EasyChair Preprint

№ 11681

Security and Privacy Challenges in Cloud-Based Data Warehousing: a Comprehensive Review

Sina Ahmadi

January 4, 2024

RESEARCH ARTICLE OPEN ACCESS

Security And Privacy Challenges in Cloud-Based Data Warehousing: A Comprehensive Review

Sina Ahmadi
Independent Researcher, USA

ABSTRACT

Nowadays, certain trends in technology have emerged, especially in cloud-based data warehousing. Organizations and associations use cloud-based data warehousing to store large amounts of data. However, this data warehousing type has many risks and challenges, such as privacy concerns. Some major security challenges are data breaches, malware attacks and data theft, which violates legal privacy frameworks, such as the Consumer Privacy Act. Certain measures like contractual agreements and data ownership can control these risks. The major object of this paper is to discuss the security and privacy challenges in cloud-based data warehousing used by private and government organizations. Some important challenges are complex cloud computing models, dynamic nature and interconnected ecosystems. The need for more resources is another major challenge for the companies which comes with the budgeting issues.

Keywords: **-** data warehousing, security, privacy, cloud

I. INTRODUCTION

Information technology has now become an important element in cloud-based data and warehousing. It has revolutionized the ways that different organizations and associations use to analyze vast datasets. Information technology has also brought certain challenges to accomplishing security and privacy matters. Due to this, it has become important to understand the benefits of cloud-based data storage and the risks of securing such sensitive information.

Cloud-based data or cloud computing plays a vital role in workflow management, providing many computing resources on demand to run data or compute critical applications. It also helps reduce costs and increase revenues while improving the quality of services. However, despite these advantages, cloud-based data is a huge area of concern that is creating restrictions and problems for organizations regarding security risks and malicious attacks. The complicated Infrastructure of cloud computing also creates risks, such as sharing stored data with other users or transferring sensitive data among other cloud components such as Data Centers. In some cases, the sensitive information is deduced from the cloud-based data and is leaked or sold to third parties by hackers. Therefore, cloud-based data warehousing has become a huge challenge for organizations to foster a secure and compliant data ecosystem.

II. CLOUD-BASED DATA WAREHOUSING OVERVIEW

A. Definition and Characteristics

Cloud-based data warehousing is a platform where the consumer can easily store a large amount of data and manage its platform and application for different purposes [1]. Some of the major examples are Google apps or IaaS-type services. There are two types of cloud-based warehousing: the private cloud and the hybrid cloud. The private cloud is the one that is handled by a single person or single organization and for a single business purpose.

On the other hand, a hybrid cloud system combines two or more infrastructures, whether public or community [2]. The cloud-based data is continuously growing in computing and has proven to be a major shift in the world of technology to improve scaling, agility, functionality, collaboration, and low-cost reduction. However, certain challenges related to cloud-based data have attracted the attention of certain service providers and researchers.

There are different types of cloud-based challenges or risks, which have been classified into two main categories: privacy and security. Both of these challenges affect the efficiency and reliability of cloud-based environments.

Figure 1: Cloud Data Warehouse [3]

B. Importance in Modern IT Infrastructure

The importance of cloud-based data warehousing in modern IT infrastructure cannot be pointed out, as it tackles critical difficulties while opening up unprecedented potential for enterprises across multiple industries [4]. Having a cloud data warehouse is essential for making fast, informed decisions. When necessary, you may extract insightful information from current, accurate, and enriched data thanks to its enhanced computing capabilities and streamlined data management.

On-premise data warehouses are more and more expensive as businesses gather more data. Expanding a business intelligence program significantly increases expenses because storage and computing cannot be acquired separately for on-premises systems. In contrast, data warehouse teams can purchase as much or as little processing power and storage as needed with cloud data warehouses. Furthermore, there is no need for networking, server rooms, or other hardware when using cloud data warehousing.

Cloud data warehouses are crucial in keeping up with increasing data sources. Businesses must combine ERP, CRM, social media, support, and marketing data while preserving speed and performance to make data-driven choices [5].

III. SECURITY AND PRIVACY CONCERNS IN CLOUD-BASED DATA WAREHOUSING

A. Overview of Security Challenges

Many companies are moving their workloads to the cloud to boost productivity and streamline processes. However, cloud computing can give businesses a competitive edge. However, it's crucial to proceed cautiously while implementing it before completely comprehending the associated risks. When shifting operations to these dynamic environments, a company might fail because it is unaware of the risks and recommended practices for cloud security.

Numerous security issues and challenges are associated with cloud computing. For example, a third-party provider gets data via the Internet and stores it in the cloud. This suggests that control and visibility over the data are restricted. It also raises the question of how secure it can be in an efficient manner. Everyone needs to be informed about their roles, the threats to cloud security, and the best practices for protecting the cloud.

Cloud computing is still revolutionizing customer offers and internal corporate procedures. Organizations can now more effectively establish remote working environments than previously, owing to the development of cloud computing architecture. It also gives teams the information and resources they need to work together.

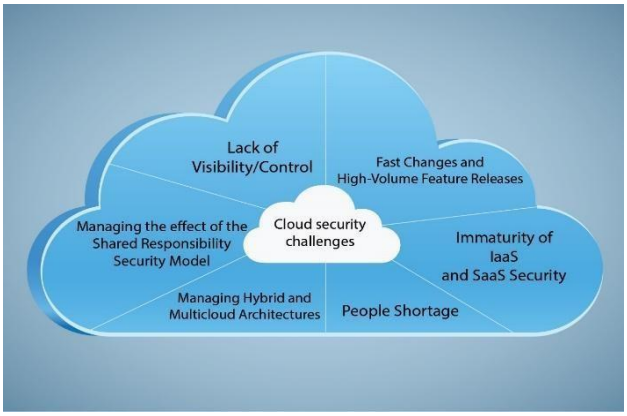

Figure 2: Cloud Security Challenges [6]

B. Overview of Privacy Challenges

The information technology industry's fastest-growing section is cloud computing, which started as a business concept. Customers' biggest concerns about cloud computing security are still tied to privacy and security concerns regarding data on the cloud. These concerns are what keep cloud computing services from taking off.

These challenges include privacy exposure, data leakage, and user data loss [7]. Several cloud computing and service models are used to increase a company's profit, financial growth, and return on investment. Cloud computing provides an appropriate environment and extra benefits or advantages for businesses or government organizations operating on the cloud. As a result, it has several issues with data security and privacy. The third parties manage and store the user's data offsite.

IV. LITERATURE REVIEW

Security Challenges in Cloud-Based Data Warehousing: The research study [8] emphasized the role of migrating a research data warehouse to a public cloud by highlighting significant opportunities and challenges [8]. This study aimed to determine the link of data warehouses with genomic pipelines so that complex data-driven strategies could be understood deeply. The researchers conducted a descriptive study where the primary materials were included, such as actual billing records of people, analysis documents, etc. This research showed that the challenges related to

storage architectures, networks, computing, etc., occur in cloud-based data warehousing. If an organization wants to enjoy the benefits of such data warehousing, it is also important to deal with these challenges.

Similarly, another study [9] also explained the security challenges related to cloud-based data warehousing. The researchers explained that cloud computing offers several benefits to different types of organizations, such as simplified IT infrastructure and management [9]. Large networks can be created with the help of cloud computing and an internet connection, but some security challenges are associated with such technologies. In this research study, the researchers surveyed to determine the cloud security issues. The findings of this study show a new classification of security solutions and explore different types of security threats related to cloud-based data warehousing.

Figure 3: Security Risks and Challenges for Cloud Computing [10]

Privacy Concerns and Regulatory Compliance: Different studies have been conducted on the privacy concerns and regulatory compliance of cloud-based data warehousing. In this case, [11] discussed the privacy and security challenges in cloud-based data warehousing. According to this study, the Internet is a vital resource in sharing data, whereas cloud computing offers certain hardware and software services [11]. The researchers have indicated that cloud-based data warehousing has been proven to be an important technology to improve certain organizations' operations. However, certain risks should be noted along with the benefits, including privacy challenges. These challenges mainly include compliance with jurisdiction, data protection regulations and data residency. In this case, organizations should take precautions in implementing enhanced privacy-preserving data techniques to avoid any privacy issues.

Authentication and Authorization Mechanisms: Authentication and authorization mechanisms related to data computing are an important area of the research. In this case, [12] discussed certain mechanisms of cloud-based data warehousing, mainly focusing on authorization and authentication. According to the researchers, it is best to use encryption techniques to protect the data warehouse in cloud networks [12]. It is important to create several firewalls and encryption codes to secure the data, as it can lead to the leaking of information and malware attacks. Therefore, the researchers have proposed a mechanism through which several encryption codes can be developed to protect the data, such as using dimension tables. The research findings suggested that an enhanced encryption model is the best way to protect the cloud network for data warehousing as it helps encrypt all the column names with the help of keys from a secure host.

Similarly, another research was carried out to understand the mechanisms of authorization and authentication in cloud security. In this case, [13] computed that cloud computing is an important network source that offers its users a high level of reliability. When users upload their data to cloud computing, it is stored in a data center [13]. However, it could be a more secure place as it can lead to data theft, privacy issues and data loss. During the study's findings, the researchers indicated that it is important to focus on identity management mechanisms to cope with the challenges of using cloud-based data warehousing. It is important to imply the protocols related to authentication and authorization in cloud computing architecture.

Emerging Trends in Cloud Security: Different trends have emerged in cloud security, which are necessary to understand and discuss. For this purpose, [14] discovered the emerging trends of security in cloud computing. The researchers mainly indicated that these trends are highly emerging in cloud computing in the computer industry. It also plays an important role in improving the overall welfare of an organization. In this study, the issues and challenges related to the increase in the number of users who want to adopt this technology have also been discussed. The researchers mainly used a mixed approach to collect the data and analyze the findings, which indicated the latest and future trends of cloud-based data warehousing. Some important trends related to cloud computing today are hybrid cloud computing, cloud-friendly enterprise frameworks, and using 547EB of data [14]. The researchers also suggested that organizations hire experts to defend the data against hackers firmly. However, the researchers also indicated some future trends in cloud computing, including quantum computing, hybrid cloud solutions, mobile cloud computing, and automation.

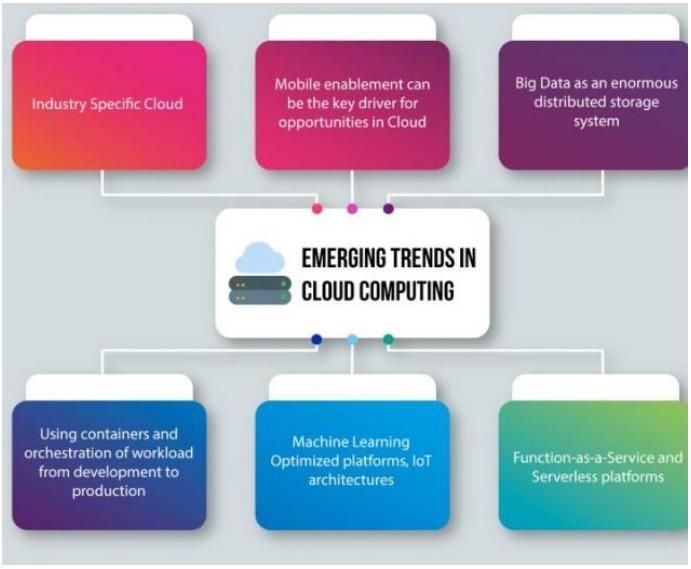

Figure 4: Emerging Trends in Cloud Computing [15]

Some other researchers have also discussed the emerging trends in cloud computing. For example, [16] indicated different levels of cloud-based data warehousing, which are important to understand to perform daily operations efficiently. Cyber attacks have become common; hackers usually steal information from government organizations and private associations [13]. One of the important reasons behind these attacks is the usage of wireless communication systems. Nowadays, people usually depend on electronic technology, but protecting that technology from cyber-attacks is a major issue. To resolve this problem, the study has suggested some of the latest developments, such as using cyberspace, as it is considered one of the important sources of power in the third millennium [16]. Furthermore, implementing different types of cyber security, such as network, cloud, application, and information security, can also help secure the data.

User Awareness and Training in Cloud-based Data Warehousing: Training is important in cloud-based data warehousing. Organizations need to develop different training programs in which they can train their employees and users to protect their data. In this case, [17] discussed that training the users plays an important role in enhancing the environment of the organization and using cloud-based data warehousing. Private or government organizations usually depend on their data, so keeping that data safe and secure is necessary [17]. The organization must provide accurate information to the users on how they can maintain the confidentiality of their data from any unauthorized access. The main purpose of this research was to indicate the significance of cloud storage and data management. However, the researchers discussed in their findings that data storage is important for data protection and record [17]. Therefore, it is necessary to give users proper training on protecting their data from bad data storage.

Another research was also conducted related to the importance of training for cloud-based data warehousing. In this case, [18] discussed the importance of human beings as they play a vital role in maintaining the privacy of a large amount of data. The researchers mainly focused on the data of healthcare organizations. They indicated that these organizations are facing problems securing their data as a large number of data breaches have occurred worldwide. In this study, the researchers have advised healthcare organizations to focus on implementing technical precaution measures for their electronic health records. It could be done by improving human behavioral interventions and introducing related training programs. The researchers mainly used exploratory analysis to conduct the results in which they showed that poor human security can lead to a large amount of data theft. It can also affect the previous records of healthcare organizations [18]. The research findings indicated that user awareness is the reason behind poor data security and phishing scams. In this case, it is important for the organization not just to hire security experts but also to give training to the users regarding data security and management.

V. PRIVACY ISSUES IN CLOUD DATA WAREHOUSING

A. Personal Data Protection

Cloud data warehousing requires the implementation of personal data protection measures. In this case, companies must implement different measures, such as adhering to legal requirements, using data minimization techniques, etc. All such actions help in gaining the trust of the clients.

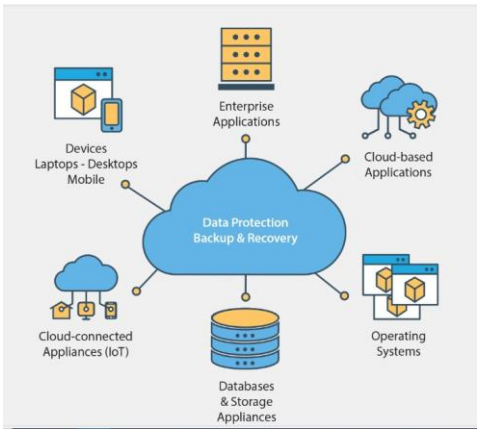

Figure 5: Personal Data Protection in Cloud Computing [19]

Legal Frameworks and Compliance: Technology firms should comply with legal frameworks to avoid security issues. The California Consumer Privacy Act is one such law that requires firms to protect the privacy of consumers and implement relevant security procedures [20]. Another important law is the GDPR or the General Data Protection Regulation. It also requires companies to use data protection measures to protect private information.

Data Minimization and Purpose Limitation: Data minimization mainly relates to collecting minimum data from

the clients. In this way, the chances of personal data breaches are reduced to a great extent. Therefore, firms should only ask for necessary data from the clients and avoid asking for unnecessary personal information. This is mainly called purpose limitation, which reduces the likelihood of breaches.

Encryption and Anonymization: Encryption mainly relates to implementing strong passwords or codes for accessing data. Thus, only authorized persons with the relevant passcodes can access the data. On the other hand, anonymization methods involve changing or eliminating clients' personal data. Both of these techniques protect the visibility of private data.

User Consent and Transparency: It is also highly significant for firms to gain consent from individuals and establish transparency measures while dealing with private data. In this case, they should always gain consent from clients before using their data for any purpose [21]. The clients should be informed of the purpose of using their data so that they can make an informed decision. In this process, transparency also helps in developing a strong level of trust among the clients. This way, they trust the firm with their private information and do not hesitate to share their data.

B. Data Ownership and Control

Data ownership is also an important method in cloud warehousing. Firms must identify data ownership before granting access to anyone. This level of control of the data assists in protecting the data in every way.

Ambiguities in Cloud Environment:

There are also some issues in cloud-based data related to data ownership. Sometimes, some pieces of information are shared by two or more users. In such situations, it takes work to identify the right owner. Companies should implement this practice very carefully to avoid data breaches.

Figure 6: Ambiguities in Cloud Environment [22]

Defining Access Controls and Responsibilities: Companies should always implement access controls to avoid data breaches [23]. In this case, access should be provided only to the relevant personnel. Not all kinds of data should be accessible to everyone. Overall, access controls help develop the proper organization of data while maintaining the highest levels of security.

Contractual Agreements and Service Level Agreements: It has been observed that using service level and contractual agreements also helps firms in overcoming security issues. These agreements help adhere to strict rules regarding using specific data. Therefore, no hackers or outsiders can access the data easily since strong security measures are adopted in compliance with the relevant agreements.

C. Adhering to Privacy Regulations

Companies implement privacy rules and regulations to ensure that clients' private information is reserved. No personnel member is allowed to access clients' private data without permission. Otherwise, legal action can be taken against anyone trying to breach the privacy rules.

Today, many firms adhere to global privacy regulations like the GDPR. It requires all companies to invest in strong privacy practices [24]. Another regulation in this regard is the Health Insurance Portability and Accountability Act. It also demands appropriate methods to maintain the accountability of sensitive data.

In compliance with these laws, companies also use innovative techniques like data classification and handling. These methods divide data into different categories based on their importance and sensitivity. In this case, many companies use methods like data masking and encryption. This is because they help in developing effective privacy rules.

VI. SECURITY AND PRIVACY MEASURES IN CLOUD-BASED DATA WAREHOUSING

A. Encryption and Data Masking

For companies involved in cloud data warehousing, it is very important to implement techniques like data masking and encryption. This is because, today, it has become a trend to hack people's sensitive information and use it for negative purposes. Therefore, companies must be highly conscious of personal data and use encryption to protect sensitive data.

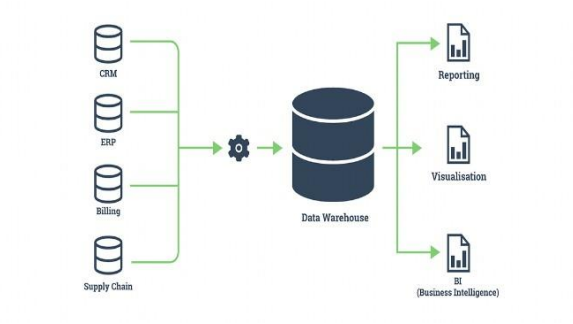

Figure 7: Data Masking in Cloud-Based Data Warehousing [25]

Encryption: Encryption is a very common but highly important technique in cloud-based data warehousing. In this method, the sensitivity is converted into an unreadable format. Therefore, hackers cannot read the data even if they gain access to it. This method has become popular among many firms and grants high security.

Data Masking: Data masking mainly relates to "masking" or "hiding" the private data in the cloud. For this purpose, a secure database is usually created. The data is not visible to hackers, who cannot leak personal information.

B. Access Controls and Authentication

It is crucial for technology firms to always focus on access controls of private databases. In this regard, an authentication procedure is utilized, which helps in protecting personal data. Strong passwords and other security protocols that cannot be easily breached are used in such cases. The use of access controls assists in limiting the number of people using a specific dataset. The fewer people use the data, the fewer the chances of breaches. This concept is very important to understand in the area of cloud warehousing. Only a single breach can put the firm's reputation at stake.

Companies should also use the authentication technique to secure their reputation in the industry. In this case, verification is always required when someone accesses the private data. With verification, access is granted to everyone. The verification can be done through different methods like strong passwords, biometrics, etc. All such measures add a level of security to the cloud data.

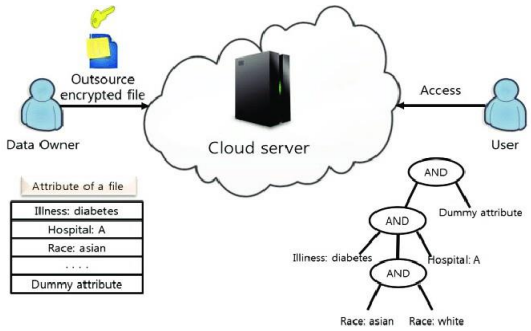

Figure 8: Access Control and Authentication in Cloud-based Data Warehousing [26]

C. Auditing and Monitoring

Companies involved in cloud data warehousing should also be aware of auditing and monitoring procedures [27]. These methods help implement transparency in security measures and detect areas of weakness. Regular auditing and monitoring procedures are crucial in identifying areas with weak security. In this case, strong security measures are immediately implemented in such areas to avoid any future breaches.

Comprehensive Audit Trails: As discussed earlier, detailed audit trails help companies track every single action taken related to the use of cloud data. Even a slight modification in sensitive data is detected immediately, and audit trails are conducted to identify the problematic areas. In this way, security violations are immediately detected, and strong security measures are put in place on time before any huge breaches.

Real-Time Monitoring: Companies should also implement real-time monitoring of cloud data warehousing [28]. If any unauthorized access occurs to sensitive data, it should be detected and monitored in real-time. The patterns of data access should be regularly monitored to identify the presence of hackers anywhere in the cloud.

Incident Response Readiness: Regular auditing mainly helps companies in preparing for any upcoming data breach incidents. Since audits help in identifying all the weak areas, the chances for future incidents of breaches are highly reduced. This is a proactive approach that helps companies develop effective strategies for the mitigation of security breaches.

VII. CHALLENGES IN IMPLEMENTING SECURITY AND PRIVACY MEASURES

Cloud-based data warehousing is a simple model for storing large data. However, it poses certain challenges when implementing security and privacy measures. One of the important challenges is saving the data from hackers and cyberattacks. The attackers mainly target the data sources and damage the whole system, which can compromise sensitive

information and cause reputational consequences to the organization. In this case, it is important to discuss these challenges in detail.

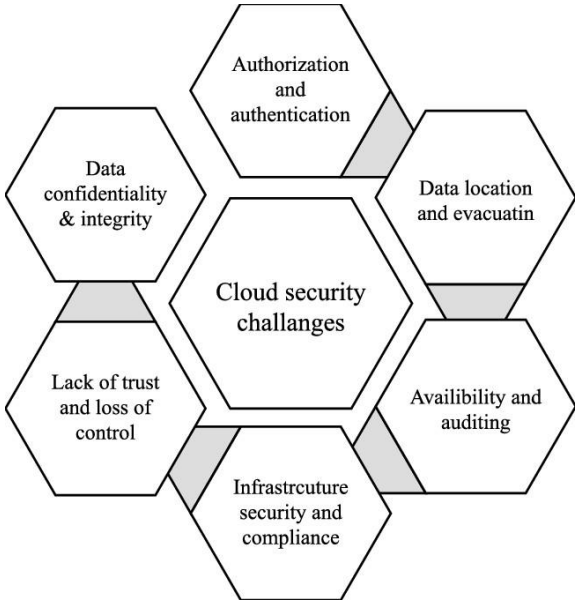

Figure 9: Challenges in Implementing Security and Privacy Measures [29]

A. Complexity of Cloud Environments

The cloud computing system is a critical system that is difficult for certain users to understand. The complex system of the cloud environments makes it a big challenge to implement security and privacy measures in cloud-based data warehousing systems. One of the main reasons is the multifaceted nature of the cloud infrastructure, which makes it more complex. It mainly involves diverse service models, shared resources, dynamic scalability and complexity of the navigation process.

Shared Responsibility Model: Shared responsibility is a complicated cloud security framework in which different security obligations of cloud computing are dictated. The main function of this model is to provide its users with the surety of accountability. One of the simple examples of this model is teaching the students both subject and pedagogical knowledge.

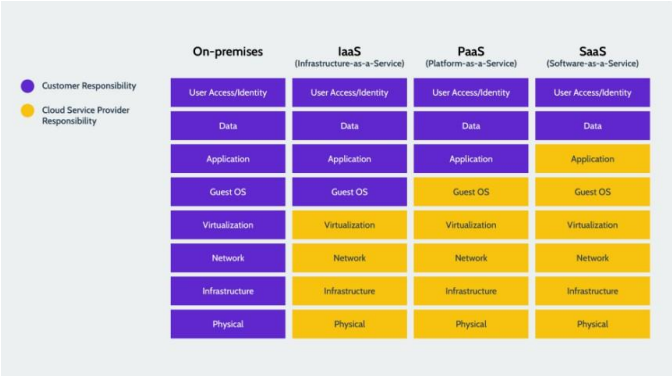

Figure 10: Shared Responsibility Model in Cloud-Based Data Warehousing [30]

Service Models: Cloud computing involves certain service models, making its Infrastructure more complex. Some models are Infrastructure as a Service (PaaS), Software as a Service (SaaS), etc. Each model has its challenges as they present unique security considerations. The organizations must understand the diversity of each of the service models before the implementation process.

Dynamic Nature and Scale: Cloud computing or cloud environments have a dynamic nature and scalability, which poses certain challenges in maintaining privacy and security controls. The resources in the cloud system are measured on different scales according to the user's demand. However, its dynamic nature makes it difficult and complex for the organization to ensure security measures, which require automation and continuous monitoring.

Interconnected Ecosystems: A cloud computing system usually uses a network of interconnected ecosystems such as platforms, infrastructure components and applications to provide an integrated cloud computing environment. This creates a huge challenge for imposing security and privacy measures, as securing the interconnectedness of vulnerabilities across the system is difficult.

B. Integration with Existing Systems

Organizations use cloud-based data computing warehousing systems to consume a large amount of data. Cloud-based data warehousing is an integrated system, which makes it more important. This Integration poses a firm challenge in the security and privacy implementation process. In this case, the organization needs to harmonize with all security protocols to make its system more cohesive and secure.

Diverse Technological Stacks: The systems that the organizations are already using operate on diverse technological stacks, each with a unique security framework. In this case, it is critical to understand each security framework while implying cloud-based data warehousing to maintain the security integrity of both cloud environments and legacy.

Data Migration Challenges: In cloud computing systems, a certain amount of data is migrated from legacy systems to the cloud. It mainly presents the complexities regarding the preservation of data integration during the transition. Therefore, organizations must ensure that sensitive information remains protected during migration, as it can lead to unauthorized access and data breaches.

C. Resource Constraints

The resource constraints are also a great challenge in implementing security and privacy measures in cloud-based data warehousing. The main reason is the budgeting system, technological incapabilities, and skilled personnel, which greatly challenge maintaining a comprehensive security posture.

Budgetary Limitations: Usually, medium or small-sized enterprises face budgeting challenges when implementing a new cloud-based system. It is difficult for them to allocate sufficient financial resources for robust security measures. To overcome this challenge, enterprises should focus on comprehensive security strategies such as employee training, continuous monitoring systems and smart investments in advanced tools.

VIII. LATEST TRENDS IN CLOUD SECURITY

Zero Trust Model: Different trends and models have been introduced to help secure information in cloud-based data warehousing. The zero trust model is one of the important and efficient models used to secure the data [31]. The model's main function is to pose strict identity security on the device. So, when a user wants to access resources on a device, they must undergo a strict identity verification process on a private network, whether outside or within the network's perimeter. The main principle behind this model is continuous monitoring and validation. It not only verifies the user's identity but also verifies the privileges and device security. Moreover, it also establishes a certain number of logins and connections time out, forcing the users to log out and verify themselves again after a certain amount of time. In this modern world of technology, it has become important for government and private organizations to implement this model.

Implementation of Artificial Intelligence and Machine Learning: Artificial intelligence has now become a vital part of data processing and securing. It can automate complex processes and decrease the downtime of finding the data by indicating maintenance needs [32]. It has improved the quality and effectiveness of the employee decision-making processes. Therefore, it is necessary to understand the importance of artificial intelligence and machine learning in securing the cloud infrastructure of certain organizations. Through AI and machine learning, organizations can easily automate security concerns as they can handle a large amount of data. AI automatically scans any threat or scamming and enhances the protection measures.

Furthermore, it is also important for organizations to use AI to manage network vulnerabilities as it identifies the weak points of security and helps to focus on critical security tasks [33]. It also handles repetitive security tasks easily by detecting security threats daily. By implementing AI, organizations can avoid certain errors that usually occur to humans or their boredom.

Cybersecurity Mesh: Cybersecurity mesh is another important emerging trend in cloud-based data warehousing. It is an important approach used in security control systems [34]. However, it is more efficient and flexible when it comes to hybrid multi-cloud infrastructures. Nowadays, different organizations are looking forward to growing in the global cybersecurity market. In this case, they are focused on expanding the business and resources outside the localized perimeters without compromising network security. In this case, a cybersecurity mesh can be an important tool to deal with data theft threats as it extends security in a hybrid Infrastructure, allowing all points of access and systems to be secured with a unified set of technologies. The consolidated dashboards in the cybersecurity mesh make it more responsible and effective. Therefore, the organization must use this tool. Moreover, data can be easily managed and controlled by taking advantage of this latest technology.

Secure Access Service Edge: Cloud-native security features like firewalls as a service, zero-trust network access, cloud access security brokers, and secure web gateways are all provided by the Secure Access Service Edge (SASE) architecture, which combines network and security as a service capability [35]. This functionality is offered as a service by the SASE vendor and is delivered from the cloud.

Through the combination of several security features, a software-defined wide area network (SD-WAN) or other WAN is secured as a whole for network traffic through the use of a SASE architecture [36]. If that's where users are, then legacy methods of inspection and verification—like sending traffic to data center firewalls via a multiprotocol label switching/forwarding service—work well.

IX. EMERGING TECHNOLOGIES AND THEIR IMPACT

Integration Platform-as-a-Service, or iPaaS, for Integration: iPaaS is used by large enterprises to integrate applications and data that reside in both public and private clouds and on-premises [37]. This makes it possible to use SaaS as a technological bridge for shared databases to create and implement intricate integration projects involving two or more connections. Traditional ETL gathers data from various systems into a single database, data store, and data warehouse to analyze and inform business decisions. Data sharing via API endpoints and security via data and authorization-based API policies are possible with iPaaS systems.

NoSQL for Big Data: When a database isn't structured and contains data in different formats, such as text, images, or videos, creating a schema can be difficult [38]. As a result, schema-less alternatives that offer more flexibility than SQL solutions have been developed to handle unstructured or big data. Since this is document-oriented rather than table-oriented (as in SQL), it is more flexible than traditional databases. Particularly when the data in the data warehouse is unstructured and changes often, big data, high-volume databases, and a wide range of online applications are well suited for NoSQL because a fixed schema model does not limit it and because it can scale horizontally, increasing compute and storage capacity.

Column-Based Storage for Advanced Analytical Query: The best choice for storing data in a cloud data warehouse for advanced analytics is column-based storage or columnar database [39]. This database management system stores data in columns instead of storing data in rows, as most relational databases do. In other words, every column in a table is kept independently, frequently in adjacent memory regions.

A columnar database's primary goal is to increase the effectiveness and speed of processes that require reading massive amounts of data. Different optimizations are possible because each column's data tends to be the same. Because the data in a column is comparable, columnar storage, for example, provides for improved data compression and more effective querying and aggregation, particularly in analytical and reporting jobs.

X. CONCLUSION

In conclusion, cloud-based data warehousing is an effective way to store large data. It helps in fulfilling the demands of the users in terms of securing the data. However, it comes with different security and privacy challenges that must be addressed to build the clients' trust. One of the important ways to maintain data integrity is data ownership, contractual agreements and access controls. However, different security and privacy challenges create hurdles for organizations in maintaining data integrity. Some challenges occur due to complex computing systems, integration issues and resource constraints. In this modern world of technology, many new technologies, such as iPaaS, column-based storage and NoSQL, offer innovative solutions to these challenges.

Moreover, cloud computing is dynamic and needs continuous monitoring and proactive security measures. Organizations should embrace advanced technologies and follow robust security protocols to cope with these challenges. This can help them ensure the integrity and confidentiality of the information in cloud-based data warehousing.

REFERENCES

[1] A. Nambiar and D. Mundra, "An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management," *Big Data and Cognitive Computing*, p. 132, 2022.

[2] M. Talaat, A. S. Alsayyari, A. Alblawi and A. Y. Hatata, "Hybrid-cloud-based data processing for power system monitoring in smart grids," *Sustainable Cities and Society*, p. 102049, 2020.

[3] Saras, "Top 3 Essential Drivers for Cloud Data Warehouse Adoption," 29 July 2022. [Online]. Available: https://sarasanalytics.com/blog/top3-reasons-to-own-a-cloud-data-warehouse/.

[4] J. Plašić, N. Stefanović and A. Gaborović, "Enterprise Business Intelligence Approach With Cloud-Based Analytics," *E-business technologies conference proceedings*, pp. 49-52, 2021.

[5] B. Sundarakani, A. Ajaykumar and A. Gunasekaran, "Big data driven supply chain design and applications for blockchain: An action research using case study approach," *Omega*, p. 102452, 2021.

[6] S. Maurya, "Data is shifting to cloud," 23 October 2021. [Online]. Available: https://www.suntechnologies.com/blogs/cloud-data-security-trends/.

[7] P. Sun, "Security and privacy protection in cloud computing: Discussions and challenges," *Journal of Network and Computer Applications*, p. 102642, 2020.

[8] M. G. Kahn, J. Y. Mui, M. J. Ames, A. K. Yamsani, N. Pozdeyev, N. Rafaels and I. M. Brooks, "Migrating a research data warehouse to a public cloud: challenges and opportunities," *Journal of the American Medical Informatics Association*, pp. 592-600, 2022.

[9] H. Tabrizchi and M. Kuchaki Rafsanjani, "A survey on security challenges in cloud computing: issues, threats, and solutions," *The Journal of supercomputing*, pp. 9493-9532, 2020.

[10] Edviser, "Eight Risks And Challenges For Cloud Computing," 21 August 2020. [Online]. Available: https://blog.skillmonks.com/eight-risks-and-challenges-for-cloud-computing/.

[11] I. S. M. Fadhil, N. B. M. Nizar and R. J. Rostam, "Security and privacy issues in cloud computing," *Authorea Preprints*, 2023.

[12] A. Arora and A. Gosain, "Mechanism for securing cloud based data warehouse schema," *International Journal of Information Technology*, pp. 171-184, 2021.

[13] J. V. Chandra, N. Challa and S. K. Pasupuletti, "Authentication and authorization mechanism for cloud security," *International Journal of Engineering and Advanced Technology*, pp. 2072-2078, 2019.

[14] N. Taleb and E. A. Mohamed, "Cloud computing trends: A literature review," *Academic Journal of interdisciplinary studies*, 2020.

[15] Nuvento, "Emerging trends in cloud computing – Hybrid Cloud, Mobile enablement and much more.," September 2019. [Online]. Available: https://nuvento.com/blog/emerging-trends-in-cloud-computing/.

[16] Y. Li and Q. Liu, "A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments," *Energy Reports*, pp. 8176-8186, 2021.

[17] O. Tanga, O. Akinradewo, C. Aigbavboa and D. Thwala, "Usage of Cloud Storage for Data Management in the Built Environment," *Advances in Artificial Intelligence, Software and Systems Engineering: Proceedings of the AHFE 2021 Virtual Conferences on Human Factors in Software and Systems Engineering, Artificial Intelligence and Social Computing, and Energy, July 25-29, 2021, USA*, pp. 465-471, 2021.

[18] L. H. Yeo and J. Banfield, "Human factors in electronic health records cybersecurity breach: an exploratory analysis," *Perspectives in Health Information Management*, 2022.

[19] Consiliant, "Data Protection, Backup & Recovery," 17 December 2023. [Online]. Available: https://www.consiliant.com/solutions/data-protection-backup-recovery/.

[20] E. L. Harding, J. J. Vanto, R. Clark, L. Hannah Ji and S. C. Ainsworth, "Understanding the scope and impact of the California consumer privacy act of 2018," *Journal of Data Protection & Privacy*, pp. 234-253, 2019.

[21] M. Ali, T. Wood-Harper and R. Ramlogan, "A framework strategy to overcome trust issues on cloud computing adoption in higher education," *Modern Principles, Practices, and Algorithms for Cloud Security*, pp. 162-183, 2020.

[22] A. Kazim and R. Varshney, "A Review: A Survey on Privacy Preserving for Secure Cloud Storage," October 2019. [Online]. Available: https://www.researchgate.net/figure/Security-threats-in-cloud-2-PRIVACY-IN-CLOUD-STORAGE-Preserving-Data-security-is-need-of_fig1_336615572.

[23] I. A. Mohammed, "The interaction between artificial intelligence and identity and access management: an empirical study," *International Journal of Creative Research Thoughts (IJCRT)*, pp. 668-671, 2021.

[24] M. L. Rustad and T. H. Koenig, "Towards a global data privacy standard," *Fla. L. Rev*, p. 365, 2019.

[25] R. Gupta, "THE NEED FOR DATA MASKING IN A DATA WAREHOUSE," 1 August 2023. [Online]. Available: https://dbsyncseo.medium.com/the-need-for-data-masking-in-a-data-warehouse-a97c196d4f06.

[26] J. Lee and J. W. J. Sungmin Oh, "A Work in Progress: Context based Encryption Scheme for Internet of Things," December 2015. [Online]. Available: https://www.researchgate.net/figure/Example-of-access-control-in-cloud-computing_fig3_282500797.

[27] R. Kumar and R. Goyal, "On cloud security requirements, threats, vulnerabilities and countermeasures: A survey," *Computer Science Review*, pp. 1-48, 2019.

[28] F. P. Knebel, R. Trevisan, G. S. do Nascimento, M. Abel and J. A. Wickboldt, "A study on cloud and edge computing for the implementation of digital twins in the Oil & Gas industries," *Computers & Industrial Engineering*, p. 109363, 2023.

[29] R. Masood, A. Shibli, Y. Ghazi and A. Kanwal, "Cloud authorization: exploring techniques and approach towards effective access control framework," April 2015. [Online]. Available: https://www.researchgate.net/figure/Security-challenges-in-cloud_fig2_273306239.

[30] Rewind, "The Shared Responsibility Model and SaaS, explained," 2020. [Online]. Available: https://rewind.com/shared-responsibility/.

[31] S. Mehraj and M. T. Banday, "Establishing a zero trust strategy in cloud computing environment," *2020 International Conference on Computer Communication and Informatics*, pp. 1-6, 2020.

[32] B. K. Mohanta, D. Jena, U. Satapathy and S. Patnaik, "Survey on IoT security: Challenges and solution using

machine learning, artificial intelligence and blockchain technology," *Internet of Things*, p. 100227, 2020.

[33] A. Bécue, I. Praça and J. Gama, "Artificial intelligence, cyber-threats and Industry 4.0: Challenges and opportunities," *Artificial Intelligence Review*, pp. 3849-3886, 2021.

[34] A. Samanta and S. S. Williamson, "A survey of wireless battery management system: Topology, emerging trends, and challenges," *Electronics*, p. 2193, 2021.

[35] S. van der Walt and H. Venter, "Research gaps and opportunities for secure access service edge," *International Conference on Cyber Warfare and Security* , pp. 609-619, 2022.

[36] S. N. L. K. Swamy, "A Study on Security Attributes of Software-Defined Wide Area Network," 2023.

[37] A. Trajkovska, T. Dimovski, R. Markoska and Z. Kotevski, "Automation and Monitoring on Integration ETL Processes while Distributing Data," 2023.

[38] H. Matallah, G. Belalem and K. Bouamrane, "Comparative study between the MySQL relational database and the MongoDB NoSQL database," *International Journal of Software Science and Computational Intelligence (IJSSCI)*, pp. 38-63, 2021.

[39] R. Liu, H. Isah and F. Zulkernine, "A big data lake for multilevel streaming analytics," *2020 1st International Conference on Big Data Analytics and Practices (IBDAP)* , pp. 1-6, 2020.

[40] B. Calzon, "13 Cloud Computing Risks & Challenges Businesses Are Facing In These Days," 6 June 2023. [Online]. Available: https://www.datapine.com/blog/cloud-computing-risks-and-challenges/.