and the accuracy of transaction analysis and fraud detection measures are high. This shows that scalpel processing does not have to come at the cost of actionable insights or business intelligence capabilities.

The experimental findings point to the fact that the trade-off between the relatively insignificant performance overhead of this mechanism (because of tokenizing and encrypting operations) and the significant security and privacy benefits and compliance benefits with the regulatory requirements is justified. In general, the suggested framework provides a realistic guide on organizations interested in instituting secure, compliant, and effective payment analytics pipelines.

Moving forward, it is possible that the future research can investigate improvements in the form of tokenization in real-time, automatic vaulting, and easy integrations with new AI-powered analytics tools. This progress would potentially streamline performance, facilitate instant fraud detection, and ensure high security standards in more multifaceted payment systems so that secure payment analytics can be robust and scalable to adverse dynamic operating conditions.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Kaur, R. (2020). PCI DSS implementation guidelines for small and medium enterprises using COBIT based implementation approach.

[2] Palmer, M. E., Robinson, C., Patilla, J. C., and Moser, E. P. (2001). Information security policy framework: best practices for security policy in the e-commerce age. Inf. Secur. J. A Glob. Perspect., 10(2), 1-15.

[3] Seaman, J. (2020). PCI DSS: An integrated data security standard guide. Apress.

[4] Elluri, L., Nagar, A., and Joshi, K. P. (2018, December). An integrated knowledge graph to automate gdpr and pci dss compliance. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 1266-1271). IEEE.

[5] Hwangbo, Y., Lee, K. J., Jeong, B., and Park, K. Y. (2021). Recommendation system with minimized transaction data. Data Science and Management, 4, 40-45.

[6] Thakur, A., and Saxena, A. (2019). 'Improved vault based tokenization to boost vault lookup performance. Int. J. Comput. Appl, 177(21), 24-32.

[7] McCallister, E., Grance, T., and Scarfone, K. A. (2010). Sp 800-122. guide to protecting the confidentiality of personally identifiable information (pii).

[8] Wieringa, J., Kannan, P. K., Ma, X., Reutterer, T., Risselada, H., and Skiera, B. (2021). Data analytics in a privacy-concerned world. Journal of Business Research, 122, 915-925.

[9] Muresan, A. (2020). Tokenization Techniques and Their Effect on Risk Reduction for Payment Data in Serverless E-Commerce Frameworks. Nuvern Applied Science Reviews, 4(1), 1-12.

[10] Rozenberg, Y. (2012). Challenges in PII data protection. Computer Fraud and Security, 2012(6), 5-9.

[11] Yulianto, S., Lim, C., and Soewito, B. (2016, May). Information security maturity model: A best practice driven approach to PCI DSS compliance. In 2016 IEEE Region 10 Symposium (TENSYMP) (pp. 65-70). IEEE.

[12] Ogigau-Neamtiu, F. (2016). Tokenization as a data security technique. Zeszyty Naukowe AON, (2 (103), 124-135.

[13] Vagadia, B. (2020). Data integrity, control and tokenization. In Digital Disruption: Implications and opportunities for Economies, Society, Policy Makers and Business Leaders (pp. 107-176). Cham: Springer International Publishing.

[14] McCallister, E. (2010). Guide to protecting the confidentiality of personally identifiable information. Diane Publishing.

[15] Liu, J., Xiao, Y., Chen, H., Ozdemir, S., Dodle, S., and Singh, V. (2010). A survey of payment card industry data security standard. IEEE Communications Surveys and Tutorials, 12(3), 287-303.

[16] Higham, L. (2016). Best Practices of Big Data Analytics Applied to PII Security (Doctoral dissertation).

[17] Rafiq, F., Awan, M. J., Yasin, A., Nobanee, H., Zain, A. M., and Bahaj, S. A. (2022). Privacy prevention of big data applications: A systematic literature review. SAGE Open, 12(2), 21582440221096445.

[18] Pasquale, L., Spoletini, P., Salehie, M., Cavallaro, L., and Nuseibeh, B. (2016). Automating trade-off analysis of security requirements. Requirements Engineering, 21(4), 481-504.

[19] Fischer, A., Janneck, J., Kussmaul, J., Krätzschmar, N., Kerschbaum, F., and Bodden, E. (2020, June). PASAPTO: Policy-aware Security and Performance Trade-off Analysis--Computation on Encrypted Data with Restricted Leakage. In 2020 IEEE 33rd Computer Security Foundations Symposium (CSF) (pp. 230-245). IEEE.

[20] Munappy, A. R., Bosch, J., and Olsson, H. H. (2021, August). On the trade-off between robustness and complexity in data pipelines. In International Conference on the Quality of Information and Communications Technology (pp. 401-415). Cham: Springer International Publishing.