RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v1 2025 | Submission: April 1, 2025 | Accepted: April 3, 2025 | Publication: April 5, 2025

Distributed architectures and data resilience: the use of big data and machine learning in detecting anomalies in financial transactions.

Distributed architecture and data resilience: the use of big data and machine learning in anomaly detection for financial transactions

Robson Alves dos Santos - MBA in Distributed Software Architecture (PUC Minas); Technologist in Systems Analysis and Development (Cruzeiro do Sul University). Researcher in Cloud Computing, Big Data, and Information Security.

Summary

This article analyzes the evolution of software architectures in the financial context, focusing on the transition from monolithic systems to distributed microservices and its impact on transactional security. The research investigates how the integration of *Big Data* and *Machine Learning* algorithms...
In cloud computing environments, this enables real-time fraud detection with low latency. The methodology addresses the challenges of the CAP Theorem, eventual consistency, and the processing of massive data streams. The results demonstrate that decoupled architectures, when combined with predictive AI models, offer superior resilience and critical responsiveness for mitigating financial risks on a global scale.

Keywords: Distributed Architecture. Big Data. Machine Learning. Fraud Detection. Cloud Computing.

Abstract

This article analyzes the evolution of software architecture within the financial context, focusing on the transition from monolithic systems to distributed microservices and their impact on transactional security. The research investigates how the integration of Big Data and Machine Learning algorithms in cloud computing environments allows for real-time fraud detection with low latency. The methodology addresses the challenges of the CAP Theorem, eventual consistency, and massive data stream processing. The results demonstrate that decoupled architectures, when combined with AI predictive models, offer superior resilience and a critical response capability for mitigating financial risks on a global scale.

Keywords: Distributed Architecture. Big Data. Machine Learning. Fraud Detection. Cloud Computing.

1. Introduction

The stability and security of global financial systems intrinsically depend on robustness of the software architectures that support them. Historically, financial institutions They operated on large mainframes and monolithic systems, where the business logic and the interface... User and data access resided in a single executable block. Although these systems offered initial simplicity of deployment and acid consistency (ACID) in databases. Relational processes have become unsustainable bottlenecks in the face of the exponential volume of transactions. Digital technologies in the 21st century. The inability to scale components individually and the fragility in the face of vulnerability. of isolated failures — where an error in one module could bring down the entire application — forced A paradigm shift towards distributed architectures. This article introduces the discussion. about how modern software engineering responds to these challenges through decomposition of Microservices systems, enabling not only horizontal scalability, but also... The resilience needed to operate in an environment of constant and sophisticated cyber threats.

Year V, v1 2025 | Submission: April 1, 2025 | Accepted: April 3, 2025 | Publication: April 5, 2025

The detection of financial fraud, which once depended on manual analysis or rules. Static solutions based on nightly *batch processing* now require a different approach. In real time. Latency, in this context, is a critical factor: the window of opportunity to block. A fraudulent transaction takes milliseconds. The introduction of *Big Data* and *Cloud* technologies. *Computing* provided the necessary infrastructure to process petabytes of historical data and streams. of real-time events. However, the implementation of these technologies brings with it significant architectural complexities, such as the need to ensure data consistency geographically distributed and the orchestration of ephemeral containers. The central hypothesis of this The work argues that distributed architecture, despite its inherent complexity, is the only viable solution for... to support the *throughput* and availability requirements demanded by modern algorithms Artificial Intelligence applied to security.

The objective of this study is to technically analyze the components that form an architecture. High-performance anomaly detection systems will be examined. Design patterns for these systems will be discussed. distributed methods, such as *Event Sourcing* and *CQRS* (Command Query Responsibility Segregation), and such as They facilitate the ingestion of data into machine learning models . The relevance of this research This is justified by the growing economic damage caused by digital fraud, estimated at billions of dollars. dollars annually, and the need for organizations to adopt proactive stances based on data, rather than reactive. The theoretical foundation is based on specialized literature in science. computing, information security standards, and case studies on *Big* Data implementations. *Data* in critical environments.

2. Evolution of architectures: from monolith to microservices

The transition from monolithic architectures to microservices represents one of the changes deeper studies in software engineering over the last two decades, driven by the need Agility and scalability. In the monolithic model, all the functionalities of a banking system... Everything from account management to credit analysis is coupled in a single codebase. This tight coupling means that any change, however small, requires recompilation and Deploying the entire system increases the risk of regressions and limits the speed of innovation . Furthermore, scalability is vertical and costly: to increase processing capacity of If a specific module is under load, it becomes necessary to duplicate the entire server, wasting resources. computational resources in modules that are idle. In contrast, the architecture of Microservices proposes the decomposition of the system into small, autonomous functional units. independent, communicating through lightweight, language-agnostic APIs, usually via HTTP/REST or gRPC.

This distributed approach allows each service to be developed, deployed, and

Year V, v1 2025 | Submission: April 1, 2025 | Accepted: April 3, 2025 | Publication: April 5, 2025

scaled independently, using the technology best suited to its specific function. (technological polyglotism). For example, a payment processing service can be written in Java for its robustness, while the risk analysis module, which requires processing Matrix intensive, it can be developed in Python to leverage *Data Science libraries.* This flexibility is crucial for fraud detection, as it allows for continuous updating of... Detection models without interrupting the main flow of transactions. If a new attack vector is Once identified, the security team can deploy a new validation microservice or update... An existing model in a matter of hours, an operational agility impossible in other architectures. rigid legacies.

However, the adoption of microservices introduces operational complexity. significant, known as "the hell of distributed complexity". Communication between services Using the network introduces latency and single points of failure that did not exist in function calls in Monolith memory. Fallacies of distributed computing, such as the assumption that the network is reliable. The idea that bandwidth is infinite, or that it has unlimited bandwidth, becomes a dangerous trap. To mitigate these risks, Resilience patterns such as *Circuit Breaker, Retry, Timeout ,* and *Bulkhead* should be implemented. rigorously. *Service Mesh* tools have emerged as solutions for managing this traffic. East-west (between services), offering observability, security (mTLS) and traffic control without Polluting the application code with infrastructure logic.

Data management in microservices also challenges the traditional paradigm. While in In a monolith, there is a centralized database that guarantees referential integrity. Microservices apply the *Database per Service pattern.* This ensures decoupling, but raises... complex issues regarding how to conduct transactions that encompass multiple services and how to maintain A consolidated view of the data is necessary for fraud detection. The solution often lies... in *event-driven architectures,* where changes of state in a Service requests are published as events on a *message broker* (such as Apache Kafka or Amazon). Kinesis), allowing other services, including machine learning engines , to consume This data is processed asynchronously and reacts in real time.

Elastic scalability is another key benefit, enhanced by orchestration of containers (such as Kubernetes). In scenarios of peak transaction volume, such as on *Black Friday,* the The infrastructure can automatically provision new replicas of fraud validation services. to handle the load, and to de-provision them when demand falls. This resource efficiency, Aligned with the cloud cost model *(Pay-as-you-go),* it makes the distributed architecture... economically viable for processing large volumes of data. The ability to Fault isolation also increases overall system availability; if the service of If product recommendations fail, the payment processing service may continue.

Year V, v1 2025 | Submission: April 1, 2025 | Accepted: April 3, 2025 | Publication: April 5, 2025

operating, ensuring business continuity.

Security in distributed architectures requires a "Defense in" approach. "Depth" *(Defense in Depth).* With the increase in attack surface due to the proliferation of APIs and robust authentication and authorization mechanisms, such as OAuth2 and OpenID Connect, make- if required. Each interaction between microservices must be authenticated and encrypted. Furthermore, Managing cryptographic secrets and keys in dynamic environments requires tools. specialized (such as *Vaults)* to prevent credential leaks. Compliance with Regulations such as GDPR and LGPD impose additional challenges in tracking data flows. Personal data is handled through dozens of services, requiring rigorous data governance and auditing. distributed.

It can be concluded that, despite its inherent complexity, microservices architecture is the foundation. indispensable technology for modern financial systems seeking to integrate intelligence. Artificial intelligence provides the agility needed to innovate, the scalability to grow, and the resilience. to survive in a hostile environment. The success in implementing this architecture depends not It's not just about technology, but about a cultural shift in engineering, adopting DevOps practices. Infrastructure automation *(Infrastructure as Code)* and continuous monitoring to tame the Distributed complexity in favor of safety and efficiency.

3. Data consistency challenges and the cap theorem

In large-scale distributed systems, the CAP Theorem (created by Eric Brewer) postulates that it is impossible for a distributed data storage system to simultaneously provide more than Two of the following three guarantees: Consistency (all nodes see the same data at the same time). Time), Availability (each request receives a success or failure response) and Tolerance to Partitioning (the system continues to operate despite message loss or network failures). In this context Regarding financial transactions and fraud detection, this theoretical limitation imposes architectural decisions. difficult. Traditional banking systems prioritize Consistency (CP systems), ensuring that the An account balance must always be accurate, even if that means denying transactions if there is a discrepancy. network failure. However, for fraud detection systems that analyze *Big Data* globally, High availability and partition tolerance (AP systems) are often preferable for To ensure that the analysis does not stop, an Eventual Consistency model is accepted.

Eventual Consistency is a consistency model used in Distributed systems are used to achieve high availability. This ensures that, if there are no new... Updates to a data item will eventually result in all subsequent accesses to that item returning the latest data. updated value. For *machine learning* algorithms that detect fraud patterns, the view Absolute instantaneous access to all global data may not be strictly necessary, provided that

Year V, v1 2025 | Submission: April 1, 2025 | Accepted: April 3, 2025 | Publication: April 5, 2025

the convergence of the data occurs within an acceptable time window (seconds or milliseconds). The use of NoSQL databases (such as Cassandra or DynamoDB), which are Designed for horizontal scaling and offering adjustable consistency models, it allows that Software architects fine-tune the system for the ideal balance between performance and accuracy. data, suitable for the massive ingestion of transaction logs.

The challenge of data synchronization between microservices is often addressed through of the *Saga* pattern , which manages distributed transactions without locking resources (unlike *Two-Phase Commit* - 2PC, which is low-performing in the cloud. Sagas are sequences of transactions. local areas where each transaction updates data within a single service and publishes an event to trigger the next transaction in the saga. If a transaction fails (for example, fraud detection) (positive), the saga executes offsetting transactions to undo the changes made by previous transactions. This mechanism is vital for maintaining the integrity of financial data without sacrificing system availability, allowing complex validation processes to occur In an asynchronous and resilient way.

Data replication across multiple availability zones or geographic regions is a This is an essential strategy for data resilience , but it introduces replication latency. In a global fraud scenario, where a credit card can be used simultaneously in two... On continents, the speed of light becomes a limiting factor. Modern architectures utilize *Conflict-Free Replicated Data Types* (CRDTs) and consensus algorithms such as Raft or Paxos (in implementations such as etcd or Consul) to resolve data conflicts in a deterministical way and efficient. A deep understanding of these algorithms is necessary to avoid data anomalies. such as "double spending" or false negatives in risk analyses, ensuring that the overall state of The system should be coherent.

Streaming transaction processing requires handling the order of events and the delivery semantics *(at-most-once, at-least-once, exactly-once).* For financial systems, the *Exactly-once* semantics is the "holy grail," ensuring that each transaction is processed only once. One time, no more, no less. Stream processing frameworks like Apache Flink or Spark. Streaming services, integrated with messaging platforms like Kafka, implement mechanisms for... *Checkpointing* and end-to-end transactional capabilities are used to provide these guarantees. This allows the Fraud detection systems maintain accurate state counters (e.g., number of transactions). in the last 10 minutes) even in the face of processing node failures, ensuring the reliability of *features* fed into AI models.

Data integrity also involves schema *validation* . Distributed environments. With multiple services producing and consuming data, the evolution of Schemas (changes to the JSON/Avro/Protobuf data structure) can break compatibility and

Year V, v1 2025 | Submission: April 1, 2025 | Accepted: April 3, 2025 | Publication: April 5, 2025

This can cause interruptions. The use of centralized *Schema Registries ,* which govern the versioning of contracts. Data management is a recommended practice to ensure that the data ingested by the *Data Lake* and by the *Machine learning* models must be structurally valid and interpretable. This avoids the problem "Data Swamp," where corrupted or poorly formatted data renders the process unusable. Predictive analytics compromise the effectiveness of security algorithms.

In short, managing data consistency and resilience in distributed architectures is not... a trivial problem solved by database software alone, but an architectural challenge. that permeates the entire application. The correct choice of persistence technologies, combined with patterns of Resilient design and a clear understanding of the trade-offs imposed by the CAP Theorem is Fundamental. For fraud detection, where the accuracy of information is the main weapon against it. Cybercrime, the ability to architect systems that maintain data integrity under Adverse conditions are what differentiate a robust infrastructure from a vulnerable one.

4. Big data and stream processing for security

The concept of *Big Data* in financial security transcends the volume of data *(Volume).* also encompassing the speed of generation and processing *(Velocity)* and the diversity of sources. *(Variety)* data . To detect sophisticated fraud, it is not enough to analyze structured transactional data (value, date, time). It is necessary to correlate this data with terabytes of information. unstructured data, such as access logs, device geolocation, browsing patterns of User *(clickstream) data,* behavioral biometrics, and even social media data. The architecture The traditional *Data Warehouse model,* based on nightly ETL *(Extract, Transform, Load)* processes , is unable to handle that speed. The technological answer is the *Lambda* or *Kappa architecture,* which It allows for hybrid processing of historical (batch) data and real-time (speed layer) data.

The Speed Layer *is* supported by processing platforms of Continuous events. Technologies like Amazon Kinesis or Apache Kafka act as the system. central nervous system, processing millions of events per second with very low latency. These *brokers* Distributed messaging ensures data durability and allows for decoupling between the... Producers (transactional systems, mobile apps) and consumers (fraud engines, dashboards). The ability to partition data enables massively parallel processing: multiple *workers.* They can analyze different fragments of the data stream simultaneously, allowing the system to... Scale linearly as transaction volume increases, while maintaining detection performance. stable.

Real-time data enrichment is a critical step. When a transaction arrives, It carries limited information. The *stream processing* system must, in milliseconds, consult it. reference databases (usually stored in distributed in-memory caches such as

Year V, v1 2025 | Submission: April 1, 2025 | Accepted: April 3, 2025 | Publication: April 5, 2025

Redis or Memcached) to add context: "Has this device already been used by another user?", "The "Is the IP address from a known VPN?", "What is the physical distance of the last transaction?". This *join* process Switching between streams and static tables in real time requires a data architecture optimized for read speed. low latency, often using NoSQL key-value databases optimized for Extreme performance.

Graph analytics emerges as a powerful tool within the *Big Data* ecosystem to identify fraud rings and dummy accounts. Oriented databases Graphs (such as Neo4j or Amazon Neptune) allow us to model the complex relationships between entities. (users, accounts, devices, addresses). Community and centrality detection algorithms These tests can be performed on these graphs to identify suspicious patterns, such as multiple users. sharing the same device or circular money transfers. The integration of these Graph analysis in the *Big Data* pipeline adds a layer of relational intelligence that Traditional statistical methods fail to capture this.

Long-term storage *(Cold Storage)* in *Data Lakes* (such as Amazon S3 or HDFS is fundamental for forensic analysis and model retraining. All raw events They are stored in an immutable and inexpensive way. Distributed query tools (such as Amazon) Athena, Presto, or Google BigQuery allow data scientists to explore these petabytes of data. Historical data can be used to discover new attack patterns *(zero-day attacks)* and test new hypotheses. The separation between computing and storage in these cloud architectures allows for a Cost and performance flexibility, where processing resources are allocated on demand. only when complex analyses are required.

Data quality *and* data lineage *are* challenges. Expanded on *Big Data.* Garbage in results in garbage out . Data engineering pipelines should include automated validation, cleaning, and... Monitoring anomalies in the data itself (e.g., a sudden shift in the distribution of values). Data observability tools help ensure that *machine learning* models do not... be poisoned by corrupted or manipulated data by attackers *(Adversarial Machine) Learning),* while maintaining the reliability of the security system.

In conclusion, the application of *Big Data* and *stream* processing is the engine that enables... Modern financial security. The ability to ingest, enrich, analyze, and store volumes. Massive amounts of heterogeneous data in real time transform fraud detection from an activity From reactive to predictive and preventative capabilities. The underlying data architecture must be Designed for elasticity, fault tolerance, and low latency, serving as a solid foundation upon in which artificial intelligence can operate to protect the financial ecosystem.

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year V, v1 2025 | Submission: April 1, 2025 | Accepted: April 3, 2025 | Publication: April 5, 2025

5. Machine learning and cloud application (AWS)

Artificial intelligence, specifically *Machine Learning* (ML), represents the brain of Fraud detection operation. While rule *-based systems* are Effective for known and simple patterns, they fail to detect new and complex frauds and They generate many false positives. Supervised ML models (such as *Random Forest, Gradient Boosting* and Deep Neural Networks) learn patterns from labeled historical data (fraud). vs. legitimate) to predict the likelihood of fraud in new transactions. The cloud, specifically Amazon Web Services (AWS) provides the ideal ecosystem for the complete ML lifecycle. *(MLOps),* from *feature* engineering to model deployment and monitoring in production.

Training robust models requires massive computing power. Services such as AWS SageMaker enables data scientists to provision clusters of high-performance GPU instances. On-demand performance to train complex models in hours, rather than days. The capability to parallelize training and perform automatic hyperparameter optimization *(Hyperparameter Tuning* accelerates experimentation and convergence toward more accurate models. Furthermore, the cloud... It facilitates the use of *AutoML techniques,* which democratize access to high-quality models. allowing engineering teams to focus on business logic and interpreting results.

Real-time inference is the critical moment where the model comes into play. Microservices architecture allows encapsulating the trained model in a container (Docker) and Expose it as a REST API (via SageMaker Endpoints or AWS Lambda). When a transaction When this happens, the data is sent to this *endpoint,* which returns a risk *score* in milliseconds. Cloud's *automatic* scaling ensures that if transaction volume triples, the system will be able to handle the situation. Suddenly, new replicas of the model are automatically started to keep latency low. without human intervention, ensuring that the end-user experience is not degraded (e.g. (delay in card approval).

The challenge of class imbalance is evident in fraud detection: transactions Fraudulent transactions represent a tiny fraction of the total (often less than 0.1%). This may to bias the models to always predict "legitimate". Advanced techniques such as SMOTE *(Synthetic Monomorphism Testing) Minority Over-sampling Technique)* or the use of weighted cost functions are applied during Training to mitigate this. Additionally, the use of unsupervised learning (such as *Isolation) Forests (or Autoencoders)* help detect anomalies that do not match any pattern. A previously known fraud, acting as a safety net against unprecedented attacks.

Explainability of models *(Explainable AI* - XAI) is a regulatory requirement and Operational growth. Financial institutions need to explain why a transaction was denied. "Black box" models like *Deep Learning* are difficult to interpret. Techniques like SHAP

Year V, v1 2025 | Submission: April 1, 2025 | Accepted: April 3, 2025 | Publication: April 5, 2025

(SHapley Additive exPlanations) or LIME are used to assign the contribution of each variable. (e.g., high value, unusual location) for the model's final decision. Integrate these explanations into The API response allows human fraud analysts to validate the model's decisions and improve them. The system continuously creates a *human-in-the-loop* feedback loop .

Monitoring models in production is vital for detecting *model drift* (degradation). (of the model). The behavior of fraudsters changes rapidly; a model trained with data A tool that was available six months ago may be obsolete today. Cloud-based monitoring tools (such as SageMaker Model Monitor continuously analyzes statistics from the input data and the Predictions, alerting if there are significant deviations. This triggers retraining pipelines. automatic *(CI/ CD for* ML), ensuring that the security system evolves organically along with the threats, maintaining the effectiveness of the protection over time.

It can be concluded that the symbiosis between *Machine Learning* and *Cloud Computing* is the state of the art in... Fraud prevention. The cloud removes infrastructure barriers, enabling the use of algorithms. computationally intensive on a global scale. The application of advanced ML techniques, Supported by a resilient distributed architecture, it allows not only reacting to fraud, but also... Anticipating them with surgical precision. For the software architect, the challenge lies in orchestrating these... complex components — data, models, infrastructure — in a cohesive, secure and High-performance solutions that protect financial assets without compromising the agility of digital businesses.

6. Conclusion

The analysis conducted throughout this article has shown that the security of transactions Financial issues in the digital age are no longer a problem that can be solved through traditional approaches. static or monolithic. The complexity and speed of modern attack vectors demand a an equally sophisticated architectural response, based on distributed processing, in Infrastructure elasticity and predictive intelligence. The migration to architectures of Microservices, while introducing significant challenges in data coordination and consistency, It has proven to be the only model capable of supporting global scale and the need for continuous innovation. of financial institutions. Resilience, in this context, ceases to be a hardware characteristic. Redundant, it is becoming an emerging property of distributed software, capable of tolerating faults. partial without systemic collapse.

Research into the role of *Big Data* has demonstrated the ability to process data streams. Real-time data *processing (stream processing)* is the competitive advantage in the fight against fraud. modern data architecture, which integrates velocity layers and batch layers , allows A holistic and contextualized view of each transaction. The millisecond latency achieved by Distributed messaging technologies and in-memory databases enable the interception of

Year V, v1 2025 | Submission: April 1, 2025 | Accepted: April 3, 2025 | Publication: April 5, 2025

Illegal activities are being targeted before they result in financial losses. It is clear that engineering... Data security is as critical to security as encryption or firewalls, as it provides the raw material. essential primary for automated decision making.

The study of the application of *Machine Learning* revealed that artificial intelligence is the The ultimate tool for dealing with the variability and sophistication of fraud. The ability of Models that learn non-linear patterns and adapt to new threats far surpass the effectiveness of systems based on rigid rules. The cloud, exemplified by the AWS platform, It democratized access to this technology, allowing the training, deployment, and... Model monitoring should be seamlessly integrated into development operations. software *(DevOps).* The explainability and continuous monitoring of the models ensure that this Automation should remain transparent, ethical, and effective over time.

It was also observed that theoretical challenges, such as the CAP Theorem, impose real commitments that must be managed by software architects. The adoption of models of Consistency, when well implemented, provides the necessary balance between Availability and accuracy for anomaly detection applications. The use of standards of Projects like Sagas and *Event Sourcing* mitigate the risks of data inconsistency in various environments. distributed, ensuring that financial integrity is maintained even in high-growth scenarios. Network concurrency and failures. Mastering these theoretical concepts and their practical application are... Essential skills for the modern technology professional.

The research highlighted the importance of in-depth security within the company itself. Distributed architecture. API protection, data encryption at rest and in transit, and... Robust identity management is a non-functional requirement that should be incorporated from the very beginning. *Security by Design.* The complexity of microservices demands complete observability. — logs, metrics, and distributed tracking — so that security teams can detect and Responding to incidents quickly. Security, therefore, becomes a responsibility. shared and integrated into the software development lifecycle.

Furthermore, it can be concluded that cloud technology acts as a facilitator and accelerator. of these architectures. The ability to provision resources on demand allows systems to Security measures can keep pace with seasonal transaction peaks without wasting capital. The use of Managed services reduce the operational burden on technical teams, allowing them to focus more on their core business. Stay focused on fraud detection and business innovation. The cloud is not just a... not just a place to stay, but a platform of capabilities that enhances the effectiveness of strategies. Cybersecurity.

In short, the convergence of distributed architectures, *Big Data* , and *Machine Learning* defines... The new standard of excellence for financial security. Organizations that master integration.

Year V, v1 2025 | Submission: April 1, 2025 | Accepted: April 3, 2025 | Publication: April 5, 2025

These disciplines are better positioned to protect their assets and the confidence of their clients.

In an increasingly digital and hostile market, scientific research and technological development are crucial.

Continuous improvements in these areas are of vital interest for economic stability, as they ensure that...

Financial infrastructure must remain resilient, reliable, and secure in the face of future challenges.

Finally, it is recommended that future research explore the impact of computing.

Quantum mechanics in cryptography and distributed systems security, as well as the use of machine learning.

Federated *Learning* to enable collaboration in fraud detection between different

institutions without compromising data privacy. Technological evolution is constant, and the

Software architecture must remain adaptable and vigilant to continue fulfilling its mission.

Criticism of safeguarding the global digital economy.

References

BREWER, EA CAP twelve years later: How the "rules" have changed. **Computer,** vol. 45, no. 2, p. 23-29, 2012.

NEWMAN, S. **Building Microservices: Designing Fine-Grained Systems.** 2nd ed. Sevastopol: O'Reilly Media, 2021.

KLEPMANN, M. **Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems.** Sebastopol: O'Reilly Media, 2017.

AMAZON WEB SERVICES. **Machine Learning on AWS.** Available at: [link]. Accessed on: October 15, 2025.

GHEMAWAT, S.; GOBIOFF, H.; LEUNG, S.-T. The Google File System. **ACM SIGOPS Operating Systems Review,** v. 37, no. 5, p. 29-43, 2003.

VERBA, N. et al. A review on fraud detection using machine learning techniques in the financial sector. **International Journal of Computer Applications,** vol. 176, no. 1, p. 34-40, 2020.

RICHARDSON, C. **Microservices Patterns: With Examples in Java.** Shelter Island: Manning Publications, 2018.

KREPS, J. I Heart Logs: Event Data, Stream Processing, and Data Integration. **O'Reilly Media,** 2014.

DEAN, J.; GHEMAWAT, S. MapReduce: Simplified Data Processing on Large Clusters. **Communications of the ACM,** vol. 51, no. 1, p. 107-113, 2008.

SHOARAFI, A. et al. Real-time credit card fraud using detection machine learning algorithms. **Journal of Supercomputing,** vol. 77, p. 1-25, 2021.