INTERNATIONAL JOURNAL OF DATA SCIENCE AND MACHINE LEARNING (ISSN: 2692-5141)
VOL.06 ISSUE 02 (2026)

Research Article

Automated Security Governance for Non-Human Identities in Cloud Identity and Access Management

Arjun Mehta[1]

[1]Department of Computer Science and Information Technology, Indian Institute of Digital Systems, India

Received: 12 Jun 2026
Revised: 2 Jul 2026
Accepted: 20 Aug 2026
Published: 02 Sep 2026

Doi:10.55640/ijdsml-06-02-07

Page No: 160-168

Abstract

The rapid adoption of cloud-native applications, distributed analytics, edge computing, automated workflows, and machine-driven services has substantially increased the number of non-human identities (NHIs) operating within contemporary information systems. Unlike conventional human identities, NHIs include service accounts, application identities, workload credentials, automated processes, machine-to-machine connections, and other programmatic security principals. Their continuous operation, high privilege requirements, and large-scale interactions create governance challenges that conventional identity and access management approaches are not sufficiently designed to address. This research and review paper develops an automated security governance framework for NHIs in cloud Identity and Access Management (IAM), drawing theoretical and technical insights exclusively from the supplied literature on distributed computing, graph processing, parallel algorithms, machine learning, edge-cloud architectures, and data analytics. The methodology synthesizes these studies into an identity-centric governance model consisting of identity discovery, relationship modeling, risk analysis, policy evaluation, privilege optimization, continuous monitoring, and automated remediation. Graph-oriented research provides a theoretical basis for representing identities, resources, permissions, and dependencies as interconnected structures, while distributed and parallel computing studies support scalable analysis of large identity environments. The resulting framework demonstrates how automated governance can improve visibility, reduce excessive privileges, detect anomalous relationships, and support continuous security decisions. The analysis also identifies limitations related to graph complexity, dynamic cloud environments, policy interpretation, false positives, computational overhead, and incomplete identity inventories. The paper contributes a research-oriented conceptual architecture for treating NHI governance as a continuously computed security problem rather than a static access-control configuration.

Keywords: Non-Human Identities, Cloud IAM, Security Governance, Identity Graph, Automated Access Control, Privilege Management, Distributed Computing, Continuous Monitoring, Cloud Security

INTRODUCTION

Cloud computing has transformed enterprise architectures from relatively centralized environments into highly distributed ecosystems involving applications, microservices, containers, APIs, automated pipelines, edge workloads, and data-processing platforms. Within these environments, access is increasingly performed not only by human users but also by software components acting autonomously on behalf of applications,

organizations, and operational processes. These identities constitute a broad class of non-human identities (NHIs). Their security significance arises from the fact that an NHI can execute continuously, access multiple resources, inherit permissions through relationships, and potentially operate at a scale exceeding that of an individual human account.

Traditional IAM governance is frequently organized around identifiable users, authentication events, role assignments, and periodic access reviews. Such approaches become difficult to scale when thousands of applications and automated workloads communicate across distributed cloud infrastructure. The problem is not simply the number of identities; it is the complexity of relationships among identities, permissions, services, resources, and computational processes. Research on graph processing demonstrates that large computational environments can be represented and analyzed through relationships among interconnected entities, providing a useful conceptual foundation for modeling cloud identity ecosystems (Bouhenni et al., 2021; Fan et al., 2021).

The challenge is intensified by the heterogeneous nature of cloud workloads. Edge-cloud architectures distribute computation across multiple locations, while parallel computing frameworks support large-scale processing and dynamically changing workloads (Babar et al., 2022; Kaur et al., 2022). Consequently, NHI governance must operate across distributed environments rather than depend on a single centralized administrative perspective. Automated governance therefore requires mechanisms capable of discovering identities, determining their relationships, assessing access risk, and continuously enforcing appropriate security policies.

The security governance problem addressed in this paper is the absence of a unified conceptual mechanism that connects identity discovery, graph-based relationship analysis, scalable computation, automated policy decisions, and continuous remediation. Ganapathy (2025) specifically emphasizes automated governance and protection of NHIs within cloud IAM, providing the direct thematic foundation for this study. Building on that foundation, this paper investigates how principles from distributed graph analytics, parallel computation, machine learning, and data-processing research can be synthesized into an automated NHI governance architecture.

The primary objective is to formulate a technically grounded governance framework capable of treating NHI security as a continuously evaluated computational problem. The study focuses on identity visibility, relationship modeling, risk evaluation, privilege governance, anomaly identification, policy enforcement, and automated remediation. Its significance lies in connecting otherwise separate research areas into a coherent model for scalable NHI security governance.

LITERATURE REVIEW

The supplied literature does not primarily address cloud IAM directly; instead, it provides complementary technical foundations for designing scalable automated governance. This distinction is important because the proposed framework does not claim that the referenced studies directly solve NHI security. Rather, their computational principles are synthesized to address the architectural requirements of NHI governance.

Graph-oriented studies provide the strongest theoretical foundation. Bouhenni et al. (2021) investigate distributed graph pattern matching in massive graphs, demonstrating the importance of identifying meaningful structures within large interconnected datasets. For NHI governance, an identity environment can similarly be modeled as a graph in which identities, services, resources, permissions, and communication relationships constitute vertices and edges. A suspicious relationship can therefore be analyzed as a graph pattern rather than as an isolated permission record.

Fan et al. (2021) present GraphScope as a unified engine for big graph processing, while Hoang et al. (2021) investigate streaming edge partitioning for distributed graph analytics. These studies are particularly relevant to cloud IAM because identity relationships can change continuously. A workload may acquire a permission, invoke a service, create a temporary connection, or terminate within a short period. Static analysis may therefore become obsolete quickly. Streaming and distributed graph-processing principles support the concept of continuously updated identity graphs.

Fuchs et al. (2022) examine a transactional graph data structure, emphasizing the importance of consistent operations over graph data. This is relevant to security governance because identity relationships should not be evaluated from inconsistent snapshots. If a permission is revoked while an automated policy engine is simultaneously evaluating access, inconsistent graph state could produce an incorrect governance decision.

Kochsiek and Gemulla (2021) address parallel training of knowledge graph embedding models. Although their work concerns knowledge-graph computation rather than IAM, the underlying concept suggests that complex relationships can be transformed into computational representations suitable for large-scale analysis. This supports the possibility of deriving identity-risk representations from graph structure rather than relying solely on manually defined rules.

Hou et al. (2021) investigate massively parallel algorithms for personalized PageRank, while Czumaj et al. (2021) study graph sparsification in massively parallel computation. Together, these works provide important insights into scalability. An enterprise cloud may contain a very large number of identities and relationships; therefore, governance algorithms must reduce computational complexity while preserving security-relevant information.

Dhulipala et al. (2021) demonstrate that theoretically efficient parallel graph algorithms can achieve practical scalability. This is directly relevant to automated NHI governance because theoretical scalability alone is insufficient. A security architecture must perform effectively under operational workloads and changing data volumes.

The literature on distributed and parallel data processing provides an additional foundation. Kumar and Mohbey (2022) review parallel and distributed approaches to pattern mining, while Yin and Shi (2022) compare big-data computing and high-performance computing architectures. These studies suggest that NHI governance should employ parallel processing for large identity datasets instead of sequentially evaluating every identity-resource relationship.

Ibtisum et al. (2023) compare MapReduce and Apache Spark approaches to big-data processing. The architectural implication is that governance workloads can be decomposed into parallel stages, such as identity discovery, relationship extraction, risk calculation, and policy evaluation. Muthu et al. (2021), through deep-learning-based text summarization, and Jain et al. (2021), through CNN-LSTM-based sentiment analysis, demonstrate broader applicability of machine-learning pipelines to complex data-processing tasks. Their relevance to IAM is methodological rather than domain-specific: security governance can similarly use learned representations and classification pipelines where deterministic rules are insufficient.

Kasinathan et al. (2021) and Ngugi et al. (2021) investigate automated classification and recognition using machine-learning and image-processing techniques. Although their application domains differ, they demonstrate the value of automated recognition in environments where manual classification becomes impractical. Applied conceptually to NHI governance, automated classification could distinguish normal workload identities from dormant, excessive, anomalous, or high-risk identities.

Dafir et al. (2021) review parallel clustering algorithms for big data, while Gheisari et al. (2023) survey data-mining techniques for web mining. These studies support clustering and pattern-discovery approaches for identifying groups of identities with similar behavior or permissions.

Collectively, the literature reveals a significant research gap. Existing supplied studies provide mechanisms for distributed computation, graph analytics, clustering, pattern mining, machine learning, and edge-cloud processing, but they do not collectively establish an automated governance model specifically centered on NHIs in cloud IAM. Ganapathy (2025) directly establishes the NHI governance problem, but a scalable technical governance architecture requires additional computational principles. The present research addresses this gap through synthesis.

METHODOLOGY

3.1 Research Design

This study adopts a conceptual research-and-review methodology based exclusively on the supplied references. Rather than conducting an empirical experiment with a production cloud environment, the research synthesizes computational concepts into an automated NHI governance framework. The methodology follows a transformation process in which distributed graph analytics, parallel computing, clustering, pattern mining, and machine-learning principles are mapped onto IAM governance functions.

The framework is organized into seven logical layers: identity discovery, identity graph construction, contextual enrichment, risk analysis, policy evaluation, continuous monitoring, and automated remediation. The layers are sequential from a conceptual perspective but operate continuously in an actual cloud environment.

3.2 Identity Discovery

The first function is comprehensive discovery of NHIs. The objective is to identify service accounts, application principals, workload identities, automation identities, machine credentials, and associated access relationships.

An identity record can conceptually be represented as:

I = {identifier, workload, owner, permissions, resources, activity, lifecycle state}

The discovery layer collects these attributes and establishes a normalized representation. The major governance problem at this stage is completeness. An identity that remains undiscovered cannot be governed, regardless of how sophisticated the subsequent controls are.

3.3 Identity Graph Construction

The second layer represents the cloud IAM environment as an identity graph:

G = (V, E)

where V represents identities, applications, services, resources, permissions, and environments, while E represents relationships such as uses, accesses, invokes, inherits, or depends on.

For example, if application A uses service identity S, and S can access database D, the graph contains a path:

Application A → Service Identity S → Database D

The security significance of an identity is therefore determined not merely by its direct permissions but also by its reachable resources. Graph pattern matching provides a theoretical mechanism for detecting meaningful structures across such relationships (Bouhenni et al., 2021).

3.4 Distributed Graph Processing

Large organizations may generate identity graphs with millions of relationships. Centralized sequential processing can become a bottleneck. GraphScope and related distributed graph-processing research indicate the feasibility of partitioning and processing large graph structures across computational resources (Fan et al., 2021; Hoang et al., 2021).

A governance engine can therefore partition the identity graph according to workloads, applications, organizational domains, or resource boundaries. Parallel computation can then evaluate multiple identity segments simultaneously. This design is particularly suitable for distributed cloud and edge-cloud environments, where workloads may already be geographically and computationally distributed (Babar et al., 2022).

3.5 Risk Analysis

Risk analysis combines graph position, permissions, activity, resource sensitivity, and behavioral patterns. A conceptual NHI risk score can be expressed as:

$$R = w_1P + w_2A + w_3D + w_4B + w_5L$$

where P represents privilege exposure, A represents access breadth, D represents dependency depth, B represents behavioral deviation, and L represents lifecycle risk.

The weights are environment-specific rather than universal. An identity with extensive privileges may be legitimate if it belongs to a core infrastructure service, whereas the same privilege profile may be inappropriate for a temporary application.

Clustering methods provide a possible mechanism for identifying groups of identities with similar behavior or access patterns (Dafir et al., 2021). An identity that substantially deviates from its peer group can be assigned a higher investigation priority.

3.6 Policy Evaluation

The policy engine converts risk information into governance decisions. Instead of evaluating only static role assignments, it evaluates contextual relationships.

For example, a workload identity may normally access a storage service during automated processing. If the same identity suddenly attempts to access unrelated administrative resources, the graph structure and behavioral context can increase its risk score.

Parallel processing architectures are useful because policy evaluation may involve large collections of identity-resource relationships. MapReduce- and Spark-oriented approaches illustrate how large datasets can be divided into processing stages and executed at scale (Ibtisum et al., 2023).

3.7 Continuous Monitoring

Cloud identities are dynamic. Permissions, workloads, and dependencies can change rapidly. Therefore, governance cannot depend exclusively on periodic certification.

Streaming graph concepts are particularly relevant because newly created or modified relationships can be processed as graph updates rather than requiring complete

reconstruction of the identity environment (Hoang et al., 2021). Transactional graph concepts also highlight the need for consistency when security decisions depend on rapidly changing relationships (Fuchs et al., 2022).

3.8 Automated Remediation

The final layer converts governance decisions into controlled actions. Depending on organizational policy, remediation may include reducing privileges, disabling dormant identities, requiring additional approval, isolating an anomalous workload, or creating an investigation event.

Automation must nevertheless be risk-sensitive. Immediate remediation may be appropriate for clearly malicious or impossible relationships but dangerous for critical infrastructure identities. Therefore, a mature framework should distinguish between automatic enforcement, human approval, and monitoring-only actions.

RESULTS

The synthesis produces five principal findings regarding automated NHI governance.

First, graph-based identity modeling is a stronger conceptual representation than isolated permission records. An NHI rarely presents risk through a single permission. Its effective exposure emerges from combinations of identity relationships, inherited access, application dependencies, and reachable resources. Graph pattern analysis therefore provides a mechanism for identifying security-relevant structures that conventional tabular IAM reviews may overlook (Bouhenni et al., 2021).

Second, scalability is a fundamental governance requirement rather than an implementation detail. Distributed graph processing, parallel graph algorithms, and large-scale data-processing architectures indicate that security analysis can be distributed across computational resources (Dhulipala et al., 2021; Yin and Shi, 2022). This is particularly significant for organizations operating large cloud estates in which identity relationships may change continuously.

Third, continuous governance requires incremental processing. A periodic access review may identify excessive privileges after they have already existed for an extended period. Streaming-oriented graph processing provides a conceptual basis for analyzing changes as they occur, allowing governance decisions to respond more rapidly to new relationships (Hoang et al., 2021).

Fourth, machine learning and clustering can complement deterministic policies. Rules are effective for known conditions, but they may not identify unusual behavior that has not been explicitly encoded. Clustering and pattern-discovery approaches provide mechanisms for detecting deviations within large datasets (Dafir et al., 2021; Gheisari et al., 2023). Machine-learning research in the supplied literature further demonstrates how complex classification pipelines can automate recognition tasks (Jain et al., 2021; Kasinathan et al., 2021).

Fifth, automation must remain context-sensitive. A governance system that automatically revokes every unusual permission could create operational instability. Conversely, excessive reliance on human approval reduces scalability. The most appropriate result is therefore a risk-tiered architecture in which high-confidence risks are automatically remediated, ambiguous cases are escalated, and low-risk deviations are monitored.

Overall, the findings support the central proposition that NHI governance should be implemented as a continuously computed security process. The framework extends the governance perspective identified by Ganapathy (2025) by connecting NHI protection with scalable graph and data-processing mechanisms.

DISCUSSION

The findings indicate that automated NHI governance should be understood as a computational security discipline rather than merely an extension of traditional IAM administration. The theoretical contribution is the integration of identity governance with graph analytics. In conventional access management, the fundamental unit is often the identity-permission relationship. The proposed approach changes the analytical unit to the identity's position within an interconnected access graph.

This distinction has important practical implications. Consider a service identity with permission to access three databases. A conventional review may classify the identity according to those three permissions. A graph-based approach additionally considers which applications invoke the identity, which resources those applications depend upon, whether permissions are inherited, and what additional resources are reachable through connected relationships. Consequently, privilege risk becomes a structural property of the identity ecosystem.

The scalability problem represents the second major issue. Large graph environments cannot be evaluated efficiently through naive sequential methods. Research on massively parallel graph computation demonstrates that graph operations can be designed for distributed execution (Czumaj et al., 2021; Dhulipala et al., 2021). This supports a governance architecture in which identity analysis is partitioned across computational nodes. However, partitioning introduces its own challenge: security relationships can cross partition boundaries. A governance engine must therefore preserve cross-domain dependencies rather than treating each partition as completely independent.

A further trade-off concerns accuracy versus automation. Machine-learning and clustering approaches may identify unusual identities that deterministic policies miss, but anomaly does not necessarily mean malicious behavior. A newly deployed service may legitimately behave differently from established workloads. Therefore, automated governance should distinguish novelty from confirmed risk. This is consistent with the broader lesson from automated classification research that recognition systems require meaningful contextual features rather than relying on simplistic classification boundaries.

Another limitation concerns data quality. An identity graph is only as accurate as the identity inventory and relationship data used to construct it. Missing identities, incomplete permission information, stale workload relationships, and inaccurate ownership metadata can produce misleading risk assessments. Transactional graph concepts are relevant because security decisions should ideally operate on consistent representations of rapidly changing relationships (Fuchs et al., 2022).

The framework also introduces computational overhead. Continuous graph updates, risk scoring, clustering, and policy evaluation consume processing resources. Edge-cloud research demonstrates the broader challenge of balancing computational requirements across distributed environments (Babar et al., 2022). An organization therefore needs to determine which governance computations require real-time execution and which can be processed asynchronously.

The literature also exposes an important contradiction between centralized governance and distributed computation. Security policy benefits from centralized visibility and consistency, whereas computational scalability benefits from distributed execution. The proposed architecture resolves this conceptually by separating policy authority from computation. Policy definitions and governance objectives can remain centrally controlled while identity analysis is distributed.

Finally, the framework extends the conceptual direction of automated NHI protection described by Ganapathy (2025). The major contribution is not the claim that graph

processing, clustering, or parallel computation individually solves NHI security. Instead, the contribution is their integration into a unified governance model. This creates a foundation for future empirical studies that can measure risk-detection accuracy, remediation latency, computational overhead, and false-positive rates in real cloud environments.

CONCLUSION

Non-human identities have become fundamental components of modern cloud architectures, but their scale, automation, and interconnected access relationships create governance challenges that traditional identity-management approaches cannot adequately address on their own. This paper developed a conceptual framework for automated security governance of NHIs by synthesizing the supplied literature on graph processing, distributed computing, parallel algorithms, machine learning, clustering, and edge-cloud architectures.

The proposed framework treats the cloud IAM environment as a dynamic identity graph and organizes governance into identity discovery, graph construction, contextual analysis, risk assessment, policy evaluation, continuous monitoring, and automated remediation. Graph analytics provides the structural foundation for understanding indirect access relationships, while distributed and parallel computing provides scalability. Clustering and machine-learning concepts support behavioral classification and anomaly identification.

The principal research contribution is the positioning of NHI governance as a continuously computed security process rather than a periodic administrative activity. Such an approach can improve visibility, privilege analysis, and responsiveness while reducing dependence on manual reviews. However, the framework remains conceptual and must be validated empirically. Future research should evaluate the model using real cloud IAM datasets, investigate graph-partitioning strategies, quantify false-positive and false-negative rates, and establish measurable thresholds for automated remediation.

Ultimately, effective NHI governance requires a balance between automation, scalability, contextual awareness, and operational safety. The integration of graph-based reasoning with distributed security analytics provides a promising foundation for that balance and extends the broader objective of automated protection of non-human identities in cloud IAM (Ganapathy, 2025).

REFERENCES

1. M. Babar et al., ''An Optimized Iot-Enabled Big Data Analytics Architecture for Edge-Cloud Computing'', IEEE Internet of Things Journal, vol. 10, no. 5, pp. 3995–4005, 2022.

2. S. Bouhenni et al., ''A Survey on Distributed Graph Pattern Matching in Massive Graphs'', ACM Computing Surveys (CSUR), vol. 54, no. 2, pp. 1–35, 2021.

3. A. Czumaj et al., ''Graph Sparsification for Derandomizing Massively Parallel Computation with Low Space'', ACM Transactions on Algorithms (TALG), vol. 17, no. 2, pp. 1–27, 2021.

4. Z. Dafir et al., ''A Survey on Parallel Clustering Algorithms for Big Data'', Artificial Intelligence Review, vol. 54, no. 4, pp. 2411–2443, 2021.

5. L. Dhulipala et al., ''Theoretically Efficient Parallel Graph Algorithms Can Be Fast and Scalable'', ACM Transactions on Parallel Computing (TOPC), vol. 8, no. 1, pp. 1–70, 2021.

6. W. Fan et al., ''GraphScope: A Unified Engine for Big Graph Processing'', Proceedings of the VLDB Endowment, vol. 14, no. 12, pp. 2879–2892, 2021.

7. P. Fuchs et al., ''Sortledton: A Universal, Transactional Graph Data Structure'', Proceedings of the VLDB Endowment, vol. 15, no. 6, pp. 1173–1186, 2022.

8. M. Gheisari et al., ''Data Mining Techniques for Web Mining: A Survey'', Artificial Intelligence and Applications, vol. 1, no. 1, pp. 3–10, 2023.

9. L. Hoang et al., ''Cusp: A Customizable Streaming Edge Partitioner for Distributed Graph Analytics'', ACM SIGOPS Operating Systems Review, vol. 55, no. 1, pp. 47–60, 2021.

10. G. Hou et al., ''Massively Parallel Algorithms for Personalized Pagerank'', Proceedings of the VLDB Endowment, vol. 14, no. 9, pp. 1668–1680, 2021.

11. S. Ibtisum et al., ''A Comparative Analysis of Big Data Processing Paradigms: Mapreduce vs. Apache Spark'', World Journal of Advanced Research and Reviews, vol. 20, no. 1, pp. 1089–1098, 2023.

12. P. K. Jain et al., ''A Hybrid CNN-LSTM: A Deep Learning Approach for Consumer Sentiment Analysis Using Qualitative User-Generated Contents'', Transactions on Asian and Low-Resource Language Information Processing, vol. 20, no. 5, pp. 1–15, 2021.

13. T. Kasinathan et al., ''Insect Classification and Detection in Field Crops Using Modern Machine Learning Techniques'', Information Processing in Agriculture, vol. 8, no. 3, pp. 446–457, 2021.

14. M. Kaur et al., ''Multi-Level Parallel Scheduling of Dependent-Tasks Using Graph-Partitioning and Hybrid Approaches over Edge-Cloud'', Soft Computing, vol. 26, no. 11, pp. 5347–5362, 2022.

15. A. Kochsiek and R. Gemulla, ''Parallel Training of Knowledge Graph Embedding Models: A Comparison of Techniques'', Proceedings of the VLDB Endowment, vol. 15, no. 3, pp. 633–645, 2021.

16. S. Kumar and K. K. Mohbey, ''A Review on Big Data Based Parallel and Distributed Approaches of Pattern Mining'', Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 5, pp. 1639–1662, 2022.

17. L. C. Ngugi et al., ''Recent Advances in Image Processing Techniques for Automated Leaf Pest and Disease Recognition-A Review'', Information Processing in Agriculture, vol. 8, no. 1, pp. 27–51, 2021.

18. B. Muthu et al., ''A Framework for Extractive Text Summarization Based on Deep Learning Modified Neural Network Classifier'', Transactions on Asian and Low-Resource Language Information Processing, vol. 20, no. 3, pp. 1–20, 2021.

19. F. Yin and F. Shi, ''A Comparative Survey of Big Data Computing and HPC: From A Parallel Programming Model to A Cluster Architecture'', International Journal of Parallel Programming, vol. 50, no. 1, pp. 27–64, 2022.

20. Z. Wu et al., ''Recent Developments in Parallel and Distributed Computing for Remotely Sensed Big Data Processing'', Proceedings of the IEEE, vol. 109, no. 8, pp. 1282–1305, 2021.

21. Ganapathy, S. K. (2025). Automated Governance and Protection of Non-Human Identities in Cloud IAM. Frontiers in Emerging Computer Science and Information Technology, 2(02), 08–20. Retrieved from https://irjernet.com/index.php/fecsit/article/view/cloud-iam-non-human-identities.