DOI: https://doi.org/10.53052/9788367652124.04

Oleh DEINEKA[1], Oleh HARASYMCHUK [2]

Supervisor: Oleh HARASYMCHUK [2]

WYZWANIA I STRATEGIE PRZECHOWYWANIA DUŻYCH ILOŚCI DANYCH WE WSPÓŁCZESNYM ŚWIECIE W OPARCIU O KONTROLE SOC2 TYPU 2

Streszczenie: Bezpieczne przechowywanie dużych ilości informacji wymaga niezawodności, bezpieczeństwa i szybkiego dostępu do danych. W tym artykule oferujemy rozwiązanie zgodne z SOC 2, które obejmuje wszystkie wymagania dotyczące klasyfikacji danych, kontroli dostępu i monitorowania danych, a także zapewnia szyfrowanie poufnych danych firmy i zmniejsza ryzyko wycieku danych.

Słowa kluczowe: cyberbezpieczeństwo, big Data, klasyfikacja i przechowywanie danych, standardy, SOC2

THE CHALLENGES AND STRATEGIES OF STORING LARGE VOLUMES OF DATA IN THE MODERN WORLD BASED ON SOC2 TYPE 2 CONTROLS

Summary: Safeguarding the storage of substantial data volumes necessitates reliability, security, and efficient data accessibility. In our operations, we provide a solution that aligns with SOC 2 standards, addressing data classification, access control, and data monitoring requirements. This solution ensures the encryption of sensitive company data, thereby mitigating the risk of data leaks.

Keywords: Cybersecurity, Big Data, Classification and Storage Data, standards, SOC2

1. Introduction

In the modern world, information becomes the most valuable global resource, which requires reliable and secure storage, as well as the ability to have quick access to the stored data. The more significant and critical information an organization has, the more important it is to manage and store it properly. The volume of data is growing exponentially, so the immediate problem of long-term storage of large amounts of data is extremely relevant today, as modern society constantly faces challenges related

[1] Lviv Polytechnic National University, Postgraduate of Information Protection Department, Oleh.R.Deineka@lpnu.ua
[2] PhD, Lviv Polytechnic National University, Associated Professor of Information Protection Department, oleh.harasymchuk@gmail.com

to choosing reliable methods and means of storage, as well as protecting these repositories from unauthorized access. Therefore, methods and means of storing large volumes of data become defining components of modern information infrastructure. The rapid development of information technology and computer equipment constantly expands the possibilities of data storage. This storage should also be accompanied by resource savings for its implementation and provide the ability to effectively manage this data.

Big data refers to datasets that are rapidly generated and received from various sources. Such data can accumulate in almost any organization: customer data, products and services, website reviews, various surveys and their results. As a result, a large amount of data accumulates, which usually has great value for the organization. The standardization in this direction contributes to the improvement of efficiency, reliability, and security of large data storage systems, as well as the overall development of this infrastructure. Security standards require the presence of access control tools, as well as data classification and storage systems. The main task of these standards is to understand what exactly the institution owns at the data level and how it ensures control over their storage and access during business processes.

There are many companies that offer ready-made solutions to address these issues. However, these solutions, on the one hand, are expensive, and on the other hand, require complexity in management and support.

A large number of works by various researchers [1-3] are dedicated to studying the problems of storing large data, implementing new methods, approaches, and standards. There are different approaches and principles that have a lot in common, but sometimes differ significantly. However, not all of them meet the requirements for reliability and security of data storage. And those that take into account such requirements are not always suitable for practical application in all cases and for all types of data.

Most organizations form their security policies based on international standards, which are mainly audited by external audit companies that certify compliance with the standard.

The following list of standards and types of data storage can be distinguished: SOC (Service Organization Control) [5-6] and ISO (International Organization for Standardization) [7]. These are two different sets of standards and requirements related to data storage and other aspects of information security.

ISO 27001 is a standard for ensuring proper management of a company's digital assets, including financial information, intellectual property, employee data, and trusted third-party information. In turn, SOC 2 certification is more recognized and is usually preferred by American and Canadian companies.

Another important point is that SOC is divided into SOC 1, SOC 2, and SOC 3. The first one is exclusively related to financial control, and the third one is mainly used for marketing purposes, so SaaS providers can focus solely on SOC 2.2. Recommended solution for providing.

2. Research products from vendors

According to Gartner Report about Data Analytics (DA) Products [8] there are a lot companies. For example, Varonis Data Security Platform, Netwrix Auditor, Data

Insight, etc. They provide products which analyze, index, search, track and report on file metadata and file content, enabling organizations to act on files according to what was identified. DA provides detailed metadata and contextual information to enable better information governance and organizational efficiency for unstructured data management. DA is an emerging solution, made of disparate technologies, that assists organizations in understanding the ever-growing volume of unstructured data, including file shares, email databases, enterprise file sync and share, records management, enterprise content management, Microsoft SharePoint, and data archives.

Based on the data of the products, we have come to the following conclusion: the company's data solutions do provide a certain level of data analysis. However, many companies use multiple products where DA products are stored without connectors, leading to the need for additional development, which consumes both resources and time. All these products are very expensive and potentially do not cover all the data classification needs across the various products used by companies. On the other hand, we need teach team to use and customize DA products which lead to additional expense.

3. SOC2 Type 2 control requirements for data storage, classification, encryption and access management

SOC 2 Type 2 is an auditing standard developed by the American Institute of CPAs (AICPA) to assess and report on the controls related to security, availability, processing integrity, confidentiality, and privacy of data. It is often used for service organizations that handle customer data.

During SOC 2 certification, a company undergoes assessment based on trust service categories, which encompass several key criteria for the final evaluation:

1. *Security*: Under the security category, the emphasis is on safeguarding systems and stored information. This involves measures to protect data from unauthorized access and disclosure, encompassing a range of security controls and protocols to ensure the confidentiality and integrity of data.
2. *Availability*: Availability ensures that information and products are readily accessible and perform their intended functions when required by customers. This means the company must have mechanisms in place to prevent and recover from disruptions to services, ensuring that they are consistently available.
3. *Confidentiality*: Within the confidentiality category, the focus is on the secure protection of non-public information. Companies must implement strict access controls and encryption to prevent unauthorized access, sharing, or exposure of sensitive data.
4. *Processing Integrity*: Processing integrity relates to the accuracy and reliability of information processing systems. It mandates that data processing systems be complete, valid, accurate, timely, and authorized, and that customer data remains correct throughout the entire operational period. This includes data validation processes and audit trails to ensure data integrity.
5. *Privacy*: The privacy category pertains to how personal information is handled. Companies must adhere to predefined rules for the collection, use, storage, disclosure,

and deletion of personal data. This ensures that customer privacy rights are respected and that data handling complies with relevant privacy laws and regulations.
Certainly, the SOC 2 process involves several key points to ensure the security and compliance of an organization's systems and data. Let's delve into more detail on each of these points:

1. ***Quarterly Vulnerability Scanning***:
– Organizations performing SOC 2 compliance engage in quarterly vulnerability scanning of all their internal systems. This practice involves the use of specialized software or services to identify security vulnerabilities within the systems.
– These scans encompass critical and high-level vulnerabilities, which are given particular attention. The organization closely monitors these vulnerabilities until they are addressed or remediated. This process ensures that known security weaknesses are promptly identified and mitigated, reducing the risk of exploitation.

2. ***Log Management and Event Detection***:
– Effective log management is a crucial component of SOC 2 compliance. Companies use log management tools to collect, store, and analyze logs generated by various systems and applications.
– The purpose of this practice is to detect events or activities that could potentially impact the security of the organization's systems. By closely monitoring logs and setting up alerting mechanisms, suspicious or unauthorized activities can be identified in real-time, allowing for rapid response and investigation of potential security incidents.

3. ***External Penetration Testing and Remediation***:
– External penetration testing is typically performed at least annually as part of SOC 2 compliance. This involves hiring external security experts or firms to simulate cyberattacks on the organization's systems and networks from the perspective of an external attacker.
– The testing aims to identify vulnerabilities that could be exploited by malicious actors. Once vulnerabilities are identified, a recovery plan is developed. This plan outlines steps to address and remediate these vulnerabilities.
– Importantly, changes to address these vulnerabilities are implemented in accordance with Service Level Agreements (SLAs) to ensure timely resolution and maintain the security and integrity of the systems.
These key points within the SOC 2 process are designed to strengthen an organization's security posture, promote vigilance in monitoring for potential threats, and proactively address vulnerabilities to protect sensitive data and maintain compliance with industry standards.

Data Storage: implement secure and controlled data storage facilities, define, and enforce data retention policies, monitor data storage capacity and performance, establish procedures for secure data disposal when no longer needed.
Data Classification: define a data classification policy, categorize data based on its sensitivity and importance, implement controls based on data classification (e.g., access controls, encryption, and monitoring), regularly review and update data classifications.

Encryption: encrypt sensitive data in transit using secure protocols (e.g., TLS/SSL), encrypt data at rest, including databases and file storage, manage encryption keys securely and rotate them periodically, implement encryption according to the organization's data classification policy.
Access Management: implement role-based access control (RBAC) to restrict access to data and systems, enforce strong password policies and multi-factor authentication (MFA), regularly review and update user access permissions, monitor and log access to sensitive data and systems.

4. Research products from vendors

Our solution is designed to handle various data types effectively. It involves the extraction of metadata and context through integration, with both sets of information stored in a data repository. Subsequently, we create a data model and provide an interface for working with this data. Access to this product is granted to Data Stewards, Auditors, and Data Owners. Utilizing a categorized catalog, users can make diverse data-related decisions while ensuring data security.

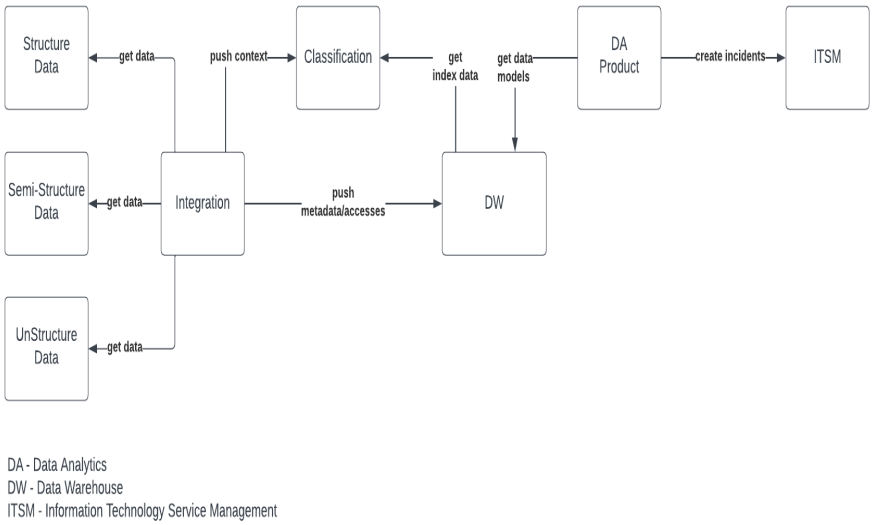

Figure 1. Document Register Framework

The proposed framework comprehensively addresses all the requirements of a Soc2 Type 2 document register. We provide flexibility in terms of technology selection, where to build the register, security measures for the framework, and ownership of this product. We are designing a document register strategy that encompasses the needs specified by the standard.

Advantages:
– Cost Effectiveness – you may build it on suitable environments cloud or on premise.

– Adapted to different sources per organization need.
– Don't need vendors support and development.
– Don't need investing on learning vendors product.
Disadvantages:
– Time to development.
– Team for development.

3. Conclusions

The modern world highly values information as a global resource. Safely storing this information requires reliability, security, and the ability to access data promptly. The continual growth of data volumes creates challenges related to selecting reliable methods and means of storage, as well as safeguarding this data from unauthorized access.

Having analyzed the main requirements of the standard, we propose a solution that encompasses all requirements concerning data classification, access control, and data monitoring. It also ensures a process for encrypting sensitive company data and reduces the risk of data breaches.

The suggested framework comprehensively addresses all the requirements for attaining the Soc2 Type 2 standard. It involves categorizing data based on its criticality, discerning which documents or data should be obfuscated or encrypted. We possess a clear understanding of data ownership, and we empower Data Stewards to manage and customize data policies within the organization.

REFERENCES:

1. AUJLA GS, CHAUDHARY R, KUMAR N, DAS AK, RODRIGUES JJ. SecSVA: secure storage, verification, and auditing of big data in the cloud environment. IEEE Commun Mag. 56(2018)1, 78-85.
2. VYAS J, MODI P. Providing confidentiality and integrity on data stored in cloud storage by hash and meta–data approach. Int J Adv Res Eng Sci Tech. (2017)4: 38-50.
3. DEIBE DAVID, AMOR MARGARITA, DOALLO RAMON. Big data storage technologies: a case study for web-based LiDAR visualization//IEEE International Conference on Big Data. – 2018. – P. 3831–3840.
4. KAITAI LIANG, WILLY SUSILO, JOSEPH K. LIU, "Privacy-Preserving Ciphertext Multi-Sharing Control for Big Data Storage," IEEE Transactions on Information Forensics and Security, vol.10, no.8, 2015.
5. SOC 2 Compliance Documentation: *https://secureframe.com/hub/soc-2/compliance-documentation/*, 20.10.2023.
6. SaaS Governance Best Practices for Cloud Customers: *https://securityxp.com/wp-content/uploads/2022/10/CSA-SaaS_Governance.pdf* , 20.10.2023.
7. ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection Information security management systems Requirement: *https://www.iso.org/standard/27001*, 20.10.2023.
8. File Analysis Software Reviews and Ratings. *https://www.gartner.com/reviews/market/file-analysis-software*, 20.10.2023.