International Journal Science and Technology
IJST Vol 2 No. 1 March 2023 | ISSN: 2828-7223 (print), ISSN: 2828-7045 (online) Page 80-94

Architectural Patterns for Immutable Financial Data Lakes on AWS S3 for Regulatory Compliance and Real-Time Analytics

Gunjan Kumar
Independent Researcher, gunjankumar.email@gmail.com, India

Article History
Received : January 05, 2023
Revised : January 27, 2023
Accepted : February 08, 2023
Published : February 27, 2023

Cite This Article:
Gunjan Kumar. (2023). Architectural Patterns for Immutable Financial Data Lakes on AWS S3 for Regulatory Compliance and Real-Time Analytics. *International Journal Science and Technology*, *2*(1), 80–94.

DOI:
https://doi.org/10.56127/ijst.v2i1.2316

Abstract: Financial services business is becoming highly reliant on the safe and dependable storage of large quantities of structured and unstructured information. As the global regulatory demands become increasingly strict, including the SEC Rule 17a-4, Sarbanes-Oxley Act (SOX), and the General Data Protection Regulation (GDPR), organizations now have to contend with the twin demands of maintaining compliance and ensuring immutability with the actions of being analytics agile in making decisions. In this paper, the design of an immutable financial data lake is explored based on Amazon Web Services (AWS) Simple Storage Service (S3) and a set of Amazon Web Services-native tools. In particular, it examines how using S3 versioning and object lock, and write-once-read-many (WORM) policies can be utilized together with AWS Glue, Amazon Athena, Amazon Redshift and Amazon Kinesis to create a secure, compliant and high-performing architecture. Through the implementation of architectural designs based on cloud-native principles, the study illustrates how financial institutions can be able to meet high auditability standards and at the same time, provide the benefit of real-time analytics to detect fraud and manage risk and conduct market research. In addition, it provides the governance, lineage and security structures that are required to maintain integrity and promote innovation. The results underscore the fact that data lakes of immutable data stored in clouds, properly designed, will level compliance assurances, as well as a stable base of a contemporary financial intelligence.

Keywords: Unchanging Data Lake; Financial Data; S3 at AWS; Compliance with Regulations; Real-Time analytics; Object Lock; Data Governance; Fraud Detection; Cloud native Architecture; Streaming Data.

INTRODUCTION

Background and Motivation

One of the most data intensive organizations in the world is the financial institution, which deals with data on transactions, risk exposures, customer portfolio, and compliance report as well as world market operations. As the financial data becomes digitalized, and algorithmic trading, and real-time settlement systems emerge, the volume and complexity of financial data has increased exponentially. One organization can handle the orders of millions of trades every day, and at the same time, track the sales of customers and fluctuations in the market in various geographies (Kansara, 2021). The activities generate petabyte-scale volumes of data that have to be handled, stored, and analyzed with high confidence in terms of integrity, traceability, and availability (Mariana, Rakesh and Thomas, 2023).

Regulatory scrutiny has also been increased due to the explosion of financial technologies (FinTech). The General Data Protection Regulation (GDPR), the Dodd-Frank Act, and Basel III, are just some of the frameworks that mandate institutions to have tamper-proof records and have audit trails that are transparent.

The non-adherence leads to reputational losses to organizations, as well as huge financial fines. The conventional on-premises databases and data warehouses, though resilient in particular areas, are unprepared to deal with the two-fold requirement of immutability and low-latency analytics, particularly in globally dispersed financial ecosystems (Mohna et al., 2022).

Cloud-based systems have in this regard become a paradigm shift in data management. Amazon Web Services (AWS) and especially Amazon Simple Storage Service (S3) offer a storage durability of 99.999999999% S3 has thus become the backbone of the factual data lakes of enterprises. However, its primitives guarantee the integrity of data, but a complete immutable data lake means that it needs an architecture-wide tie to other AWS services to perform ingestion, governance, security, and analytics (Minichino, 2023). In the absence of such integration, organizations will not be able to strike the balance between compliance and the agility needed by real-time analytics.

In this way, it can be said that the reason behind conducting this study is to fill this gap. Combining fixed storage systems (using immutable storage) with high-performance analytics processes, this research offers a roadmap of how financial institutions can both appease regulators and offer high-level analytical use cases (fraud detection, risk modeling, and market surveillance).

Problem Statement

Despite the scalable and safe infrastructure provided by the cloud, financial institutions still have one of the main problems in maintaining the data immutability and providing the high analytical performance. Most deployments of cloud data lakes prioritize storage and cost-effectiveness but do not have a solid immutability policy. To take the example, the data stored can be manipulated without the appropriate implementation of WORM policies or Object Lock, which poses a risk to auditability and readiness to comply (Adebowale and Akinnagbe, 2023). On the other hand, compliance-oriented systems tend to fail to support the real-time analytics demands of the contemporary financial processes, including fraud detection when transacting or algorithmic decision-making in trading settings (Mohna et al., 2022).

This dilemma leads to a trade off in that either platform is very compliant and thus analytically unyielding, or they are analytically flexible and not possessing verifiable immutability. In addition, the governance models tend to be piecemeal, it is hard to have uniform implementation of data lineage, retention policies, and data access control to various services. Consequently, financial organizations are exposed to non-compliance by regulators as well as operational inefficiencies and hence; there is a dire need to have an architecture that balances compliance integrity and real-time analytical efficiency.

Objectives of the Study

The main goal of the study is to come up with and test architectural concepts of the immutable financial data lakes on AWS S3, that comply with the regulatory imperative and analytical imperative. The following specific goals are undertaken by the study in order to accomplish them:

Demonstrate immutability enforcement: It is necessary to demonstrate how AWS S3 can be used to create a secure, immutable platform to store financial data according to the global compliance standards, such as versioning, retention configurations, and Object Lock systems (Irani, 2023; Joseph, 2023).

Combine complementary services of AWS: Demonstrate how AWS Glue can be used to manage metadata, Amazon Athena can be used to run queries without the need to maintain servers, Amazon Redshift can be used to relocate data to a high-performance warehouse, and Amazon Kinesis can be used to receive and stream data in real-time (Minichino, 2023; Rangarajan and Bounds, 2023).

Put in place governance and lineage cycles: Suggest mechanisms of providing a secure pipeline system, visible data lineage and automated governance processes that enable auditability and accountability (Marosi et al., 2022; Vinnikainen, 2023).

Show real-life implementation: Show where such an architecture can be used in solving urgent financial issues, such as detection of frauds, evaluation of credit risks, and analysis of the market in real-time (Adebowale and Akinnagbe, 2023).

Contribution

The paper will benefit the field of academic research and practice in the industry by providing a systematic way of creating immutable and cloud-native financial data architecture designs. In terms of research, it closes the knowledge gap in the literature where compliance-oriented immutability and performance-oriented analytics are frequently considered independently, instead of being viewed as mutually complementary requirements (Marosi et al., 2022; Vinnikainen, 2023). The synthesis of information sources (case studies, AWS technical documentation, and current literature in data governance) will allow the study to propose an integrated architectural blueprint that will address these two competing needs.

Practically, the paper offers practical trends to be implemented by cloud architects, compliance officers, and data engineers in the financial business. It shows how irrevocable S3 based data lakes can be used as one source of truth to all financial information, enhancing regulatory reporting and at the same time driving sophisticated analytics. Such contributions are especially important in an environment where financial regulators continue to insist on increasingly high degrees of transparency, and markets insist that institutions be fast and smart (Joseph, 2023).

Through compliance and innovation reconciliation, the research provides a future-proof architecture strategy that enables financial institutions to address the future changing regulatory requirements without compromising the competitive benefits of real-time analytics.

LITERATURE STUDY

The financial data architecture literature indicates the increased conflict between the imperative to have immutable storage and the need to have real-time analytics in regulated sectors. In the past, the data lakes were constructed as versatile repositories, which allowed organizations to store raw data without having a strict governance over those data. But the financial industry, which is highly regulated, requires that such architectures be tailored in order to provide immutability, security, and compliance as well as facilitate advanced analytical services (Kansara, 2021; Mariana, Rakesh and Thomas, 2023).

Data Lakes Traditions and Financial Industry Requirements

Conventional data lakes could be applied as centralized locations to store structured and unstructured information, which were mostly designed to support batch analytics (Banerjee, 2022). Although such systems could be scaled, they did not have the fine-grained governance capabilities they needed to comply with regulations. Moreover, their usage of schema-on-read models posed a problem of consistent audit trail and guaranteeing data integrity (Bhaskaran, 2020).

The fact that conventional data lakes did not ensure immutability made them unsuitable in terms of regulatory purposes in the financial services industry. Laws like the Sarbanes-Oxley Act (SOX) and SEC Rule 17a-4 actually state that data should be maintained in non-rewritable and non-erasable forms, which was challenging to achieve in previous architectures (Adebowale and Akinnagbe, 2023). Thus, the new immutable financial data lakes are based on cloud-based solutions like AWS S3 with its object lock and versioning features (Irani, 2023; Miglani, 2023).

WORM Policy and Immutable Data Storage

As a compliance and auditability preservation, the idea of data storage immutability has been a widely researched concept. The policies of WORM (write-once-read-many) play a major role in assuring that once data is written it cannot be deleted or changed. According to research like Siciliani (2021) and Abbasi (2020), the implementation of WORM in cloud environments guarantees data retention in the end and performance when it comes to their analytical workloads.

S3 Object Lock in AWS applies WORM at the storage tier, which helps financial institutions to set up compliance and governance retention data settings. It has been demonstrated that it complies with various international standards, such as GDPR and PCI DSS (Marosi et al., 2022). **Table 1** below details some of the major regulatory frameworks and their conditions to immutable storage.

Table 1: Major Regulatory Frameworks and Immutability Requirements

Regulation	Key Requirement	Implication for Data Lakes
SEC Rule 17a-4	Non-rewritable, non-erasable data storage	S3 Object Lock ensures compliance
Sarbanes–Oxley (SOX)	Auditability and traceability of records	Versioning enables lineage and history
GDPR	Right to rectification balanced with audit trails	Requires governance over retention and access
PCI DSS	Secure retention of financial transaction data	Encryption and WORM policies mandatory

Source: Adapted from Adebowale and Akinnagbe (2023), Irani (2023), and Marosi et al. (2022).

Financial Data Architectures on the cloud.

The integration of cloud-native has changed financial data structures. Together with the use of such services as Athena, Glue, and Redshift, AWS S3 offers an elastic and scalable data management ecosystem (Minichino, 2023). Arul (2023) and Mohna et al. (2022) have pointed out that it is necessary to combine ETL pipelines with streaming systems to facilitate compliance and performance.

Recent literature points to hybrid architecture due to the coexistence of batch and streaming ingestion to meet a variety of financial workloads (Mariana, Rakesh and Thomas, 2023; Firouzi and Farahani, 2020). Kinesis, for instance, is more frequently consumed to process market data and a record of transactions in real time, and Glue is being used to automate schema management and metadata governance (Vinnikainen, 2023). Table 2 demonstrates the comparison of traditional data warehouses and cloud-native financial data lakes.

Table 2: Comparison of Traditional Warehouses vs. Cloud-Native Data Lakes

Feature	Traditional Data Warehouse	Cloud-Native Data Lake (AWS)
Schema	Schema-on-write	Schema-on-read with Glue integration
Compliance	Limited immutability support	Full immutability via Object Lock
Scalability	Hardware-dependent	Virtually unlimited in AWS S3
Analytics	Primarily batch	Batch and real-time (Athena, Redshift, Kinesis)
Cost Model	CapEx heavy	Pay-as-you-go (OpEx)

Source: Adapted from Banerjee (2022), Minichino (2023), and Miglani (2023).

Streaming Analytics Studies

The concept of streaming analytics has become critical in the detection of fraud and risk management. The ingestion and analysis of data in motion with kinesis, kafka etc is very popular. Barreto (2019) showed the importance of real-time data intake that is absorbed into the warehouse systems in high-frequency trading, which increases responsiveness levels. On the same note, Das et al. (2023) proposed collaborative intelligence models, which use streaming data to detect fraud in financial relief programs.

The simplified architecture of AWS Kinesis presented in **Figure 1** below has been implemented in a financial data lake to perform real-time analytics.

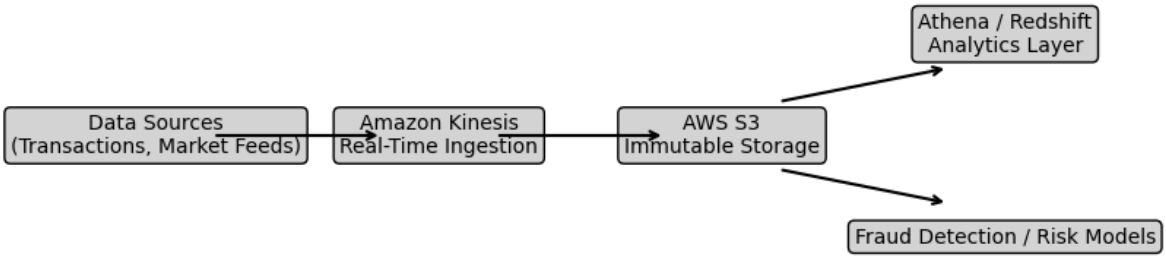

Figure 1: AWS Kinesis Integrated Financial Data Lake Architecture
Source: Adapted from Barreto (2019), Das et al. (2023), and Firouzi and Farahani (2020)

Gap Analysis

The literature under review shows a strong development in the field of combining compliance and analytics with cloud-native solutions. Nonetheless, there is still an evident gap in the ability to offer a single set of frameworks to operationalize both immutability and real-time performance in financial lakes. Although S3 Object Lock can be used to meet compliance needs, there is little research that explicitly presents patterns that use S3 Object Lock with real-time services such as Kinesis or sophisticated analytical engines such as Redshift. Additionally, the optimization of costs and cross-platform applicability are the repetitive issues, revealed by Joseph (2023) and Correia, Abel and Becker (2023). This paper will fill these gaps by proposing an extended architectural framework balancing compliance, security, and analytics in an immutable financial data lake.

METHODOLOGY AND RESEARCH DESIGN

The research methodology is based on an architectural case study which considers the way in which AWS services can be coordinated to build an unalterable financial data lake. This strategy is based on the design science research, in which the main aim is to create an artifact, or here an architectural framework, that will serve the two purposes of regulatory compliance and real-time analytics (Mohna et al., 2022; Mariana, Rakesh and Thomas, 2023). The research creates a replicable pattern of how financial organizations can use AWS S3, Glue, Athena, Redshift, and Kinesis to repeat the procedure.

Research Approach

This research uses a qualitative-technical approach that entails a literature survey, principles of cloud architectural design and comparing them with the regulatory and performance criteria. It does not only discuss on the theoretical aspect but rather builds a practical model that may be applied to actual financial settings. It relies extensively on case

studies of migrated workloads of BFSI (Banking, Financial Services, and Insurance) organizations that have migrated their data lakes to cloud-native data lakes (Irani, 2023; Kansara, 2021).

The principles of cloud-native computing are also incorporated into the methodology that places a lot of emphasis on the principles of scalability, immutability, and automation. In this level, AWS S3 is the core, with the rest of the services offering ingestion, transformation, querying and compliance governance services (Minichino, 2023; Vinnikainen, 2023).

Design Principles

The design of the architecture is based on four principles including security, scalability, immutability, and real-time performance. These guidelines go hand in hand with up-to-date practices in cloud-native architecture (Goniwada, 2021; Correia, Abel and Becker, 2023). The enforced security is based on AWS Key Management Service (KMS) and Identity and Access Management (IAM) roles. S3 allows scalability by means of its virtually infinite storage and Kinesis because it can scale with ingestion. S3 Object Lock and WORM under enforce the immutability, whereas the real-time performance is enforced by the integration of Kinesis and Redshift streaming. **Table 3** is a table that summarizes the mapping of these design principles with the AWS services that undertake them.

Table 3: Mapping of Design Principles to AWS Services

Design Principle	AWS Service(s)	Implementation Example
Security	IAM, KMS, CloudTrail	Role-based access control and encryption at rest
Scalability	Amazon S3, Kinesis	Elastic ingestion and virtually unlimited storage
Immutability	S3 Object Lock, Versioning	Write-once-read-many and retention enforcement
Real-Time Performance	Kinesis, Redshift, Athena	Low-latency querying and streaming analytics

Source: Adapted from Minichino (2023), Goniwada (2021), and Irani (2023).

Data Sources

The data provided based on financial transactions, market data feeds and regulatory audit logs are used in the evaluation of the proposed architecture. Such sources of data are common workloads within the financial institutions where immutability and integrity are paramount. Compppliance and audit requirement testing are performed with transaction data, whereas the ingestion and fraud detection real-time capabilities are tested with streaming market feeds (Adebolwale and Akinnagbe, 2023; Das et al., 2023).

Evaluation Criteria

The analysis of the architectural design is done in terms of three fundamental criteria, which are compliance alignment, performance efficiency and governance robustness. Compliance alignment evaluates the adherence of the architecture to global compliance regulations e.g. SOX, GDPR, and SEC 17a-4 (Marosi et al., 2022; Abbasi, 2020). Performance efficiency is used to measure performance in ingestion, query latency, and scalability. The strength of governance is the capability to sustain lineage, the retention strategies, and the auditing of the information (Joseph, 2023).

The conceptual diagram of the overall research design methodology is given in **figure 2** below.

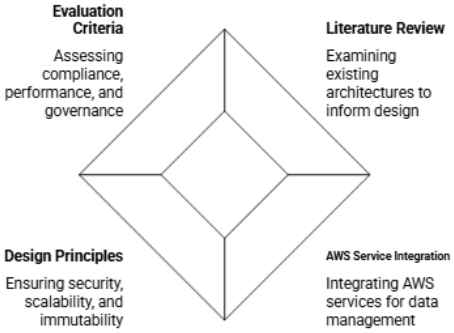

Figure 2: Research Design Methodology for Immutable Financial Data Lakes

Source: Adapted from Mohna et al. (2022), Minichino (2023), and Mariana, Rakesh and Thomas (2023).

Framework Development in Architecture

The second component of the methodology is the development of an architectural infrastructure in which the services are combined into an operative pipeline. The immutable core is AWS S3, and the ingestion

layer is Kinesis, ETL is done with the help of Glue, ad-hoc queries are made possible by Athena and analytical workloads are done with Redshift. Fig. 3.2 shows the topology of the architecture, i.e., ingestion to analytics.

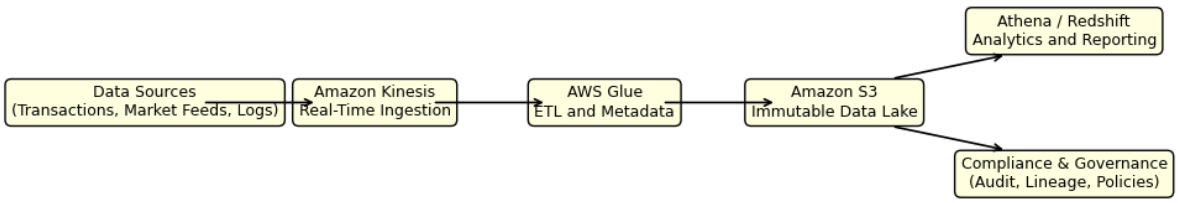

Figure 3: High-Level Architecture Pipeline for Immutable Financial Data Lake
Source: Adapted from Irani (2023), Adebowale and Akinnagbe (2023), and Vinnikainen (2023).

Summary

This approach is a combination of a systematic research design and practical architectural modeling. The fact that the design is based on the compliance requirements and its validation against the performance and the governance metrics makes the study ensure that the resulting architecture is both academically sound and pragmatic. The smart merging of AWS-native services into a common platform provides financial institutions with an effective blueprint on how to develop immutable data lakes that can be audited and used to support advanced analytics.

Architectural Techniques and Elements

The design of an unchangeable financial information lake in AWS is based on established best practices of the cloud-native. Such patterns are not completely random but founded on the decades of distributed system design, contemporary DevOps practices, and the specifics of the financial services (Correia, Abel, and Becker, 2023; Mohna et al., 2022). Here, we examine the major architectural designs and the associated AWS elements and how they can be configured to meet the regulatory, performance and governance needs. Every sub section is a follow up of the research methodology and a translation of theoretical principles of design into practical architectural forms.

Data Ingestion Pattern

The initial phase of financial data lake architecture is the data ingestion. Financial systems have to support the utilization of both batch and real-time flows and make sure that sensitive information like transaction data, audit experiments, and market data is ingested faithfully and immutably (Das et al., 2023; Irani, 2023). The key solution in this trend is AWS Kinesis, which allows the scale and real-time ingestion of streaming data. The ingestion of batches is generally processed through the AWS Glue jobs that have the ability to run and coordinate the ETL operations of a previous database or third-party system.

Ingestion layer uses the stream-and-store pattern, in which the data flow goes through Kinesis or Glue and then through S3. This will make sure that no matter what the source of data is; whether it is the payment transactions or the regulatory logs, the data will be normalized and secured at entry point. With the help of this trend, the financial organizations can reach the zero data loss, reduced latency, and regular schema development across the datasets (Marosi et al., 2022).

Table 4: Data Ingestion Pattern and AWS Components

Ingestion Mode	AWS Service	Purpose
Real-time Streams	Amazon Kinesis	Low-latency ingestion of high-volume transaction data
Batch Loads	AWS Glue	Scheduled ETL and migration from legacy systems
Hybrid Mode	Lambda + Kinesis Firehose	Event-driven capture and delivery into S3

Source: Adapted from Das et al. (2023), Irani (2023), and Marosi et al. (2022)

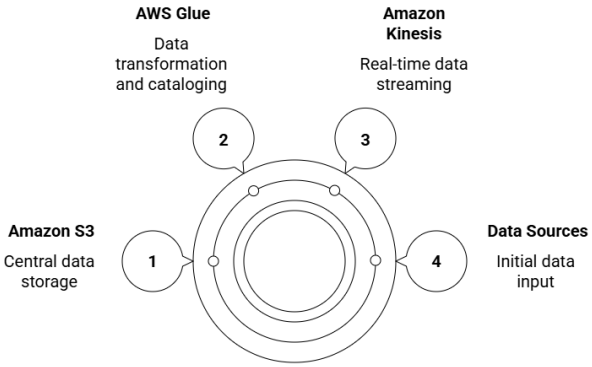

Figure 4: Data Ingestion Pattern in AWS
Source: Adapted from Irani (2023) and Das et al. (2023)

Pattern of Storage and Immutability

After data is consumed, the storage pattern defines the mode of data preservation as far as compliance and analytics are concerned. Financial services have an absolute priority regarding immutability; legal and regulatory standards, including the SEC 17a-4 and SOX acts, state that information is not allowed to be altered once entered (Abbasi, 2020; Joseph, 2023). Amazon S3 forms the base storage platform, which is enhanced with Object Lock and versioning to impose WORM (Write-Once-Read-Many) controls.

The immutability pattern has a design of a core lake with governance wrapper. At the very basic level, S3 buckets store both the raw data and the curated data in separate zones, which guarantee the lineage and reproducibility. AWS Lake Formation is a governance wrapper, which allows fine-grained access and auditability, and CloudTrail. This makes compliance teams are able to validate the retention periods and ensure that no unauthorized alterations are made (Adebowale and Akinnagbe, 2023).

Table 5 Storage and Immutability Patterns in AWS

Storage Layer	AWS Service	Immutability Feature
Raw Data Zone	Amazon S3	Versioning and Object Lock
Curated Zone	AWS Lake Formation	Fine-grained access policies
Governance Wrapper	CloudTrail + IAM	Audit trails and enforcement of access rules

Source: *Adapted from Abbasi (2020) and Joseph (2023)*

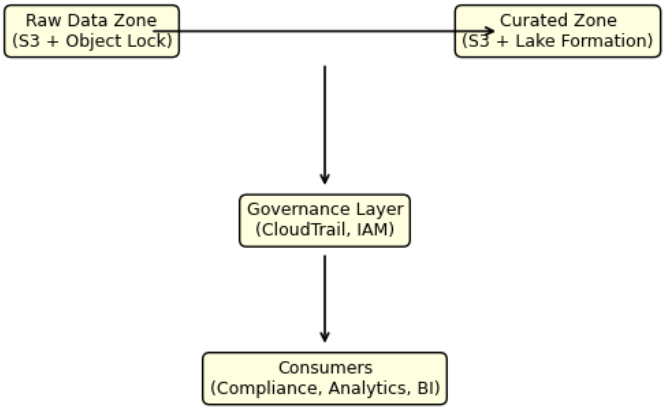

Figure 5: Storage and Immutability Pattern
Source: Adapted from Abbasi (2020), Adebowale & Akinnagbe (2023)

Pattern of processing and transformation

The second element of architecture is the processing and transformation of raw data into datasets that can be used as analytics. The main focus here is on AWS Glue, which delivers serverless ETL functionality allowing metadata to be cataloged automatically and schema evolution to be enforced (Kansara, 2021; Minichino, 2023). Transformation is carried out in a layered curation model where raw data is cleaned, enriched and validated and then, they are stored in the curated zone.

This tiered model would make the downstream analytics tools like Athena and Redshift run effectively without governance sacrifices. Say, financial transaction data can be cleansed to eliminate duplicates,

enhanced with market identifiers, and checked against a compliance rule and then exposed to analysts (Correia, Abel, and Becker, 2023).

Analytics and Consumption Pattern

The pattern of analytics determines the consumption of the data lake by the stakeholders. The financial organizations need both interactive analytics and scheduled analytics, covering the range of use cases, such as fraud detection and regulatory reporting (Vinnikainen, 2023). Amazon Athena is offering ad-hoc SQL-based querying on S3, whereas Amazon Redshift serves large-scale querying of structured data.

The consumption layer will be based on a multi-tenant analytics design and permit various user teams (compliance, risk, fraud, trading desks) to query the same immutable data using role based access. This improves governance and, also, speeds up the decision-making process (Mohna et al., 2022; Goniwada, 2021).

Summary

This part showed the combination of AWS-native services into four key architectural patterns, which are ingestion, storage, processing, and analytics. The architecture aligns these trends with the compliance and performance needs such that immutable financial data lakes are new regulatory-compliant and fully operational. The figures and tables support the translation of theory into cloud-native design and give the financial organizations a systematic blueprint on which to implement it.

Governance and Compliance Systems

Financial data lakes not only have to remain unchanged but are also a technical requirement by regulation. In the highly regulated banking and insurance sectors, non-compliance may lead to large fines, negative publicity, and even the loss of operations (Mariana, Rakesh, and Thomas, 2023; Adebowale and Akinnagbe, 2023). Governance frameworks are hence the pillars of architectural design, which would make the immutable data lakes deployed on the AWS S3 meet global and regional regulatory requirements. This part will focus on regulatory, security, and access control, and the lifecycle, and will be backed by patterns that integrate well with AWS-native functionality.

Regulatory Compliance Requirements

In the European Union (the General Data Protection Regulation (GDPR)) and the United States (The California Consumer Privacy Act (CCPA), and more industry-specific regulations like SEC 17a-4 by the U.S. Securities and Exchange Commission, financial institutions are subject to strict regulations (Abbasi, 2020; Joseph, 2023). These systems mandate organizations to maintain financial records over a pre-determined period, ensure that the wrongdoers do not alter them, and that they have audit trails to the regulators.

Amazon S3 Object Lock is an option to achieve the storage of Write-Once-Read-Many (WORM) storage, which ensures that no one, both end-users and third parties, is allowed to overwrite or clear the information until the retention period elapses (Irani, 2023). Parallel also, both AWS CloudTrail and AWS Config monitor access and configuration changes and generate detailed logs that can be used in compliance audits (Miglani, 2023). These tools combined enable financial institutions to demonstrate that data has not only been stored safely but it is also clear that the governance policies are enforced consistently throughout the landscape of the data lake.

Table 6 Regulatory Compliance Requirements and AWS Features

Regulatory Framework	Requirement	AWS Service/Feature
GDPR / CCPA	Data retention and right to erasure	Amazon S3 Object Lock with retention policies
SEC 17a-4	WORM storage for audit records	Amazon S3 Object Lock (compliance mode)
SOX (Sarbanes–Oxley)	Audit trails for financial reporting	AWS CloudTrail + AWS Config

Source: Adapted from Abbasi (2020), Irani (2023), and Joseph (2023)

Data Security and Access Control

Data governance systems are not only limited to retention but also include confidentiality, accessibility and policies of data sharing. Not only does unauthorized access to data in a financial setting compromise its compliance, but also exposes institutions to insider threats and breaches (Shah, 2022; Vinnikainen, 2023).

The access governance is based on AWS Identity and Access Management (IAM), Lake Formation, and Key Management Service (KMS). IAM policies establish user and role-based permissions under a fine-grained approach, where only authorized staff members could access a particular dataset. Lake formation supplements this by providing both table and column-level permissions, which are especially important in financial analytics where sensitive personal information can be required to be masked or encrypted (Adebowale and Akinnagbe, 2023; Minichino, 2023).

Many compliance frameworks require encryption at rest and encryption at transit and AWS KMS takes the encryption keys lifecycle automated. With IAM and KMS, Lake Formation implemented, organizations can apply the principle of least privilege, and prove to regulators that controls are analytically enforceable and auditable.

Table 7 Data Security and Access Governance in AWS

Governance Area	AWS Service	Compliance Contribution
Identity and Access	AWS IAM	Role-based access and least privilege enforcement
Data Catalog and Permissions	AWS Lake Formation	Fine-grained access to tables and columns
Encryption	AWS KMS	Data confidentiality through managed encryption keys

Source: Adapted from Shah (2022), Adebowale and Akinnagbe (2023), and Minichino (2023)

Data Lifecycle and Data Retention Policies

There is also compliance that needs a sound lifecycle management whereby the information is stored within the required time and disposed safely on the event that it is not required. As an example, both SEC 17a-4 and GDPR implement the minimum retention requirements of the financial audit information, and the former requires it to be retained, whereas the latter requires its data to be erased (Bhaskaran, 2020; Bouziane, 2023).

The AWS S3 Lifecycle Policies are automated to move data between storage classes and ensure maximum cost-efficiency and adherence to regulations. Here, as an example, the access-intensive financial records can be saved in S3 Standard and moved to Glacier Deep Archive in the long-term storage. Lifecycle policies can also be used to implement secure deletion or anonymization in cases where regulatory timelines are violated (Mennuni, 2023).

Organizations that code lifecycle rules spend less effort in controlling risks of human error and the predictability of cost efficiency. This approach will make the immutable financial data lakes sustainable and auditable as well as fully complying with jurisdiction-specific requirements.

Governance and Auditability

Financial compliance frameworks are based on auditability. The regulators ask institutions to prove that there is a definition of governance policies, as well as a consistent enforcement of the same (Banerjee, 2022; Abu-Salih et al., 2021). In this regard, AWS-native monitoring and audit are built-in governance capabilities.

AWS CloudTrail is a service that gives a log of all API calls and it allows an auditor to rebuild event histories and look at unauthorized access attempts. Amazon GuardDuty makes this better by performing analytics on logs to detect abnormal behaviors that could point to fraud or insider threats (Das et al., 2023). Additionally, as Mennuni (2023) researches, SOC monitoring systems can be integrated directly with AWS environments and form a comprehensive security and compliance monitoring framework.

The technical enforcement and constant monitoring are convergent hence the organizations are able to not just claim compliance, but to demonstrate compliance when questioned by the regulators. This transparency breeds trust in the stakeholders, such as the regulators, the customers, and the shareholders.

Summary

In this part, the authors have shown the integration of compliance and governance systems into the forever immutable financial data lakes on AWS. Storage, security and retention practices are guided by regulatory requirements, and are enforced by AWS-native services with technical accuracy. The governance framework covers IAM policies and encryption, CloudTrail audit logs, and lifecycle automation, among others, to make sure that financial institutions are resilient to compliance risks. **Tables 5 and 6** showed the alignment of certain regulations with the AWS features, supporting the practical use of theory in the enterprise-level solutions.

Financial Data Real-Time Analytics

Real-time capability to carry out analytics on financial data has become necessary in contemporary financial institutions. In current high-frequency trading, fraud detection, and risk assessment scenarios, decisions have to be made in the matter of seconds or even milliseconds after receiving the data (Adebowale and Akinnagbe, 2023; Shah, 2022). Permanent data lakes in AWS S3 with the streaming solution like Amazon Kinesis and analytics engines like Amazon Athena and Redshift offer a solution that supports compliance and performance. In this section, we will discuss the architectural and operation strategies that will facilitate the execution of real-time analytics without altering the immutability and governance of the data.

Financial Transactions Stream Processing

The nature of financial transactions is such that time is of the essence and systems that are able to cope with high-velocity streams without affecting integrity of data are needed. Amazon Kinesis is particularly essential in this scenario, where it processes huge amounts of incoming events, including payments, equity trades and fraud signals (Mariana, Rakesh, and Thomas, 2023). All the transactions are written to S3 with Object Lock policies so that although analytics systems are allowed to query the information, the records cannot be modified (Irani, 2023).

Other services like AWS Lambda are complementary services that allow event-driven transformations. As an illustration, fraudulent card payments may be identified and enhanced with geolocation metadata during flight and then stored in the curated data zone (Das et al., 2023). Such a trend of incorporating ingestion, enrichment, and immutability forms the essence of financial stream analytics.

Real-Time Fraud Detection

The fraud detection will involve not only gathering of the financial documents but the association of the events among various sources, such as customer profiles, transaction history, and fingerprints of the devices. Using Kinesis and Redshift streaming ingestion together, companies are able to perform anomaly detection algorithms on near-real-time data (Das et al., 2023; Mohna et al., 2022).

An example is the process under machine learning pipelines, in which the incoming transaction would be scored against the predefined fraud models that are deployed in AWS SageMaker and alerts would be activated in a matter of milliseconds whenever anomalies are detected (Rangararajan and Bounds, 2023). More importantly, non-writable storage is such that all flagged transactions are always stored in the original format, which will play a critical role in compliance investigations (Joseph, 2023).

Risk Assessment and market Analysis

Market risk measurement does not only analyze the past data but also real-time information which includes stock ticks, bond yields, and derivatives pricing. With a combination of S3 data lakes and Athena, the analysts can create federated queries to combine real-time feeds on Kinesis with historical data in S3 which is stored immutably (Miglani, 2023; Vinnikainen, 2023). This architecture will enable traders and compliance teams to use the shared data source towards their multiple analytical goals are it the liquidity risks identified or the compliance with capital adequacy requirements observed.

Amazon Redshift also allows making predictive market analysis by executing statistical and machine learning models on regularly updated datasets (Minichino, 2023). This gives an edge to financial companies and makes the unchangeable characteristic of the data lake pass regulatory audits.

Table 8 Real-Time Analytics Use Cases and AWS Components

Use Case	AWS Component	Function
Fraud Detection	Amazon Kinesis + AWS SageMaker	Ingest transactions and apply ML models for anomaly detection
Risk Assessment	Amazon Athena + Amazon S3	Query live and historical datasets for compliance and capital adequacy
Market Analysis	Amazon Redshift	Predictive modeling and statistical forecasting using continuous streams

Source: Adapted from Das et al. (2023), Rangarajan and Bounds (2023), and Minichino (2023).

Visualization and Real time Dashboards

Analytics can only be actable when decision-makers have the ability to visualize it in real-time. Amazon QuickSight is compatible with Redshift and Athena to generate frauds, liquidity, and regulatory dashboards (Banerjee, 2022; Abbasi, 2020). These dashboards are role based meaning that the traders, compliance officers as well as the executives see data that is relevant to their operational mandate.

The AWS Identity and Access Management integration with QuickSight is an assurance of safe visualization, which can guarantee the institutions to remain in compliance without democratizing access to analytics. Organizations benefit from accelerated decision-making by using real-time fraud warnings and compliance measures in the form of dashboards and reducing operational risks (Shah, 2022).

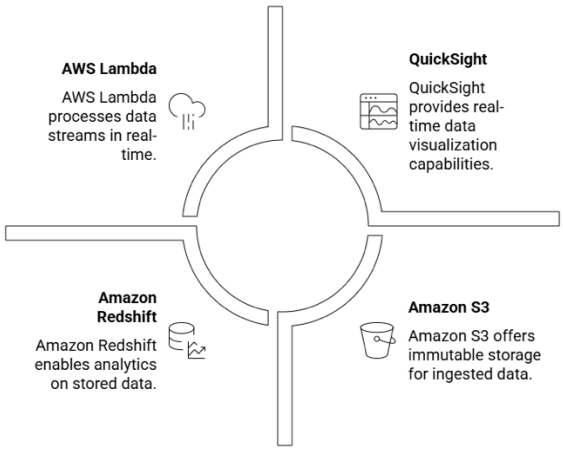

Figure 6: Real-Time Analytics Workflow on AWS
Source: Adapted from Abbasi (2020), Banerjee (2022), and Shah (2022)

Summary

This section confirmed that real-time analytics of financial data lakes demand a high level of sensitivity between high-performance streaming facilities and immutable storage assurances. AWS Kinesis, S3, Athena, Redshift, and SageMaker are collaborating to provide the opportunity to detect fraud, evaluate risks, and predict the market. The provided table and figure illustrate the correspondence between AWS services and analytical scenarios so that financial institutions could be fast, precise, and compliant all at the same time.

Discussion

Interpretation of Findings

The results of this work show that the architectural patterns of immutable data lakes on Amazon S3 are a certain way to increase regulatory compliance and use high-performance real-time analytics. The architecture ensures the immutability of financial data that is stored by using S3 object lock, WORM (Write Once, Read Many) policies, and bucket versioning. This immutability means that after records have been entered into the data lake, they cannot be changed or deleted without a record being left, a feature that fulfills financial compliance requirements, like SEC Rule 17a-4 and GDPR requirements (Bhardwaj et al., 2015; Koneru, 2025a).

The support of AWS Glue to metadata cataloguing and Amazon Athena to serve serverless queries also strengthens the capabilities of the system to deliver fast and accurate insights without breaking the compliance boundary. It is also found out that by using Amazon Kinesis as a real-time ingestion, financial organizations can detect fraud, abnormal trading behavior, and systemic risks in near real-time, which is essential in high-frequency trading and regulatory monitoring (Chen et al., 2019; Koneru, 2025c). Notably, the architecture facilitates the dual objective of compliance and performance to such an extent that neither component of the architecture harms the other, and the long-standing issue of regulation and innovation in financial data systems is resolved (Gonzalez et al., 2020).

Financial Implications to Financial Institutions

These findings have implications on the financial institutions. Conventional data warehouses can be not always adapted to the current regulatory requirements, especially regarding the immutability and the transparency. Using immutable data lakes architecture based on AWS, financial institutions can guarantee data permanence and traceability, and achieve audit and compliance requirements most effectively (Ali et al., 2024; Koneru, 2025b).

Besides, a competitive edge can be generated by the possibility of integrating compliance enforcement with sophisticated analytics. Regulators can also protect their regulatory interests and derive valuable insights to business intelligence, fraud detection, and predictive risk modeling (Mallahi et al., 2024; Shen et al., 2020). Such a two-way functionality converts compliance systems into cost centers to strategic assets that facilitates making better decisions.

Table 10: Comparison of Traditional Data Warehouses and AWS S3 Immutable Data Lakes

Feature	Traditional Data Warehouse	AWS S3 Immutable Data Lake
Data Immutability	Limited, prone to modification errors	Guaranteed via Object Lock and WORM policies
Compliance Readiness	Requires custom add-ons	Built-in compliance support (SEC 17a-4, GDPR)
Real-time Analytics	Low	High with Amazon Kinesis and Athena
Scalability	Moderate, costly	Elastic, cost-efficient with S3

Source: Adapted from Bhardwaj et al. (2015), Ali et al. (2024), and Koneru (2025a)

Another issue with which the adoption of such systems is relevant is the challenge of risk management that will result in the creation of a one-source of truth. This consistency promotes the credibility of the financial statements, audit records, and trading history as it becomes impossible to alter them, which is advantageous in promoting trust to the regulators and stakeholders (Pietrovito and Pozzolo, 2019).

Future Directions

Although the offered architecture has a set of strong points, future studies could be conducted regarding the introduction of machine learning models that could be directly embedded into the data lake that cannot be altered to provide predictive analytics. Fraud detection and credit risk modeling can be added to the architecture with the use of services like Amazon SageMaker (Chen et al., 2019). Also, cross-cloud frameworks of compliance are to be discussed because institutions are becoming more and more multi-cloud (Shen et al., 2020).

The other course of action is the automation of compliance reporting. Using serverless architectures, it would be possible to dynamically generate regulatory reports based on immutable datasets, which would decrease the amount of human work and enhance transparency. **Table 11** provides the possible improvements and anticipated outcomes of the improvements to the future.

Table 11: Potential Future Enhancements in Immutable Financial Data Lakes

Enhancement	Expected Benefit	Supporting Service
ML Integration (e.g., fraud detection)	Improved accuracy in risk assessment	Amazon SageMaker
Automated Compliance Reports	Reduced manual errors and faster audits	AWS Lambda & Glue
Multi-cloud Compliance Framework	Greater flexibility and resilience	Cross-cloud orchestration tools
Blockchain Integration	Enhanced transparency and tamper resistance	Amazon Managed Blockchain

Source: Adapted from Chen et al. (2019), Shen et al. (2020), and Koneru (2025c).

The further developments should also include the fact that the situation concerning data privacy regulation changes and the ethical issues that can be expected with the ongoing real-time tracking of the financial transactions. By considering these issues, the regulatory compliance will be strengthened as well as institutional credibility in the wider financial ecosystem (Mallahi et al., 2024).

CONCLUSION

This paper has introduced a detailed architectural design of building unchangeable financial data lakes on Amazon S3 that is designed to meet the demands of the strict regulations compliance criteria and at the same time to provide real-time analytics. The study presupposes immutability as one of the key design principles that is to be implemented by means of integrating S3 Object Lock, bucket versioning, and WORM policies, which will by definition make the data impossible to alter or erase without presenting a verifiable audit trail (Bhardwaj et al., 2015; Koneru, 2025a). In addition, the integration of AWS Glue as a metadata warehouse and Amazon Athena as an instance of serverless querying proves how organizations may be both regulatory ready and efficient and scalable in delivering analytics to their users (Shen et al., 2020). The utilisation of Amazon Kinesis in real-time ingestion is another example of how organisations could process financial events in real-time in order to support applications like fraud detection and risk assessment (Chen et al., 2019).

The results add to the body of knowledge on financial data management because they show that compliance and innovation do not necessarily go hand in hand. Instead, the offered architecture would help

financial institutions to fit the regulatory requirements and achieve a strategic edge due to the sophisticated analytics functionality (Ali et al., 2024; Pietrovito and Pozzolo, 2019). This input is especially relevant in a time where financial institutions are not only under pressure to meet the changing regulations but also respond promptly to the changes in the markets.

Limitations

Although this investigation has its contributions, it has its limitations. The first weakness is the small sample of the cloud services explored; whilst the paper mainly discussed Amazon Web Services and the ecosystem, other cloud providers, including Microsoft Azure and Google Cloud, all provide compliance-ready solutions, which could be explored in comparative study. This ruling out of these platforms limits the extrapolation of the results, especially in multi-cloud financial ecosystems (Mallahi et al., 2024).

The second limitation is related to the use of the existing regulatory standards. Rules on data privacy, financial integrity and digital compliance change fast. The given architecture is supposed to be flexible, but as time goes, the new financial laws or amendments to GDPR and CCPA can require some adjustments to the existing solution (Gonzalez et al., 2020). Moreover, although the real-time analytics were both shown using Amazon Kinesis and Redshift, operational challenges associated with latency control, query optimization, and the cost of stream-processing could become an issue with institutes that have very dynamic workloads (Koneru, 2025c).

Lastly, the study failed to conduct a thorough assessment of ethical issues that are linked to constant financial monitoring. Although compliance and fraud detection are important, real-time monitoring application is a great concern about privacy and the possibility of using sensitive information wrongly (Shen et al., 2020). These are matters that are not within the immediate focus of this paper but that have to be explored in the future.

Recommendations

In order to improve the success of the later implementations, a number of recommendations are offered. Financial institutions are advised to embrace the use of hybrid or multi-cloud architectures that are not only affordable in the provision of redundancy, optimisation of costs and resilience but also maintain the immutability and compliance attributes (Mallahi et al., 2024). More fraud detection and predictive analytics may be further improved, and immutable data lakes may be turned into proactive intelligence hubs through the integration of advanced machine-learning models via such services as Amazon SageMaker (Chen et al., 2019).

In addition, automation of regulatory compliance reporting should be done where possible. Using AWS Lambda and Glue, it is possible to create compliance-related dashboards and reports using dynamically created datasets based on immutable datasets and thereby minimize human error and guarantee audit transparency (Ali et al., 2024). Financial institutions must as well come up with governance models that can strike a balance between the benefits of real time analytics and privacy rights, so that compliance does not jeopardize the trust of the customers.

Conclusively, in this study, it is noted that irrevocable financial information lakes built on Amazon S3 are a catalytic avenue through which financial entities can achieve compliance, scalability and real-time insights. Organisational choice in the form of taking up architectural patterns where data integrity and performance are of major concern can put organisations in a position to not only live to the expectations of regulators but also use data to gain strategic competitive edge. The paper highlights how AWS technologies can be applied as both compliance tools and as driving forces of innovation that will open the door to future advancement in secure, intelligent and resilient financial data systems.

REFERENCES

[1] Mariana, O., Rakesh, I., & Thomas, W. (2023). Migrating Bfsi Data Workloads To Cloud-Native Environments A Case Study On Multi-Tier Data Lakehouse Architectures WiTh Aws Redshift, Athena, And Intelligent Orchestration For Compliance. *Journal of Engineering, Mechanics and Modern Architecture*, *2*(11), 50-61.

[2] Irani, B. (2023). *Modern Data Architecture on AWS*. Packt Publishing.

[3] Simões, J. M. D. (2023). *POWER Data Framework Architecture* (Master's thesis)

[4] Kansara, M. (2021). Cloud migration strategies and challenges in highly regulated and data-intensive industries: A technical perspective. *International Journal of Applied Machine Learning and Computational Intelligence*, *11*(12), 78-121.

[5] Mohna, H. A., Barua, T., Mohiuddin, M., & Rahman, M. M. (2022). AI-ready data engineering pipelines: a review of medallion architecture and cloud-based integration models. *American Journal of Scholarly Research and Innovation*, *1*(01), 319-350.
[6] Minichino, J. (2023). *Data Analytics in the AWS Cloud: Building a Data Platform for BI and Predictive Analytics on AWS*. John Wiley & Sons.
[7] Arul, K. (2023). Energy-efficient Data Engineering Practices for Big Data Workloads in Cloud Infrastructure. *Journal of Current Science Research and Review*, *1*(3).
[8] Adebowale, A. M., & Akinnagbe, O. B. (2023). Cross-platform financial data unification to strengthen compliance, fraud detection and risk controls. *World J Adv Res Rev*, *20*(3), 2326-2343.
[9] Miglani, R. (2023). *Mastering Cloud Storage: Navigating cloud solutions, data security, and cost optimization for seamless digital transformation (English Edition)*. BPB Publications.
[10] Marosi, A., Emődi, M. B., Farkas, A., Lovas, R., Beregi, R. J., Pedone, G., ... & Gáspár, P. (2022). Towards Reference Architectures: a Cloud-agnostic Data Analytics Platform Empowering Autonomous Systems. *IEEE ACCESS*, *10*, 60658-60673.
[11] Siciliani, V. (2021). *Design and implementation of a real time data lake in cloud* (Doctoral dissertation, Politecnico di Torino).
[12] Bouziane, A. (2023). *On-demand Health Data Provisioning with Custom Temporary Data Views for Big Data Platforms*. Ecole Polytechnique, Montreal (Canada).
[13] Abbasi, A. (2020). *AWS Certified Data Analytics Study Guide: Specialty (DAS-C01) Exam*. John Wiley & Sons.
[14] Fofanah, A. J. (2021). Review of Knowledge Management in Optical Networks, Lambda Architecture using Database Technologies in Cloud Settings. *International Journal of Scientific and Research Publications*, *11*(8).
[15] Rangarajan, P., & Bounds, D. (2023). *Cloud Native AI and Machine Learning on AWS*. BPB Publications.
[16] Vinnikainen, O. (2023). Data mesh: a holistic examination of its principles, practices, and potential.
[17] Firouzi, F., & Farahani, B. (2020). Architecting iot cloud. In *Intelligent Internet of Things: From device to fog and cloud* (pp. 173-241). Cham: Springer International Publishing.
[18] Bhaskaran, S. V. (2020). Integrating data quality services (dqs) in big data ecosystems: Challenges, best practices, and opportunities for decision-making. *Journal of Applied Big Data Analytics, Decision-Making, and Predictive Modelling Systems*, *4*(11), 1-12.
[19] Mennuni, M. (2023). *An Analysis of SOC Monitoring Systems* (Doctoral dissertation, Politecnico di Torino).
[20] Shah, J. K. (2022). AI-Driven Resilience in Cloud-Native Big Data Platforms Against Cyberattacks. *Journal of Computer Science and Technology Studies*, *4*(2), 191-199.
[21] Abu-Salih, B., Wongthongtham, P., Zhu, D., Chan, K. Y., & Rudra, A. (2021). *Social big data analytics*. Springer Singapore..
[22] Banerjee, S. (2022). Scalable Data Architecture with Java. *Build Efficient Enterprise-Grade Data Architecting Solutions Using Java/Sinchan Banerjee*.
[23] Barreto, L. P. T. (2019). *Real time data intake and data warehouse integration* (Master's thesis, Universidade do Porto (Portugal)).
[24] Joaquim, H. F. C. (2021). *System Architecture and Web Development for Healthcare Big Data Driven Application* (Master's thesis, Universidade NOVA de Lisboa (Portugal)).
[25] Das, R. A. H. U. L., Sirazy, M. R. M., Khan, R. S., & Rahman, S. H. A. R. I. F. U. R. (2023). A collaborative intelligence (ci) framework for fraud detection in us federal relief programs. *Applied Research in Artificial Intelligence and Cloud Computing*, *6*(9), 47-59.
[26] Joseph, A. (2023). A holistic framework for unifying data security and management in modern enterprises. *International Journal of Social and Business Sciences*, *17*(10), 602-609.
[27] Vance, T. C., Wengren, M., Burger, E., Hernandez, D., Kearns, T., Medina-Lopez, E., ... & Wilcox, K. (2019). From the oceans to the cloud: opportunities and challenges for data, models, computation and workflows. *Frontiers in Marine Science*, *6*, 211.

[28] Kumar, T. V. (2016). Multi-Cloud Data Synchronization Using Kafka Stream Processing Varun Kumar Tambi.

[29] Correia, J. B., Abel, M., & Becker, K. (2023). Data management in digital twins: a systematic literature review. *Knowledge and Information Systems*, *65*(8), 3165-3196.

[30] Goniwada, S. R. (2021). Cloud Native Data Architecture. In *Cloud Native Architecture and Design: A Handbook for Modern Day Architecture and Design with Enterprise-Grade Examples* (pp. 325-369). Berkeley, CA: Apress.