International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

AI-Driven Operational Intelligence and Lineage Automation in Distributed Federal Cloud Systems

Aditya Swaprakash Gadepalli Sri Pratyak

Submitted: 02/03/2022 **Revised:** 07/05/2022 **Accepted:** 15/05/2022

Abstract : Federal cloud systems have seen the introduction of Artificial Intelligence (AI) more than ever before — but pre-2022, that meant that the monitoring of these systems and their data lineage tools were manual and disconnected. This is a secondary study reviewing literature published in Google Scholar, IEEE Xplore, ScienceDirect, Springer and government reports that investigates operational intelligence and lineage automation powered by AI in a Distributed Federal Cloud System. The key takeaways highlight faster anomaly detection, real-time analytics capabilities, and the decrease in manual effort required, as well as lineage automation for boosted data traceability, higher audit preparedness, and better compliance with regulatory requirements such as FedRAMP and NIST standards. These technologies combined provided security, governance and operational benefits. Yet, there was always a hurdle to jump like integration complexity, lack of data silos, lack of skills, and algorithmic bias issues. Overall, AI and intelligent automation have proved to be valuable for advancing federal cloud operations before 2022, but fragmented governance and interoperability challenges held back their full value. Standardization and transparency will be essential to future progress of standardized policies and explainable AI.

Keywords: *Artificial Intelligence, Operational Intelligence, Data Lineage Automation, Federal Cloud Systems.*

1. Introduction

1.1 Background of the Study

Cloud computing has now found its way into the operations of the federal government since this technology assists agencies to effectively store, retrieve and manipulate data that is in bulk. Most of the federal organisations embraced a distributed cloud system pre-2022 to enhance the scalability, flexibility, and digital services. These systems were however made more complicated as the usage of several cloud platforms, mass data streams and networked networks. It was this growing complexity that gave rise to a great demand for operational intelligence systems that would help track its performance and identify issues on run-time [1]. Meanwhile, technologies of Artificial Intelligence (AI) began to assist automation, enhancing monitoring of the system, predictive analysis, and data management procedures.

1.2 Problem Statement

The distributed federal cloud system complicated the management of cloud systems since the existing monitoring systems were usually slow and highly performed manually. Agencies struggled to track data flow, have correct data lineage and have a transparent system. Besides that, federal organisations were also required to comply with rigorous security and compliance regimes, further pressure was exerted on their operations. The human element was also impacted negatively, as

Sr Technical Director and Sr Solutions Architect

Global Alliant Inc. USA.

AI and Healthcare Technology Expert.

PrakashA.GadepalliSP@outlook.com

operational processes were being done manually in order to decrease efficiency and probability of a human error.

1.3 Research Aim

This study is designed to discuss the operational intelligence AI-enhanced lineage automation that enhanced management, monitoring, governance, and efficiency of distributed federal cloud systems up to 2022.

1.4 Research Objectives

- To examine the AI use in operational intelligence.
- To assess automation of lineages.
- To see the positive aspects of federal cloud systems.
- To explore issues of implementation.

1.5 Research Questions

Q1. What was AI applied to in federal cloud systems in terms of operational intelligence?

Q2. How did lineage automation contribute to governance and compliance?

Q3. What were the key advantages and problems prior to 2022?

1.6 Significance of the Study

The importance of the current study is that it outlines how AI can be utilized to transform the government digitally, enhance cybersecurity, and govern the clouds [2]. It also renders valuable learning content in responding to future research on cloud infrastructure systems which are intelligent.

2. Literature Review

2.1 Distributed Federal Cloud Systems

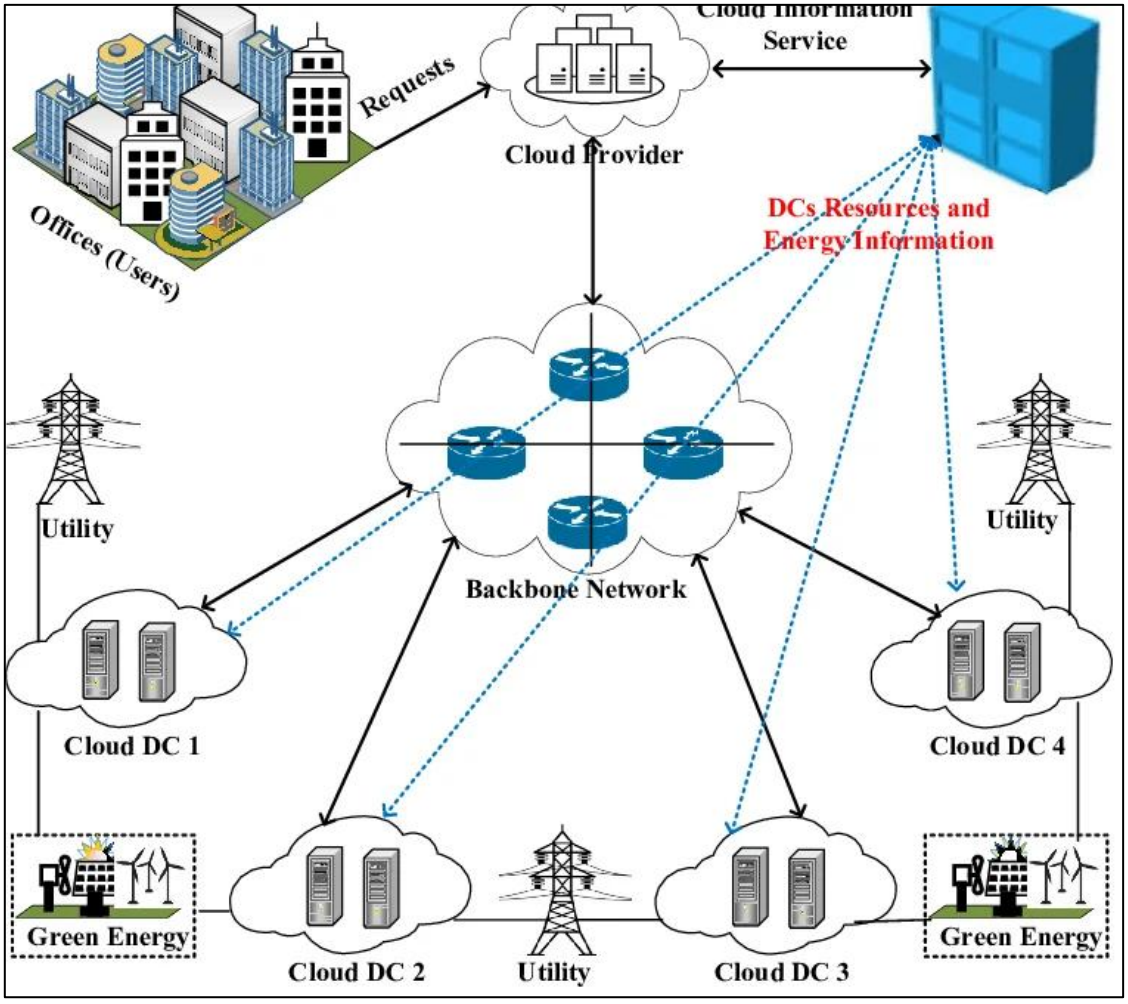

Figure 1: Architecture of geographically distributed cloud data centers

(Source: https://www.researchgate.net)

Distributed federal cloud systems can be defined as cloud computing systems where data, programs, and services are distributed among several interconnected servers and sites. The federal agencies embraced the implementation of these systems in order to enhance the efficiency of digital services, storage of data and efficiencies in their operations. Prior to 2022, most government organisations deployed multi-cloud and hybrid cloud architecture in order to integrate both the public and the private cloud. These strategies enhanced the flexibility, scalability and resource management [3]. Interoperability was also facilitated by distributed systems whereby various information systems within the federal government were able to communicate and share data successfully. Cloud migration within the government was amplified as the agencies needed solutions that are secure and economical to handle them in bulk. There was also the introduction of cloud orchestration tools that automated processes of managing workloads, enhanced coordination of the systems, and sustain service performance even in a complex distributed environment.

2.2 AI-Driven Operational Intelligence

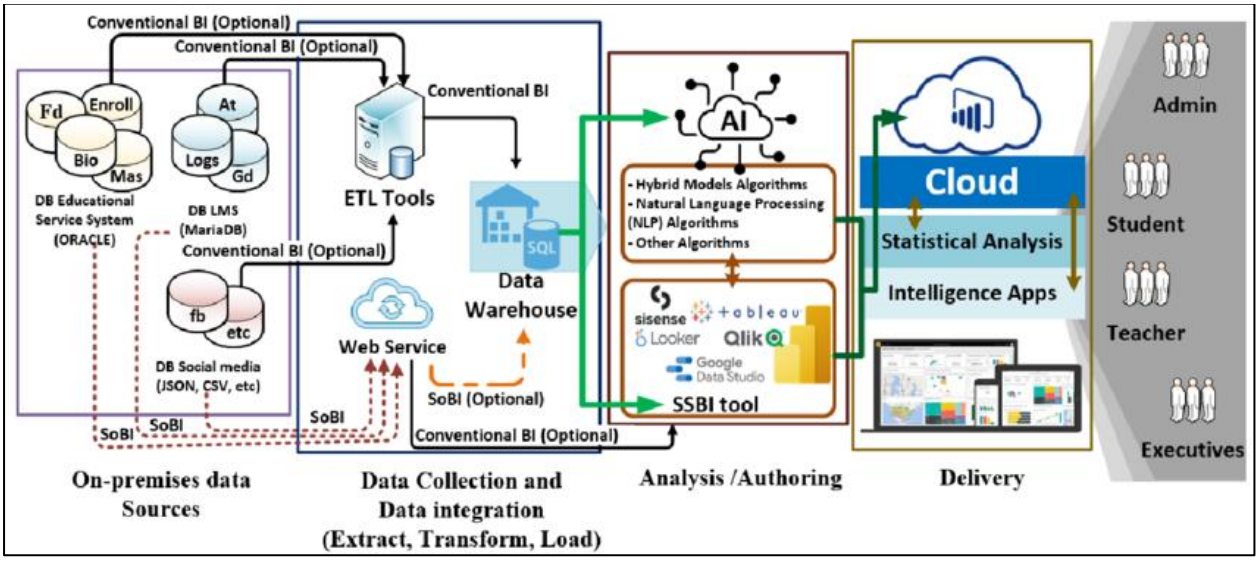

Figure 2: The architecture of the Business Intelligence management system

(Source: https://encrypted-tbn0.gstatic.com)

AI-driven operational intelligence is the practice of using technologies based on artificial intelligence and machine learning to oversee, examine and enhance system processes in real-time. The federal cloud systems processed vast volumes of operational data using AI to point out patterns that would in turn have been difficult to spot during the manual process [4]. Predictive analytics assisted agencies in predicting system failures and cutting down on downtime prior to occurrences of the problem. The use of AI in the detection of anomalies enhanced cybersecurity by detecting suspicious activities and threats. Log analytics provided the accelerated system logs and network events analysis. The automated incident detection minimized the number of hands-on to monitor as well as accelerating response. Smart automation also facilitated self-healing of infrastructures where the infrastructures could automatically rectify some technical problems. The technologies addressed distributed federal cloud challenges by enhancing operational efficiency, reliability and decision-making.

2.3 Data Lineage and Lineage Automation

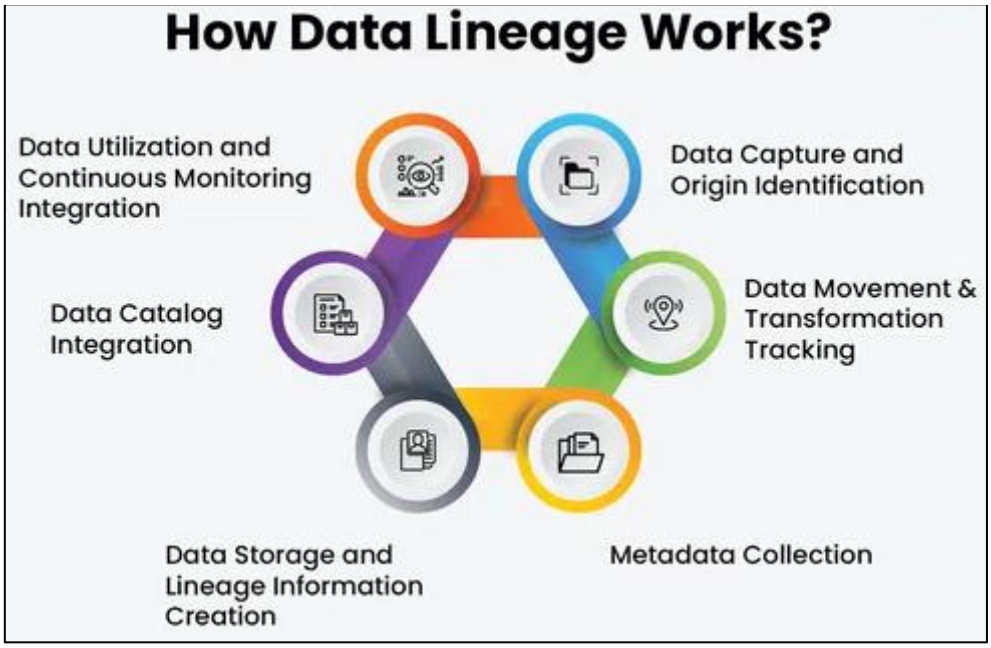

Figure 3: Data Lineage and Lineage Automation

(Source: https://encrypted-tbn0.gstatic.com)

Data lineage is the journey of data, through who gets involved in it over time and its pathways through systems. Lineage will play an important role in a compliance audit or forensic investigation in a federal cloud environment. Lineage automation eliminates manual tracking and automatically adapts itself to map the data flow using AI features, across distributed nodes [5]. When coupled with automated provenance tracking, every data transaction is recorded, which helps to support data transactions with governance, such as FedRAMP. AI-powered lineage mapping also helps in this aspect by mapping undocumented relationships and notifying administrators of lineage gaps. With workflow automation, these capabilities work seamlessly throughout the various processes in everyday work, minimising the risk of human error. This means agencies have visibility, audit turnaround and can prove data integrity across complex multi-cloud environments - all in real-time.

2.4 Security and Compliance in Federal Cloud Systems

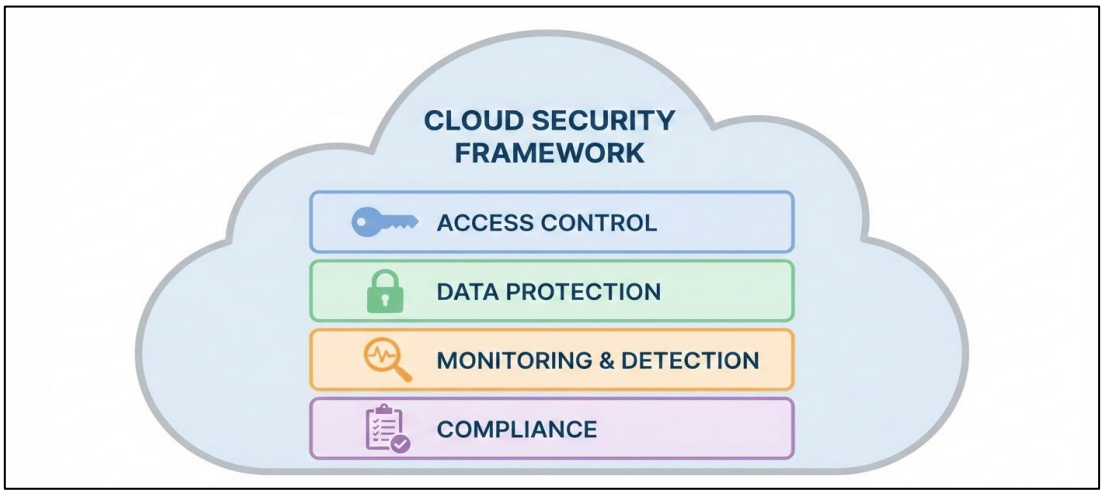

Figure 4: 7 Cloud Security Frameworks to Protect Digital Assets

(Source: https://tuxcare.com)

Federal cloud systems have to fulfill high security requirements, such as FedRAMP authorization and NIST-suggested risk-management ways. Issues anticipated prior to 2022 were compliance-related, such as areas where compliance policies were not consistently followed between hybrid and cross-cloud environments, and non-conforming audit trails [6]. No trust was assumed and an ever-constant verification of each access request was the solution: the Zero-trust architecture. Imported with AI-driven threat intelligence to complement this, and identified behavioral patterns for zero-day exploits and insider threats. Automated classification and encryption of data and access control to sensitive workloads became a requirement for data governance policies. However, the adoption of legacy systems and models was still challenging for zero trust adoption. However, with AI-driven security analytics, the ability to do real-time threat detection and auto-generated security compliance reports is greatly enhanced.

2.5 Challenges Identified in Previous Studies

Six major challenges for AI-driven federal cloud systems have been identified in earlier studies before 2022. The many past federal structures were the primary hurdles, simply because they seemed to be resistant to working alongside more contemporary artificial intelligence surveillance systems. Data privacy issues emerged due to data distributed around nodes, with sensitive data being shared for AI training. Compliance audits were critical to the fairness of decision-making, and were threatened by algorithm bias. Federal auditors struggled to ensure the validity of AI-driven decisions because of the lack of transparency in AI models.Federal auditors faced challenges due to limited transparency in AI models, making it difficult to ensure the validity of AI-driven decisions [7]. The costs of scaling up the infrastructure needed for AI implementation posed a challenge for several agencies. Last but not least, cloud clouds lacked interoperability between cloud providers and made it difficult to get a unified operational intelligence. These obstacles helped hinder the rollout of 100% automated lineage and governance technology solutions.

2.6 Research Gap

As of 2022, there is a paucity of literature that crosses the borders of operational intelligence and lineage automation as an integrated system to support federal cloud systems in the pre-2022 published literature. The majority of the works evaluated only manual lineage tracking or only auto-monitoring of AI, but not with the combination of both. In addition, there are no consistent IT Governance approaches available to take a holistic view of AI both during its use and in hindsight, regarding its accountability, efficiency, quality, and reliability. There is not enough research work dedicated to specially deal with the implementations required in Federated Arch having security and compliance requirements that are very different from those found in commercial clouds. Prior to 2022, there wasn't a standard architecture or a best practice guideline to include automated lineage in AI-driven operational intelligence in a distributed federal way.

3. Methodology

3.1 Research Design

In this research, the secondary qualitative research is used, which in this study is a literature-based analytical study. It conducts a systematic review of the existing research from peer-reviewed literature, conference papers and technical reports, and generates a synthesis of knowledge related to operational intelligence and lineage automation with the help of AI. Primary data collection was not done. The design helps to perform multiple data source comparisons of results without experimental intervention.

3.2 Data Sources

Peer-reviewed scientific articles from Google Scholar, IEEE Xplore, Science Direct and Springer databases were used as data sources. In addition, some government resources such as NIST and industry resources like cloud service providers and federal consulting companies were also incorporated. Until 2022, these sources helped provide -- perhaps more than any others -- an expansive view of theoretical models and federal cloud integrations.

3.3 Inclusion Criteria

The studies included were published up through 2022, were based on peer-reviewed research published in journals, conference papers and reputable governmental or industry reports. The studies needed to include federal cloud systems, distributed computing, intelligent monitoring or automatic data lineage studies. Only those publications that explicitly relate issues of operational intelligence in federal settings were

considered for analysis—that is, English-language publications.

3.4 Exclusion Criteria

Non-English papers, studies from after 2022 and papers that were data only about the commercial cloud were excluded. These are not included in the sample since they are opinion pieces, editorials or non-peer reviewed blog posts [8]. In addition, research was excluded from review that was specifically geared toward applications of general AI (GAI) that are not directly operational or lineage specific to distributed federal systems.

3.5 Data Analysis Method

Using the selected literature, thematic analysis was used to identify any repeated patterns, concepts and findings to find themes. A comparative review framework allowed for the crossing of studies which included benefits, challenges, and strategies for implementation. Systematic coding and categorisation of key themes identified, including aspects of AI monitoring performance, benefits of automation of lineage, impacts on security and barriers identified.

4. Results / Findings

4.1 AI Improved Operational Monitoring

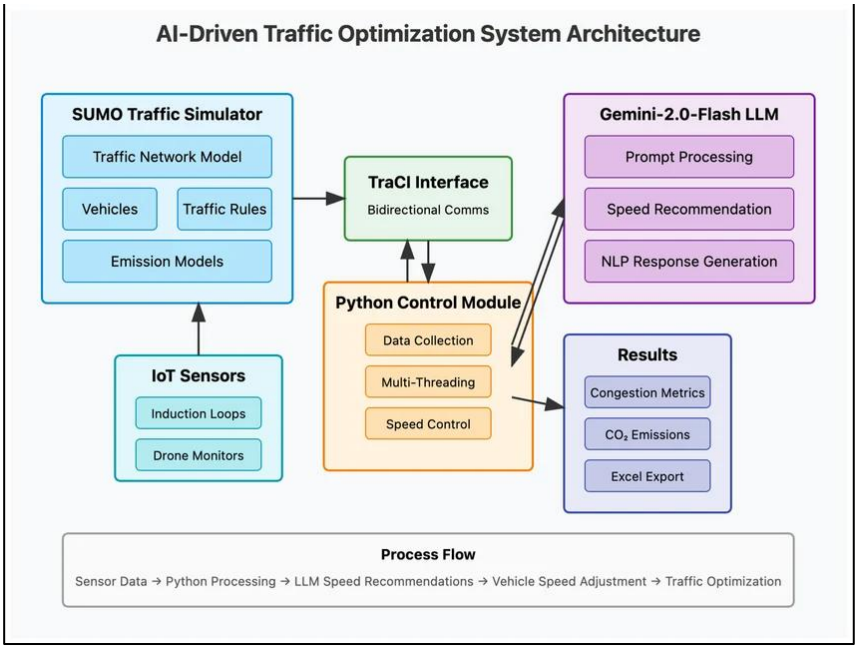

Figure 5: System architecture of the AI-driven traffic optimization framework

(Source: https://www.researchgate.net)

The reviewed research has shown that AI has been proven to be a valuable enabler to monitor operations in distributed federated cloud systems before 2022. With the help of AI algorithms, real-time system logs and performance metrics could be used to detect anomalies in hours rather than seconds, further minimizing the average time to detect them [9]. Real-time analytics supported operators to see how their infrastructure was doing at a given time, and predictive allows them to foresee infrastructure issues prior to disruption. Since then, the infrastructure performance was improved, the latency time decreased and the optimization of resources allocation was observed. Hand-on monitoring activities started to decline significantly when automated monitoring and self-healing tools started to take over the traditional hand-on approach to monitoring for federal cloud deployments.

4.2 Benefits of Lineage Automation

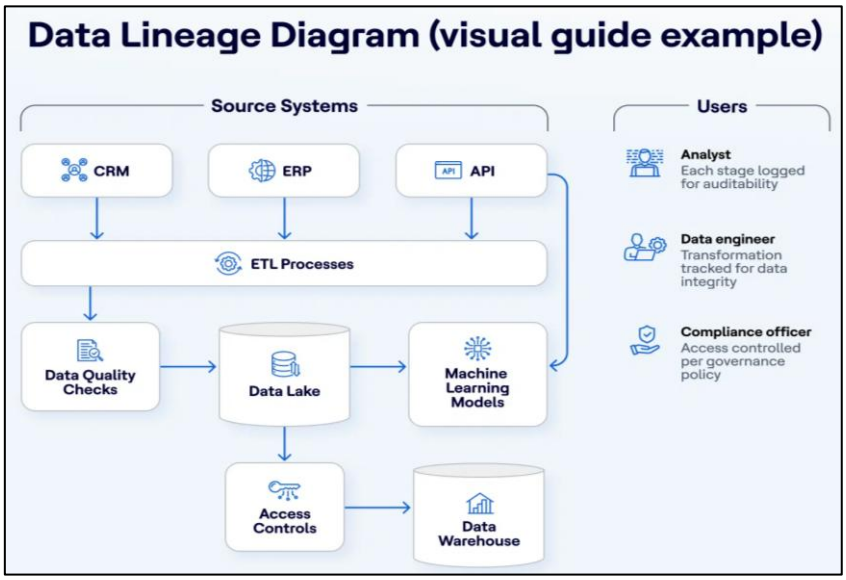

Figure 6: Data Lineage Diagram

(Source: https://www.researchgate.net)

Lineage automation brought about a tremendous governance improvement across the federal cloud. The improved capability to trace data helped agencies to follow the movement of data and all its transformations through distributed nodes to create complete provenance chains. Once the automated lineage-log capability was adopted in the compliance-reporting system, compliance reporting subsequently improved because auditors' requirements were met without having to manually compile the logs. A compliance visualization tool for on-demand audits was used to showcase compliance for both FedRAMP and NIST compliance frameworks to improve audit readiness [10]. Moreover, decentralized spreadsheets and manual record management were eliminated, further lightening the governing burden, whereas the compliance officer can now concentrate on dealing with exceptions in policies, instead of routine tasks involved in checking the lineage.

4.3 Security and Governance Improvements

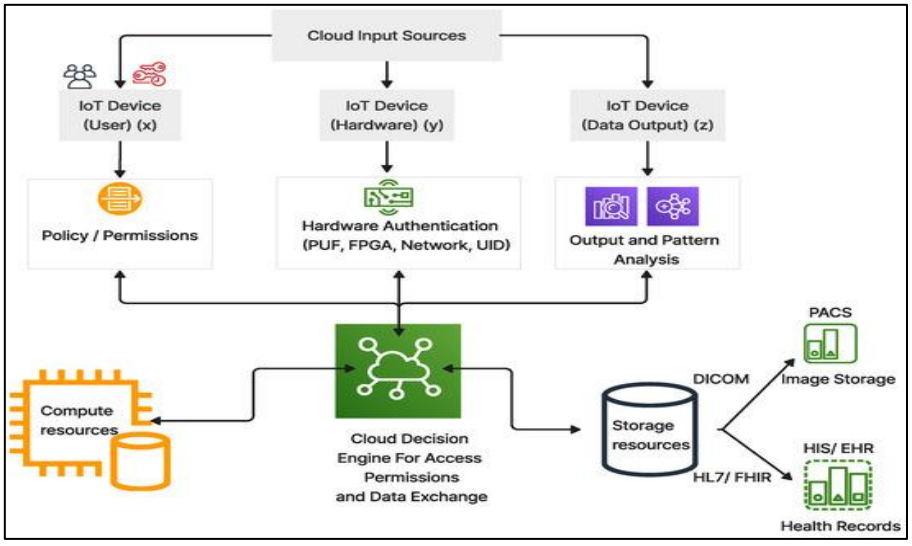

Figure 7: Security and Governance Improvements

(Source: https://www.mdpi.com)

One of the most prominent security features was the real-time identification of unusual access patterns and possible data exfiltration attempts by AI powered threat monitoring, where machine learning algorithms monitored the activities on the website.AI-powered threat monitoring proved to be an essential addition to security, where machine-learning models tracked the activity on the website in real time and alerted to abnormal patterns that could indicate data exfiltration or unauthorized access. Continued detection of system gaps in vulnerabilities by scoring and automated prioritization of remediation actions led to better risk management [11]. Enforcement of improved policies was enabled by AI-driven rules engines which enforced federal data governance policies on hybrid cloud boundaries, consistently. All these features combined shortened the time required for response to security incidents and gave zero-trust architectures a boost, but complete automated compliance with policies was a major hurdle till 2022.

4.4 Key Challenges Identified

However, there were indeed a number of difficulties that remained. Federal data silos hindered the functioning of an operational intelligence model that could be unified to develop AI models. The complexities in meeting with legacy federal systems drove a lot of customization and extended deployments. Innovative programmatic solutions slowed and stressed agency budgets, largely due to the resource and infrastructure challenges of implementing artificial intelligence and systems. A shortage of AI, cloud engineering, and data lineage automation workers put obstacles in the way of adoption. The potential for algorithm bias and absence of transparency or understanding of AI generated actions were ethical concerns, making governance issues pertinent especially with compliance-oriented federal workloads that justify audit to ensure decisions are traceable.

5. Discussion

5.1 Interpretation of Findings

The results highlight that AI completely transformed operational management in cloud systems of federal organisations, moving them from reactive to predictive monitoring mode. One of the elements that must have been important was intelligent automation, which meant that networks were automated and had no person to track manual lineage, nor were they required to be self-healing [12]. Governance and AI became clearly linked as AI was both used to automatically implement policies and create easy-to-audit evidence of actual policy compliance. Despite having technical capabilities, however, governance systems did not keep pace prior to 2022. Finally, AI became an operational enabler and a governance accelerator, albeit in an incomplete way; with the absence of federal standards on AI.

5.2 Comparison with Previous Research

Several common trends emerged across the reviewed studies, emphasizing the significant benefits of AI, including quicker anomaly detection and fewer manual workloads. But the strategies used were found to differ – some agencies placed their AI monitoring in the central location, others in the edge. Prior to 2022, data indicated that cloud providers had more interoperability issues and that users continued to increase their use of open-source AI technologies [13]. Of note also, commercial cloud studies reported less integration problems than government-specific research – reflecting the challenges facing governments that are different. In general, the findings from the previous studies indicated a need for one set of standards.

5.3 Implications for Federal Agencies

Skalable governance of AI that keeps pace with the increased ‘cloud use’ needs to be a priority for Federal Agencies to address without compromising audit integrity. Lineage automation is essential; it directly saves audit preparation time and human error, and is no longer an option for automation. What makes strategic use of operational intelligence so special, is its ability to convert telemetry data into actionable data, which then allows proactive risk management. Agencies need to invest in high-quality AI systems that are compatible and train staffers. These steps will otherwise further widen the technical-savvy-gov-prepared divide that continues to be a digital transformation inhibitor.

6. Conclusion

According to this research, the benefits of enhanced distributed federal cloud systems before 2022 were large, driven by AI's operating (OI) and lineage automation capabilities. AI was found to facilitate quicker anomaly detection, real-time analytics, and

less manual effort, and lineage automation facilitated better traceability of the data, audit readiness, and compliance reporting. For making the shift from reactive to predictive operations, AI-driven intelligence systems were invaluable. The value to lineage automation also had a huge impact as it provides automatic governance and creates data lineage everywhere while in hybrid environments. Reflecting this, the Government's overall performance in cloud transformation pre-2022 was positive on efficiency and secure use, but limited by complexity in integration, lack of skilled individuals and an unintegrated governance framework. This requires the adoption of an AI strategy and seamlessly interoperable Automation standards for future success.

7. Recommendations

7.1 Technical Recommendations

Cloud-based federal systems need more and better transparency of AI, which means that there is a need for an explainable model that validates the automated decision-making processes. To ensure that lineage is tracked across multi-cloud and hybrid setsups, with no more proprietary silos, better interoperability standards need to be established. The role of AI in threat intelligence and in Zero Trust architectures should be inherently connected, with AI playing a pivotal part in real-time policy adjustments [14]. Moreover, cloud providers would also enjoy standardized APIs for exchanging metadata with other providers, thereby enabling easy lineage automation. These technological breakthroughs lay the groundwork for trustworthy and scalable AI operations in the federal environment.

7.2 Organisational Recommendations

To succeed in their efforts, federal agencies need to invest in workforce training to develop lineage knowledge and proficiency in AI. Robust AI governance policies should be established which outline how different stakeholders are responsible, bias mitigation mechanisms, and auditing processes of automated decision-making processes. Continuous investment in the automation infrastructure is crucial, covering metadata management platforms and AI monitoring tools. To succeed in adopting cultural change, leadership engagement in moving from manual to automated compliance workflows is key. If there is no readiness in the organization, technically solutions cannot render the desired effect.

7.3 Future Research Recommendations

Moving forward, the study of per-2022 trends in Generative AI and Large Language Models for automated lineage documentation and compliance reporting would be beneficial. The mix of biases and their detection methods for federal cloud workloads needs to be explored in an ethical governance. For advanced autonomous cloud management such as fully self-healing infrastructures and policy-driven automation, it is essential to validate these through the lens of real data [15]. The effects of implementing the plan would show up over time, in a longitudinal study that looked at the plan before and after it was put into place in 2022. Lastly, resource-limited federal agencies would be able to plan strategically with the help of the cost-benefit analysis of investments in AI automation.

Reference List

[1] Amina, E.S., Rohan, I. and Thomas, R., 2021. Designing Federated Compliance Data Platforms: Leveraging Multi-Region Snowflake Warehousing and Distributed Governance Frameworks for Global BFSI Risk Analytics. International Journal of Trend in Scientific Research and Development, 6(1), pp.1988-2000.

[2] Bega, D., Gramaglia, M., Perez, R., Fiore, M., Banchs, A. and Costa-Pérez, X., 2020. AI-based autonomous control, management, and orchestration in 5G: From standards to algorithms. IEEE Network, 34(6), pp.14-20.

[3] Gaffar, O., Sikiru, A.O., Otunba, M. and Adenuga, A.A., 2020. Cloud-Native Data Lake Architectures for Advanced Financial Modelling and Compliance Analytics. Journal of Frontiers in Multidisciplinary Research, 1(1), pp.145-155.

[4] Goethals, T., Volckaert, B. and De Turck, F., 2021. Enabling and leveraging AI in the intelligent edge: A review of current trends and future directions. IEEE Open Journal of the Communications Society, 2, pp.2311-2341.

[5] Guntupalli, B., 2021. The Evolution of ETL: From Informatica to Modern Cloud Tools. International Journal of AI, BigData, Computational and Management Studies, 2(2), pp.66-75.

[6] Khan, A., Bilal, E.M., Rodríguez, E. and Chen, D., 2021. The Intersection of AI, Big Data, and Cloud Computing in Modern ICT Solutions. International

Journal of Information and Communication Technology Trends, 1(1), pp.127-135.

[7] Kothandapani, H.P., 2021. Integrating robotic process automation and machine learning in data lakes for automated model deployment, retraining, and data-driven decision making. Sage Science Review of Applied Machine Learning, 4(2), pp.16-30.

[8] Machireddy, J.R., Rachakatla, S.K. and Ravichandran, P., 2021. AI-Driven business analytics for financial forecasting: Integrating data warehousing with predictive models. Journal of Machine Learning in Pharmaceutical Research, 1(2), pp.1-24.

[9] Nyberg, A.O.H., 2021. Transformer-Augmented AI Framework for ERP-Integrated Cloud Security Multi-Factor Authentication, Multivariate Classification, and Real-Time Threat Detection. International Journal of Research Publications in Engineering, Technology and Management (IJRPETM), 4(5), pp.5588-5594.

[10] Padur, S.K.R., 2020. AI augmented disaster recovery simulations: From chaos engineering to autonomous resilience orchestration. International Journal of Scientific Research in Science, Engineering and Technology, 7(6), pp.367-378.

[11] Shaffi, S.M., 2020. Intelligent emergency response architecture: A cloud-native, ai-driven framework for real-time public safety decision support. The Artificial Intelligence Journal, 1(1).

[12] Sharma, A., 2019. A Multi-Layered Framework for Secure Distributed Computing in Heterogeneous Cloud–Edge Environments Using Adaptive AI Orchestration. American International Journal of Computer Science and Technology, 1(3), pp.1-11.

[13] Singh, B., 2017. Enhancing Real-Time Database Security Monitoring Capabilities Using Artificial Intelligence. International Journal of Current Engineering and Scientific Research (IJCESR).

[14] Subramanya, T. and Riggio, R., 2021. Centralized and federated learning for predictive VNF autoscaling in multi-domain 5G networks and beyond. IEEE Transactions on Network and Service Management, 18(1), pp.63-78.

[15] Thota, M.R., 2020. AI-Augmented Database Administration: From Reactive Operations to Predictive, Self-Optimizing Data Ecosystems. European Journal of Advances in Engineering and Technology.