The American Journal of Interdisciplinary Innovations and Research
ISSN 2642-7478 Volume 07 - 2025

SECURING CLOUD-NATIVE BIG DATA WAREHOUSES: A DISTRIBUTED SYSTEMS AND PRIVACY-PRESERVING ANALYTICS PERSPECTIVE

Prof. Isabela Correa
Department of Computer Engineering, University of Barcelona, Spain

Received: 12th Oct 2025 | Received Revised Version: 14th Oct 2025 | Accepted: 24th Nov 2025 | Published: 30th Nov 2025

Volume 07 Issue 11 2025 |

Abstract

The contemporary data ecosystem is defined by the explosive growth of heterogeneous, high-velocity, and high-volume datasets that are increasingly processed within cloud-native data warehousing platforms. These environments promise unprecedented scalability, elasticity, and analytical sophistication, yet they simultaneously introduce profound security and privacy challenges that extend well beyond the concerns of traditional on-premise data management. Distributed architectures, multi-tenant infrastructures, and complex data life cycles generate an intricate threat surface that demands systematic, theoretically grounded, and empirically informed approaches to protection. This article develops a comprehensive and original analysis of security and privacy in cloud-based big data warehousing by synthesizing perspectives from distributed systems theory, big data security scholarship, and modern data warehouse engineering practices. In particular, the architectural and operational principles articulated in contemporary cloud data warehouse platforms, as exemplified by Amazon Redshift, are treated not merely as engineering choices but as socio-technical constructs that reconfigure trust, accountability, and risk within data-driven organizations (Worlikar, Patel, & Challa, 2025).

The study begins by situating cloud-native data warehouses within the historical evolution of distributed systems, tracing how reliability, fault tolerance, and security principles originally developed for tightly controlled enterprise networks have been transformed by the rise of virtualized, globally distributed cloud infrastructures (Birman, 2005; Tanenbaum & van Steen, 2007). It then integrates big data security and privacy research that highlights the vulnerability of the entire data life cycle, from ingestion and storage to analytics and sharing (Koo, Kang, & Kim, 2020; Venkatraman & Venkatraman, 2019). Through a qualitative, literature-driven methodological design, this article interprets how architectural components such as shared-nothing clusters, columnar storage, massively parallel processing, and serverless elasticity alter the classical assumptions of access control, encryption, auditing, and trust boundaries.

By grounding its analysis in both distributed systems theory and modern data warehouse practice, including the operational recipes and architectural patterns discussed by Worlikar et al. (2025), this research contributes a holistic framework for understanding and governing security and privacy in the era of cloud-based analytics. The article concludes by outlining implications for system designers, data governance professionals, and researchers, emphasizing that future progress will depend not only on stronger cryptography or access controls but also on transparent architectures, accountable service models, and ethically informed data practices.

Keywords

Cloud-native data warehousing, Big data security, Distributed systems, Privacy-preserving analytics, Data governance, Cloud computing, Amazon Redshift

© **2025** Prof. Isabela Correa. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). The authors retain copyright and allow others to share, adapt, or redistribute the work with proper attribution.

Cite This Article: Prof. Isabela Correa. (2025). SECURING CLOUD-NATIVE BIG DATA WAREHOUSES: A DISTRIBUTED SYSTEMS AND PRIVACY-PRESERVING ANALYTICS PERSPECTIVE. The American Journal of Interdisciplinary Innovations and Research, 7(11), 110–117. Retrieved from https://theamericanjournals.com/index.php/tajiir/article/view/7344

INTRODUCTION

The last two decades have witnessed a profound transformation in how organizations collect, store, and analyze data. What began as enterprise data warehouses hosted on centralized servers has evolved into globally distributed, cloud-native platforms capable of processing petabytes of information in near real time. This evolution has been driven by both technological innovation and economic necessity: organizations seek to harness the value embedded in massive datasets while avoiding the prohibitive costs and inflexibility of traditional infrastructure (Venkatraman & Venkatraman, 2019). At the center of this transformation lies the convergence of big data analytics, cloud computing, and distributed systems, a convergence that fundamentally reshapes not only performance and scalability but also the nature of security and privacy (Koo et al., 2020).

From a theoretical standpoint, distributed systems have always posed unique challenges for security and reliability because they operate across multiple nodes that may fail, behave unpredictably, or even act maliciously (Birman, 2005; Lynch, 1996). Classical distributed systems theory was developed in contexts where nodes were often owned and controlled by a single organization, or at least operated within a relatively stable trust domain (Tanenbaum & van Steen, 2007). Cloud computing disrupts these assumptions by introducing virtualization, multi-tenancy, and geographically dispersed data centers that are owned and operated by third-party providers (Bos, 2019). In such an environment, the question of who controls data, who is responsible for protecting it, and how trust is established becomes significantly more complex (Anderson, 2008).

Big data compounds these complexities. Unlike traditional transactional databases, big data systems ingest information from a wide range of sources, including sensors, social media, enterprise applications, and external data brokers. These data streams often contain personal, sensitive, or commercially valuable information, and they are processed through complex pipelines involving extraction, transformation, storage, and analytics (Matturdi et al., 2014). Each stage of this life cycle introduces potential vulnerabilities, from insecure data ingestion points to improper access controls on analytical results (Gahi, Guennoun, & El-Khatib, 2015). As a result, security and privacy cannot be treated as afterthoughts but must be integrated into the very architecture of data platforms.

Cloud-native data warehouses such as Amazon Redshift exemplify this new paradigm. They combine massively parallel processing, columnar storage, and elastic resource management to provide high-performance analytics on large datasets (Worlikar et al., 2025). These platforms are designed to be accessible to a broad range of users, from data scientists to business analysts, and they abstract away much of the underlying infrastructure complexity. While this abstraction is a key source of their value, it also creates new challenges for security governance, as users may not fully understand where their data is stored, how it is replicated, or who has administrative access to it (Bertino, 2015).

The literature on big data security and privacy reflects a growing awareness of these challenges. Surveys and systematic reviews consistently identify issues such as data confidentiality, integrity, availability, and accountability as central concerns (Nelson & Olovsson, 2016; Ye et al., 2016). At the same time, scholars emphasize that traditional security mechanisms, which were designed for relatively static and bounded systems, are often ill-suited to the dynamic, scalable, and heterogeneous environments of cloud-based big data platforms (Alsulbi et al., 2021). Encryption, for example, can protect data at rest and in transit, but it does not automatically prevent misuse by authorized insiders or inference attacks on aggregated results (Lu et al., 2014).

Despite this rich body of work, there remains a significant gap in how security and privacy research is integrated with the practical realities of modern cloud-native data warehouses. Much of the big data security literature treats storage and processing platforms in an abstract or generic manner, without engaging deeply with the specific architectural patterns and operational practices that characterize systems like Amazon Redshift (Worlikar et al., 2025). Conversely, technical manuals and engineering-focused texts often emphasize performance optimization and cost efficiency while relegating security and privacy to a set of configuration options rather than treating them as core design principles (Gollmann, 2019). This disconnect limits our

ability to develop holistic, actionable frameworks for protecting data in contemporary analytics environments. This article addresses that gap by developing an integrated analysis of security and privacy in cloud-native big data warehouses that is grounded both in distributed systems theory and in the concrete architectures of modern platforms. By drawing on the comprehensive engineering perspective provided by Worlikar et al. (2025) alongside a wide range of security and privacy scholarship, this study seeks to move beyond high-level generalities toward a nuanced understanding of how technical, organizational, and regulatory factors interact. The central research problem can be articulated as follows: how do the architectural and operational characteristics of cloud-native data warehouses reshape the security and privacy landscape of big data analytics, and what implications does this have for governance and system design?

Addressing this problem is not merely an academic exercise. Organizations across sectors, from healthcare and finance to government and retail, increasingly rely on cloud-based analytics to inform strategic decisions. Breaches, misuse of data, or violations of privacy regulations can have severe legal, financial, and reputational consequences (Lafuente, 2015). At the same time, overly restrictive security controls can undermine the very value of big data by making it difficult to derive timely and meaningful insights (Venkatraman & Venkatraman, 2019). Understanding the trade-offs and synergies between security, privacy, and analytical performance is therefore essential for both practitioners and policymakers.

In the sections that follow, this article develops a comprehensive methodological and analytical framework to explore these issues. The methodology section explains how a qualitative, literature-driven approach can be used to synthesize insights across diverse domains, from cryptography and access control to cloud architecture and data governance (Cachin, Guerraoui, & Rodrigues, 2011). The results section then presents a detailed interpretive analysis of how security and privacy manifest across the data life cycle in cloud-native data warehouses, drawing on empirical and conceptual findings from the literature (Koo et al., 2020; Matturdi et al., 2014). The discussion section provides an extended theoretical interpretation of these results, engaging with competing scholarly perspectives and exploring the broader implications for the future of data-driven societies (Anderson, 2008; Bertino, 2015).

By situating Amazon Redshift and similar platforms within this broader intellectual landscape, this study aims to demonstrate that security and privacy are not peripheral concerns but central dimensions of modern data warehousing. They shape not only how data is protected but also how it is valued, shared, and ultimately used to make decisions that affect individuals and societies (Worlikar et al., 2025; Gahi et al., 2015). Through this integrated perspective, the article contributes to a more holistic and theoretically grounded understanding of cloud-native big data security.

METHODOLOGY

The methodological foundation of this study is rooted in a qualitative, interpretive research design that draws systematically on the existing body of scholarly literature in distributed systems, big data security, and cloud-native data warehousing. This approach is particularly appropriate for a domain characterized by rapid technological change, heterogeneous architectures, and complex socio-technical interactions, where purely quantitative metrics or isolated case studies would be insufficient to capture the full range of relevant dynamics (Nelson & Olovsson, 2016). By synthesizing theoretical insights, empirical findings, and engineering practices, the methodology seeks to construct a coherent analytical framework that can explain how security and privacy emerge in cloud-based big data environments.

A central pillar of the methodological rationale is the recognition that cloud-native data warehouses such as Amazon Redshift cannot be meaningfully analyzed in isolation from the distributed systems principles that underlie them (Worlikar et al., 2025; Tanenbaum & van Steen, 2007). These platforms are not simply databases hosted in the cloud; they are complex ecosystems of compute nodes, storage layers, networking components, and management services that operate under conditions of partial failure, latency, and asynchronous communication (Birman, 2005). Consequently, the study draws heavily on the distributed systems literature to frame issues such as fault tolerance, consistency, and trust, which are directly relevant to security and privacy (Lynch, 1996).

In parallel, the methodology integrates the extensive body of research on big data security and privacy, which provides conceptual models and taxonomies for understanding threats, vulnerabilities, and countermeasures across the data life cycle (Alsulbi et al., 2021; Koo et al., 2020). This literature is particularly valuable for identifying the points at which data is most

exposed, such as during ingestion from external sources, during transformation and aggregation, and during sharing with downstream users or applications (Matturdi et al., 2014). By mapping these stages onto the architectural components of cloud-native data warehouses, the study can analyze how specific design choices amplify or mitigate particular risks (Bertino, 2015).

The analytical process involved a close reading and thematic coding of the provided references, with particular attention to how they conceptualize security, privacy, and trust. For example, works such as Gahi et al. (2015) and Venkatraman and Venkatraman (2019) emphasize the multi-layered nature of big data security, spanning hardware, software, network, and organizational levels. These insights were juxtaposed with the architectural patterns described by Worlikar et al. (2025), such as shared-nothing clusters, elastic scaling, and managed service models, to identify points of convergence and tension. This interpretive synthesis allows for the generation of new theoretical propositions about how cloud-native data warehouses reconfigure traditional security paradigms.

An important methodological consideration is the avoidance of technological determinism. Rather than assuming that specific technologies inherently produce certain security outcomes, the study treats platforms like Amazon Redshift as socio-technical systems embedded in organizational, regulatory, and economic contexts (Anderson, 2008). This perspective is informed by the cyber security body of knowledge, which highlights the interplay between technical controls and human factors such as authentication practices, authorization policies, and accountability mechanisms (Gollmann, 2019; Jha, 2019). By incorporating these dimensions into the analysis, the methodology seeks to provide a more realistic and comprehensive account of security and privacy.

The study also adopts a comparative lens, drawing on a wide range of sources to identify both consensus and disagreement within the scholarly community. For instance, some authors argue that cloud computing enhances security by centralizing expertise and enabling economies of scale in protection (Lu et al., 2014), while others warn that it creates single points of failure and attractive targets for attackers (Lafuente, 2015). By systematically comparing these viewpoints, the methodology allows for a nuanced assessment that goes beyond simplistic claims about the benefits or risks of cloud-based data warehousing.

In terms of limitations, a literature-based methodology cannot capture the full diversity of real-world implementations or organizational practices. While the references provide rich conceptual and empirical insights, they inevitably reflect the contexts and assumptions of their authors (Nelson & Olovsson, 2016). Moreover, rapidly evolving technologies and regulatory environments mean that some findings may become outdated. However, by grounding the analysis in fundamental principles of distributed systems and security engineering, the study aims to produce insights that remain relevant even as specific platforms and tools change (Cachin et al., 2011).

The interpretive nature of the methodology also requires careful attention to bias and subjectivity. To mitigate this risk, the analysis draws on multiple sources for each major claim, seeking convergence across independent studies and theoretical frameworks (Ye et al., 2016; Matturdi et al., 2014). Where disagreements exist, they are explicitly acknowledged and explored rather than glossed over. This approach not only enhances the credibility of the findings but also reflects the inherently contested nature of security and privacy in complex technological systems (Bertino, 2015).

By combining distributed systems theory, big data security scholarship, and cloud-native data warehouse engineering, the methodology provides a robust foundation for the results and discussion that follow. It enables the study to move beyond surface-level descriptions of features or threats and to engage deeply with the structural and conceptual forces that shape security and privacy in modern data analytics environments (Worlikar et al., 2025; Venkatraman & Venkatraman, 2019).

RESULTS

The results of this study emerge from the systematic synthesis of distributed systems theory, big data security literature, and cloud-native data warehousing practices. They reveal a complex and often paradoxical landscape in which cloud-based analytics platforms simultaneously offer unprecedented opportunities for robust security while introducing novel and deeply intertwined risks (Bertino, 2015; Worlikar et al., 2025). Rather than yielding a single, linear conclusion, the analysis highlights a set of interrelated patterns that characterize how security and privacy are enacted across the data life cycle in cloud-native environments.

One of the most salient findings is that cloud-native data warehouses fundamentally alter the boundaries of trust.

In traditional on-premise data warehouses, organizations typically exercised direct control over hardware, networks, and administrative access, which allowed them to define relatively clear trust domains (Tanenbaum & van Steen, 2007). In contrast, platforms such as Amazon Redshift operate on infrastructure owned and managed by cloud providers, meaning that a significant portion of the security perimeter is effectively outsourced (Worlikar et al., 2025). This does not necessarily weaken security, but it does reconfigure accountability: organizations must now rely on contractual agreements, compliance certifications, and shared responsibility models to ensure that their data is protected (Anderson, 2008).

Another key result concerns the dual role of abstraction in cloud-native architectures. Abstraction layers, such as managed storage services, automated scaling, and serverless query execution, simplify system administration and reduce the likelihood of misconfiguration by end users (Bos, 2019). At the same time, these layers can obscure critical details about where data is stored, how it is replicated, and who has access to it, thereby making it more difficult for organizations to perform rigorous risk assessments or to demonstrate compliance with privacy regulations (Koo et al., 2020). This tension is particularly evident in multi-tenant environments, where data from different organizations may reside on the same physical infrastructure even if logical isolation mechanisms are in place (Matturdi et al., 2014).

The analysis also shows that security in cloud-native data warehouses is increasingly embedded in automated processes rather than manual controls. Encryption at rest and in transit, for example, is often enabled by default and managed by the platform rather than by individual users (Lu et al., 2014). Similarly, patch management, intrusion detection, and resource monitoring are typically handled by the cloud provider using centralized systems that benefit from large-scale data collection and machine learning (Bertino, 2015). These capabilities can significantly enhance protection against known vulnerabilities and common attack vectors, but they also create dependencies on the provider's operational integrity and transparency (Lafuente, 2015).

From a privacy perspective, the results highlight that risks extend well beyond unauthorized access or data breaches. In big data analytics, sensitive information can be inferred from aggregated or anonymized datasets through sophisticated statistical and machine learning techniques (Gahi et al., 2015). Cloud-native data warehouses, with their powerful analytical capabilities and ease of data integration, can inadvertently facilitate such inference attacks if appropriate governance and access controls are not in place (Venkatraman & Venkatraman, 2019). The literature suggests that traditional privacy-preserving techniques, such as simple anonymization, are often insufficient in these contexts (Koo et al., 2020).

Another important finding concerns the dynamic nature of data in cloud environments. Data in platforms like Amazon Redshift is not static; it is constantly being ingested, transformed, replicated, and queried by multiple users and applications (Worlikar et al., 2025). Each of these operations creates new copies or representations of data, increasing the potential attack surface and complicating efforts to track and control sensitive information (Nelson & Olovsson, 2016). This dynamicity challenges traditional security models that assume relatively stable data locations and access patterns (Birman, 2005).

Finally, the results underscore the importance of organizational and governance factors in shaping security outcomes. Even the most sophisticated technical controls can be undermined by poor authentication practices, inadequate authorization policies, or lack of accountability mechanisms (Gollmann, 2019; Jha, 2019). In cloud-native data warehouses, where access is often granted through web-based interfaces and programmatic APIs, the risk of credential theft, privilege escalation, and insider misuse is particularly acute (Anderson, 2008). The literature consistently emphasizes that effective security and privacy require not only robust technology but also clear policies, training, and oversight (Ye et al., 2016).

Taken together, these results depict cloud-native big data warehouses as environments in which security and privacy are deeply intertwined with architectural design, operational practices, and organizational governance. Platforms such as Amazon Redshift exemplify both the promise and the peril of this new paradigm: they offer powerful tools for protecting and analyzing data, yet they also demand new ways of thinking about trust, responsibility, and risk (Worlikar et al., 2025; Bertino, 2015).

DISCUSSION

The findings presented above invite a deeper theoretical and critical examination of what security and privacy mean in the context of cloud-native big data warehouses. Rather than viewing these concepts as static properties

that can be achieved through the application of specific technical controls, the literature suggests that they should be understood as emergent qualities of complex distributed socio-technical systems (Anderson, 2008; Birman, 2005). This perspective is particularly important when analyzing platforms such as Amazon Redshift, which integrate advanced distributed computing techniques with managed service models that blur traditional organizational boundaries (Worlikar et al., 2025).

One of the central theoretical tensions in the literature concerns the role of centralization versus decentralization in security. Classical distributed systems theory often emphasizes the benefits of decentralization for fault tolerance and resilience, arguing that systems with multiple independent nodes are less vulnerable to catastrophic failure (Lynch, 1996; Tanenbaum & van Steen, 2007). Cloud computing, however, introduces a form of logical centralization, in which vast amounts of data and computational power are concentrated within a small number of provider-operated platforms (Lafuente, 2015). From one perspective, this concentration enables providers to invest heavily in security expertise, infrastructure, and monitoring, potentially delivering higher levels of protection than most individual organizations could achieve on their own (Lu et al., 2014). From another perspective, it creates attractive targets for attackers and raises concerns about systemic risk and power asymmetries (Bertino, 2015).

The analysis of cloud-native data warehouses suggests that both perspectives contain important insights. On the one hand, managed platforms such as Amazon Redshift benefit from standardized security architectures, automated updates, and large-scale threat intelligence that can significantly reduce the likelihood of successful attacks (Worlikar et al., 2025). On the other hand, the reliance on a single provider for critical infrastructure means that a vulnerability or misconfiguration at that provider can have far-reaching consequences across many organizations (Gollmann, 2019). This duality underscores the need for shared responsibility models that clearly delineate the roles of providers and customers in maintaining security and privacy (Anderson, 2008).

Another key theme in the discussion is the evolving nature of privacy in big data analytics. Traditional privacy frameworks were largely developed in contexts where data was collected for specific, well-defined purposes and processed in relatively isolated systems (Matturdi et al., 2014). In contrast, cloud-native data warehouses enable the integration and analysis of diverse datasets for a wide range of purposes, often far removed from the original context of data collection (Koo et al., 2020). This raises profound ethical and legal questions about consent, purpose limitation, and the potential for harm through inference and profiling (Gahi et al., 2015). The literature reflects a growing recognition that technical measures alone cannot fully address these concerns. While encryption, access control, and differential privacy techniques can mitigate certain risks, they do not resolve the underlying tension between the drive to extract value from data and the obligation to respect individual rights (Venkatraman & Venkatraman, 2019). In cloud-native environments, where data can be easily shared and re-purposed, governance frameworks and regulatory oversight play an increasingly important role in shaping acceptable practices (Bertino, 2015).

The architectural characteristics of cloud-native data warehouses also invite reflection on the concept of accountability. In traditional IT environments, it was often relatively clear who was responsible for a given system or dataset: a specific department or organization owned the hardware, managed the software, and controlled access (Tanenbaum & van Steen, 2007). In the cloud, responsibility is distributed across multiple actors, including providers, customers, and third-party service integrators (Bos, 2019). This diffusion of responsibility can create gaps in accountability, particularly when security incidents occur or when regulatory requirements are violated (Gollmann, 2019).

Worlikar et al. (2025) highlight how modern data warehouse platforms provide extensive logging, auditing, and monitoring capabilities that can, in principle, support strong accountability. However, the effective use of these tools depends on organizational commitment and expertise, as well as on the willingness of providers to offer transparency and cooperation (Anderson, 2008). The literature suggests that without clear contractual and legal frameworks, technical audit trails may be insufficient to ensure meaningful accountability (Lafuente, 2015).

A further dimension of the discussion concerns the relationship between performance and security. Big data analytics thrives on the ability to process large volumes of data quickly and flexibly, often by distributing workloads across many nodes and by caching or replicating data for efficiency (Venkatraman & Venkatraman, 2019). These same mechanisms, however, can increase the risk of data exposure by creating multiple copies of sensitive information and by widening

the attack surface (Nelson & Olovsson, 2016). Cloud-native data warehouses seek to manage this trade-off through features such as role-based access control, network isolation, and encryption, but the tension remains inherent in the architecture (Worlikar et al., 2025).

Scholarly debate continues over whether it is possible to achieve both high performance and strong privacy in big data systems. Some researchers advocate for advanced cryptographic techniques, such as homomorphic encryption or secure multi-party computation, that would allow data to be analyzed without being decrypted (Lu et al., 2014). Others argue that these techniques are currently too computationally expensive for large-scale, real-time analytics and that more pragmatic governance and risk management approaches are needed (Bertino, 2015). The analysis presented here suggests that cloud-native data warehouses, with their vast computational resources, may eventually make some of these advanced techniques more practical, but their adoption will require careful integration with existing architectures and workflows (Worlikar et al., 2025).

The discussion also highlights the importance of viewing security and privacy as ongoing processes rather than one-time achievements. In cloud-native environments, software is continuously updated, resources are dynamically allocated, and data flows change rapidly (Bos, 2019). This dynamism means that security configurations and privacy safeguards must be constantly reviewed and adapted in response to new threats, technologies, and regulatory requirements (Koo et al., 2020). Static compliance checklists or periodic audits are unlikely to be sufficient in such a context (Gollmann, 2019).

Looking toward future research, the literature points to several promising directions. One is the development of more integrated security and privacy frameworks that span the entire data life cycle, from collection and ingestion to analytics and sharing (Alsulbi et al., 2021). Another is the exploration of how emerging technologies, such as blockchain or trusted execution environments, might be used to enhance transparency and trust in cloud-based data platforms (Ye et al., 2016). Finally, there is a growing need for interdisciplinary research that combines technical, legal, and ethical perspectives to address the societal implications of large-scale data analytics (Anderson, 2008; Bertino, 2015).

In this broader context, Amazon Redshift and similar platforms can be seen as both laboratories and battlegrounds for the future of data security and privacy. They embody the cutting edge of cloud-native analytics, yet they also expose the limitations of existing theories and practices (Worlikar et al., 2025). By examining these systems through the lenses of distributed systems theory and big data security scholarship, this study contributes to a more nuanced and critical understanding of how we might build data infrastructures that are not only powerful and efficient but also trustworthy and ethically grounded.

CONCLUSION

This article has developed a comprehensive and theoretically grounded analysis of security and privacy in cloud-native big data warehouses, drawing on distributed systems theory, big data security scholarship, and the practical architectures of modern platforms such as Amazon Redshift (Worlikar et al., 2025). The findings demonstrate that cloud-based analytics environments fundamentally reshape the boundaries of trust, the nature of accountability, and the dynamics of risk. Rather than simply transferring traditional data warehouses to a new hosting model, cloud-native platforms create new socio-technical configurations in which technical controls, organizational practices, and regulatory frameworks are deeply intertwined (Bertino, 2015; Anderson, 2008).

By synthesizing insights across multiple domains, the study shows that security and privacy in these environments cannot be reduced to a checklist of features or compliance requirements. They emerge from the interaction of architectural design choices, such as abstraction layers and elastic scaling, with governance mechanisms, such as access control policies and audit trails, and with broader social and legal expectations about data use (Koo et al., 2020; Gollmann, 2019). This perspective challenges simplistic narratives that portray the cloud as either inherently secure or inherently dangerous, instead highlighting the contingent and dynamic nature of protection in distributed data systems (Lafuente, 2015).

The analysis also underscores the importance of viewing modern data warehouses as distributed systems in the fullest sense of the term. Concepts such as fault tolerance, consistency, and trust, which have long been central to distributed computing, are directly relevant to understanding how data is protected and how privacy is preserved in cloud-native analytics platforms (Birman, 2005; Lynch, 1996). By reconnecting these theoretical foundations with contemporary engineering practice, as exemplified by Worlikar et al. (2025), the study offers a

more coherent framework for future research and practice.

Ultimately, the challenge of securing cloud-native big data warehouses is not merely technical. It is also organizational, legal, and ethical. As data becomes an ever more powerful driver of decision-making, the stakes of getting security and privacy right continue to rise (Venkatraman & Venkatraman, 2019; Gahi et al., 2015). Addressing these challenges will require not only stronger technologies but also more transparent architectures, more accountable service models, and more thoughtful governance. By illuminating the complex landscape in which these efforts must take place, this article contributes to the ongoing effort to build data infrastructures that are worthy of trust in an increasingly data-driven world.

REFERENCES

1. Koo, J., Kang, G., & Kim, Y.-G. (2020). Security and privacy in big data life cycle: A survey and open challenges. Sustainability, 12(24), 10571.
2. Birman, K. (2005). Reliable Distributed Systems. Springer.
3. Lafuente, G. (2015). The big data security challenge. Network Security, 2015(1), 12–14.
4. Venkatraman, S., & Venkatraman, R. (2019). Big data security challenges and strategies. AIMS Mathematics, 4(3), 860–879.
5. Worlikar, S., Patel, H., & Challa, A. (2025). Amazon Redshift Cookbook: Recipes for building modern data warehousing solutions. Packt Publishing Ltd.
6. Bos, H. (2019). The Cyber Security Body of Knowledge: Operating Systems & Virtualisation. University of Bristol.
7. Nelson, B., & Olovsson, T. (2016). Security and privacy for big data: A systematic literature review. Proceedings of the IEEE International Conference on Big Data, 3693–3702.
8. Matturdi, B., Zhou, X., Li, S., & Lin, F. (2014). Big data security and privacy: A review. China Communications, 11(14), 135–145.
9. Anderson, R. J. (2008). Security Engineering: A guide to building dependable distributed systems. Wiley.
10. Gollmann, D. (2019). The Cyber Security Body of Knowledge: Authentication, Authorisation & Accountability. University of Bristol.
11. Ye, H., Cheng, X., Yuan, M., Xu, L., Gao, J., & Cheng, C. (2016). A survey of security and privacy in big data. Proceedings of the International Symposium on Communications and Information Technologies, 268–272.
12. Gahi, M., Guennoun, M., & El-Khatib, K. (2015). Big Data Analytics: Security and Privacy Challenges. IEEE Communications Surveys & Tutorials.
13. Lu, R., Zhu, H., Liu, X., Liu, J. K., & Shao, J. (2014). Toward efficient and privacy-preserving computing in big data era. IEEE Network, 28(4), 46–50.
14. Cachin, C., Guerraoui, R., & Rodrigues, L. (2011). Introduction to Reliable and Secure Distributed Programming. Springer.
15. Tanenbaum, A., & van Steen, M. (2007). Distributed Systems: Principles & Paradigms. Prentice Hall.
16. Steen, M., & Tanenbaum, A. (2017). Distributed Systems. Prentice Hall.
17. Alsulbi, K., Khemakhem, M., Basuhail, A., & Eassa, F. (2021). Big data security and privacy: A taxonomy with some HPC and blockchain perspectives. International Journal of Computer Science and Network Security, 21(7), 43–55.
18. Bertino, E. (2015). Big data – Security and privacy. Proceedings of the IEEE International Congress on Big Data, 757–761.
19. Jha, S. (2019). The Cyber Security Body of Knowledge: Network Security. University of Bristol.
20. Lee, W. (2019). The Cyber Security Body of Knowledge: Malware & Attack Technology. University of Bristol.
21. Verissimo, P., & Rodrigues, L. (2001). Distributed Systems for System Architects. Kluwer.
22. Hartman, B., Flinn, D., & Beznosov, K. (2001). Enterprise Security with EJB and CORBA. Wiley.
23. Wang, C., et al. (n.d.). Secure Data Storage and Processing in Cloud Computing. IEEE Transactions on Cloud Computing.
24. Singla, A., & Goyal, V. (n.d.). Security in Distributed Systems. Journal of Network Security.
25. Lynch, N. (1996). Distributed Algorithms. Morgan Kaufmann.