A STRUCTURED APPROACH TO LEGACY DATABASE TRANSFORMATION IN HYBRID CLOUD ENVIRONMENTS

N. SIVA KUMAR
Research scholar, Department of Computer Science & Engineering, University of Technology, Vatika, Jaipur, E-mail: Siva733.uot@gmail.com

Professor SUNEEL PAPPALA
Department of Computer Science & Engineering, University of Technology, Vatika, Jaipur, E-mail: suneelpappala@gmail.com

ABSTRACT

A comprehensive review of literature suggests seven levels of complexity in transforming legacy systems, including, being a stand-alone system, being part of a larger system, and data incompatibility, each depicting unique criteria. Using a mixed-method approach, we surveyed and interviewed stakeholders across five Saudi Arabian universities. These systems were built for a different era, lacking the agility and scalability required to support modern business demands. Most of the new systems and applications are built to be cloud native applications but many large organizations still depend on legacy systems that have been built tens of years ago. Nonetheless, very little is known as to what degree these complexities implicate the implementation of digital transformation efforts in public administration (PA). Legacy systems remain critical to the operations of many organizations but often pose significant challenges in a fast-paced, technology-driven landscape. In addition, a comprehensive understanding of the systems to be transformed, the policies which they are serving, and the broader PA setting in which they are implemented were deemed central to succeeding in digital transformation efforts. Cloud modernization-the process of migrating these systems to the cloud-has emerged as a key solution. By incorporating AI, organizations can enhance the effectiveness of cloud modernization initiatives. AI brings automation, advanced analytics, and real-time decision-making capabilities that streamline migration and improve operational outcomes.

Keywords: *mixed-method approach, public administration (PA), transforming legacy systems, technology-driven landscape, real-time decision-making.*

INTRODUCTION

Data virtualization has existed for many years but only recently have new capabilities come online that enable companies to leverage disparate legacy and modern data across the hybrid cloud, bringing it together for BI teams and the greater business. Specifically, as part of an overall adaptive analytics fabric (the virtualized data and associated tools to aid analytics speed, accuracy, and ease of use), virtualization empowers companies to treat all their disparate data repositories as a single, unified data source that's extensible to support future technologies. A fabric provides a bridge across data warehouses, data marts, and data lakes, delivering a single view of an organization's data without having to physically integrate, engineer, or re-architect it. This abstraction enables enterprises to instantly surface usable data, no matter where it's actually stored, to produce fast, timely insights. The ability to merge data from different sources reveals another advantage. Rather than combining data into a single system that necessitates formatting data for the lowest common denominator of capability, adaptive analytics fabrics enable enterprises to store data in the data structures that best fit its use (e.g., in legacy systems). For example, time-series data can be stored one way and relational data can be stored another way, enabling companies to use the specialized analytics formats for

that data. Data can live in the format best suited for its utility while the analytics fabric adapts to business or operational analytics users by translating and presenting the data as needed. With regard to querying necessary legacy systems, autonomous data engineering, as part of an adaptive analytics fabric, shortens these query times from days (on large data sets) to hours or minutes. As queries are run against data sets in the analytics fabric, machine learning is applied to determine what data within the larger set is needed, bypassing extraneous data altogether during the query process.

LITERATURE REVIEW

Chelliah Paramasivan (2024) The banking sector is undergoing a digital revolution that is drastically changing how banks run, interact with clients, and handle internal procedures. But the advent of digital technologies has forced a change in direction toward more adaptable, scalable, and economical solutions. Banks are currently utilizing cloud technologies more and more to accomplish their many goals and to establish an adaptable and agile banking environment that can react rapidly to changing business requirements. The main purpose of this paper is to examine how cloud computing is conceptualized through digital transformation, and how it performs smooth adoption of cloud computing in the banking industry. With the help of previous research studies, models of cloud computing, cloud operating models, best practices, and challenges from the viewpoint of the customer pointed out clearly.

Gabriel Mafura (2023) Currently, Kenyan research institutions Information Technology operations use external storage, within or without institutional network environments. This study presents a structured literature review of a cloud computing adoption strategy for research institutions in Kenya. The reviews objectives are: 1) to establish the characteristics/peculiarities and IT environments of research institutions in Kenya, 2) current cloud computing technologies, adoption approaches and drivers in the adoption of cloud computing. The findings are that there is need for an adoption strategy for Cloud computing in the Kenyan research Institutions that will use a DOI and TOE combined approach while addressing technological, organizational and environmental factors. The findings are used to propose a cloud computing adoption strategy for the Kenyan Research Institutions in future research.

Ramakrishnan Manjula (2023) In recent years, technological advancements have provided the world with cloud computing which can transfer, store, and process huge data chunks in the form of video, audio, images, and text efficiently. In spite of the universal hype on the subject across the information technology world, protecting sensitive data stored in the cloud server is one of the crucial problems. The large volume and sophistication of cyberattacks conclude to the fact that private pictures need exceptional care than other forms of data on the cloud. Since the user who has stored their private pictures in the cloud has no control over the privacy protection of data, the cloud vendors have to assure a greater level of security in terms of authentication and prevention from cyberattacks. This work aims to develop a method for enhancing the security of user photographs on a cloud platform by means of cryptography algorithms.

Durga Prasad Sharma (2020) In the modern arena, the Information and Communication Technologies (ICTs) have been playing a vital role in every walk of our day-to-day life. In order to enhance the ICT capacity and align with the global technology transformations, the developing countries have started introducing the computerization and automation processes at different levels of the governments. The several research studies revealed that the existing legacy of governance system and their services in current state have several issues and challenges in terms of timeliness, cost of services, delay in service delivery, time-bound availability of services (24/7/365), inefficiency services, ease of service and discomforts, poor service collaboration, absence of responsiveness, and limited security of sensitive information/documents. A significant question is still unanswered that how to bring the Citizens and Government bodies closer for alleviating the aforementioned issues and challenges of existing government system services. This research paper aims to investigate the issues and challenges in the current status of Governance and partial E-Governance systems which encompass the computerization or automation process.

Al Qadami, S. (2018) China has entered the stage of excess supply, which provides a material basis for sharing economy. In order to solve the backward management information of the existing catering enterprises, the idle and wasteful problems of human technology, restaurant spaces and food resources, this paper studies and develops a shared restaurant platform based on cloud computing, aiming at overcoming the shortcomings of existing restaurant management, and further integrating and optimizing the food and beverage supply chain resources. Integrating the idea of "sharing economy", this paper proposes a shared restaurant business model, and uses cloud computing to integrate and share software and hardware resources and information to build a shared restaurant platform, through designing three different subsystems of the client, restaurants management background, and platform management background to complete the phased management of shared restaurant platform.

Hybrid Cloud

The public cloud might consist of at least two or more distinct cloud architecture (private, community, or public) that remain different entities but were bound together by institutionalised or rigid entrepreneurship that licences data and processing pocket ability.

Role of IDS in Cloud Environment

Intrusion Detection System (IDS) is a type of security software that can monitor, mitigate, and perhaps respond to system threats. IDSs have shown to be successful techniques in traditional distributed networks in the past. An IDS creates vulnerability analysis warnings and archives them for later analysis in a common internet setting. The system administrator can then determine whether or not to act on the IDS decision and prohibit user actions. Many firms that use cloud systems have an autonomous procedure to track numerous events that occur on their network assets. As a result, IDS can offer the necessary security against foreign invaders and internal users abusing their authentication systems. As a result, including IDS into every network is critical. The placement of the IDS is crucial to enable effective event detection. The IDS, along with other critical security software such as the admission wiring harness and virus protection, can be deployed directly in behind stateful firewall in a conventional network topology.

Scalability and Performance in Hybrid Environments

A well-designed hybrid architecture must be elastic, efficient, and resilient. The key is correctly distributing workloads and taking advantage of the best of each environment.

Load Balancing between Cloud and On-Premise

Implementing intelligent load balancing mechanisms allows traffic to be distributed between on-premises systems and the cloud. This ensures availability, reduces latency, and maximizes resource usage. Load balancers, application gateways, or specific hybrid cloud solutions like Azure Front Door, AWS Global Accelerator, or GCP Cloud Load Balancing can be used.

Use of Containers and Orchestrators (e.g., Kubernetes) Containers allow packaging applications in a portable and lightweight way. By deploying workloads on Kubernetes, automatic scalability, failure recovery, and uniform distribution between on-premises and cloud environments can be achieved (thanks to solutions like Red Hat OpenShift, SUSE Rancher, Azure Arc, or GKE On-Prem). This also facilitates the adoption of DevOps and continuous integration/continuous deployment (CI/CD).

Latency and Throughput Optimization

Latency is critical in hybrid integrations. Some recommendations include:

- Placing services close to consumers (edge computing).
- Minimizing network hops through flat architectures.
- Compressing payloads and reducing unnecessary calls.

Additionally, optimizing throughput requires proper resource sizing, middleware tuning, and efficient use of communication protocols.

Overview of legacy system modernization

Legacy system modernization has emerged as a critical priority for organizations aiming to remain competitive in the rapidly evolving digital landscape. Although once considered state-of-the-art, many enterprises rely on outdated infrastructure,

which now imposes significant operational and financial burdens. These systems are often monolithic, built on outdated programming languages, and unable to integrate seamlessly with newer technologies. As businesses scale and evolve, the limitations of such infrastructure become more apparent, leading to reduced efficiency, increased security risks, and rising maintenance costs. One of the primary challenges associated with legacy systems is their inability to support modern business needs. As data volumes grow exponentially and real-time processing becomes a necessity, traditional architectures struggle to keep up with the demands of contemporary applications. Furthermore, these systems often lack interoperability with cloud environments and emerging technologies such as artificial intelligence and machine learning, limiting an organization's ability to leverage data-driven insights. Enterprises often find themselves locked into vendor-specific solutions with high licensing costs and limited flexibility, further exacerbating modernization challenges. Security vulnerabilities pose another critical concern for organizations relying on legacy systems. Many of these infrastructures were built before modern cybersecurity threats became widespread, leaving them susceptible to data breaches, ransomware attacks, and compliance violations. As regulatory requirements evolve, businesses must ensure that their IT environments meet industry standards, such as data protection laws and governance frameworks. Failing to modernize can result in non-compliance, leading to reputational damage and financial penalties.

Limitations of legacy infrastructure

Once considered cutting-edge, legacy systems now represent significant technological liabilities for organizations striving to remain competitive in a digital-first economy. These outdated infrastructures create multiple challenges, particularly in the areas of scalability, security, and operational efficiency. One of the most critical concerns is the accumulation of technical debt, which arises when older systems are continuously patched and modified to accommodate new business needs without a strategic overhaul. This incremental approach often leads to a complex and fragile IT environment that becomes increasingly difficult to maintain and upgrade. Technical debt manifests in various forms, including outdated programming languages, inefficient database structures, and rigid system architectures. Many legacy systems were designed using monolithic frameworks lacking modularity, making implementing updates without disrupting core functionalities challenging. As a result, organizations experience slow software development cycles, limited agility in responding to market demands, and high maintenance costs. The inability to integrate modern development practices, such as continuous deployment and automated testing, further exacerbates inefficiencies, forcing IT teams to devote substantial resources to maintaining obsolete technologies.

RESEARCH METHODOLOGY

This method is more effective, and it may be used to anticipate incursion. The duplicate data is deleted and the supervised learning is delivered. Fuzzy modelling is used to quantify correntropy amongst attributes for unsupervised learning. This second research work focused on

developing an intrusion detection system, which is critical for creating computer security and anticipating harmful activity. For assessing the variable's interconnections and featured acquisition to eliminate outliers over the collection, the Reference model and auto-encoder are utilised. This model improves accuracy rate by predicting malicious activity or aberrations that are approaching the cloud. The event association and learn how to navigate among the endpoints in a cloud infrastructure are determined by related to collective, which specifies inter-dependencies. The transfer possibility is determined by the network model, which itself is made up of directed graph and components. Data protection is one of most prominent problems in cloud computing. As a result, in this third research work, an appropriate intrusion detection approach was proposed. The events are determined using mathematical techniques in order to achieve a fair categorization. The AN's teaching process is comprised of teaching and developing the attachment description of the information in order to lower the model complexity in between recovered information and the input variables.

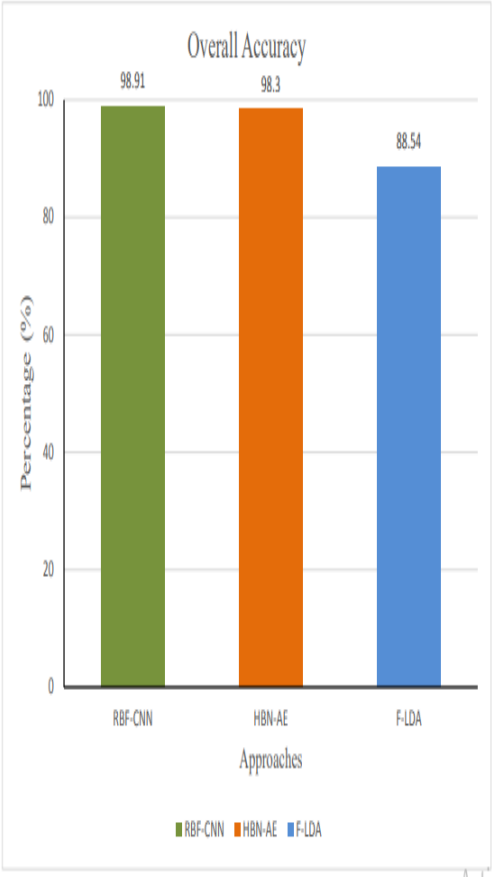
RESULTS AND DISCUSSIONS

The behaviour rule constructed in this manner is too dependent on a particular variable. This strategy loses the precision and breadth of information given by aspects. As a result, F-LDA, with the exception of RBF-CNN, are unable to identify as many assaults as the proposed model can identify. The proposed model has a limitation in that it identifies fewer probing assaults like port sweep, queso, and ipsweep. The reason for this is that proposed model is host-based, whereas other are infrastructure, which has the advantage of being able to identify probing assaults.

Table 1: Comparative result on Accuracy

	RBF-CNN	HBN-AE	F-LDA
	Accuracy (Optimal model)	Accuracy	Accuracy
Over All Accuracy	98.91	98.30	88.54

This issue fixed by increasing the weight of specific characteristics, such as port number and TCP flag. Table 1. Shows that the proposed model RBF-CNN has accuracy of 98.91%.



Graph 1: Results of Proposed Works

The comparison of our suggested approach RBF with the CNN combined rule had

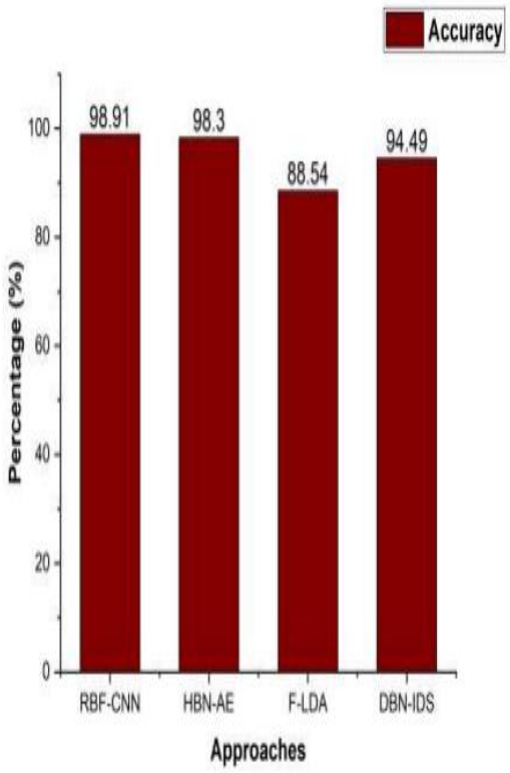
improved prediction data. Since the HBN-AE classification learner and the F-LDA classifications learning both anticipate that certain sequences are aberrant. During such period, some standard packets are blocked by the VM. Table 2 demonstrates the overall accuracy of HBN-AE, F-LDA, and RBF-CNN when compared to other methods, while graph 1 shows the average precision of HBN-AE, F-LDA, and RBF-CNN.

For intrusion prediction, the proposed attention-based RBF-CNN is compared to state-of-the-art approaches like Hybrid Bayesian Network with Auto Encoder (HBN-AE), Fuzzy based Latent Dirichlet Allocation (F-LDA), and Deep Belief Network (DBN) IDS.

Table 2: Overall Comparison

Methods	Accuracy	Sensitivity	Specificity	FAR
RBF-CNN	98.91	98.13	98.99	3.89
HBN-AE	98.30	96.72	97.08	4.11
F-LDA	88.54	94.8	95.00	4.31
DBN-IDS	94.49	93.99	94.32	5.01

The proposed technique exceeds all other current approaches in terms of FAR, range, inaccuracy, and responsiveness. As a consequence, Table 2 examines the effectiveness of current intrusion detection approaches.



Graph 2: Comparison of accuracy

CONCLUSIONS

The research large datasets were all from the KDD CUP 99 data set, which includes intrusion and nonintrusion datasets. The proactive security methodology was implemented using JAVA software. It has a more significant influence on the long-term growth of cloud computing. Modeling effective IDS for cloud environments is a significant approach for securing the cloud from susceptible and malicious assaults. Several researchers have recently demonstrated that proper and effective modelling of IDS for the cloud may be built. A hybrid Bayesian network and an auto-encoder can be used to analyse variable interconnectedness and classification techniques in order to predict attacks on nodes. The suggested RBF-CNN intrusion detection model was compared to state-of-the-art techniques. When compared to several current detection equipment, the suggested system will

increase intruder detection performance. The proposed model was additionally distinguished by a low false alarm rate and good precision. The proposed model was put to the test on a typical dataset, and the findings demonstrate that the process was suitable at identifying network intrusions. The proposed model can create networking predetermined guidelines immediately and then safeguard internet traffic. It is appropriate for usage in networks with diverse data formats and sources, and therefore it provides a means of cybersecurity. Without any annotated information, learning algorithms can "learn" the network's normal pattern and identify irregularities. It can identify novel sorts of intrusions, although false positive alarms are common.

REFERENCES

1. *Evgeny V. Nikulchev (2013), "Use of Dynamical Systems Modeling to Hybrid Cloud Database", International Journal of Communications, Network and System Sciences, issn:1913-3723, Vol. 6, No. 12, Pages. 505-512.*
2. *Gabriel Mafura (2023), "Adoption Strategy for Cloud Computing in Research Institutions: A Structured Literature Review", Journal of Computer and Communications, issn: 2327-5227, vol.11, pages.63-78.*
3. *Chelliah Paramasivan (2024), "Navigating Digital Transformation in Banking with Cloud Computing Solutions", Open Journal of Business and Management, issn:2329-3292, vol. 12, pages. 4227-4253.*
4. *Frank Maurer (2012), "Emerging Issues & Challenges in Cloud Computing—A Hybrid Approach", Journal of Software Engineering and Applications, issn:1945-3124, vol. 5, pages. 923-937.*
5. *Durga Prasad Sharma (2020), "Cloud-Enabled E-Governance Framework for Citizen Centric Services", Journal of Computer and Communications, issn: 2327-5227, vol.8, pages. 63-78.*
6. *Ramakrishnan Manjula (2023), "Implementation of a Hybrid Triple-Data Encryption Standard and Blowfish Algorithms for Enhancing Image Security in Cloud Environment", Journal of Computer and Communications, issn: 2327-5227, vol. 11, pages. 135-149.*
7. *Al Qadami, S. (2018), "Research and Development of Shared Restaurant Platform Based on Cloud Computing", American Journal of Industrial and Business Management, issn: 2164-5175, vol. 8, pages. 2321-2333.*
8. *Deb M, Choudhury A. (2021), "Hybrid cloud: A new paradigm in cloud computing", Machine Learning Techniques and Analytics for Cloud Security, issn: 2959-6386, pages. 1–23.*
9. *Gade KR (2021), "Migrations: Cloud migration strategies, data migration challenges, and legacy system modernization", Journal of Computing and Information Technology, issn: 1846-3908, vol. 1(1).*
10. *Rapanotti, L. (2020), "Requirements analysis gamification in legacy system replacement projects", Requirements Eng., issn: 1432-010X, vol. 25, pages. 131–151. https://doi.org/10.1007/s00766-019-00311-2*