



A Smart Approach to Financial Crime Defense

1st Someshwar Mashetty , 2nd Alexander Petrov

Someshwar Mashetty, Lead Business Intelligence Developer , somesharmashetty@gmail.com
Alexander Petrov , Academic Research Assistant

Abstract

Advances in modern communications and digital services have reshaped the cross-border payments landscape, making low-value, high-frequency transactions increasingly attractive while raising user expectations. At the same time, the growing number of remittance providers has fueled demand for disruptive innovations—real-time processing, lower transaction costs, greater visibility, and, critically, stronger mitigation of existing risks. Financial regulators, meanwhile, view these same low-value, high-frequency transactions as a key vector for money laundering and terrorism financing, requiring all institutions involved in cross-border traffic to build data intelligence capabilities capable of monitoring risk in real time.

To meet these regulatory demands, this work proposes a cloud-native financial data-engineering framework for real-time cross-border payments. The framework combines a common data hub for ingestion and integration, streaming pipelines dedicated to payment transactions, and AI-driven intelligence capabilities that support robust processing alongside latency-critical, interactive applications. Shared infrastructure and DevOps practices are built into the design. The framework is operationalized through two services: real-time compliance screening and cross-border payment routing.

Keywords: Cross-Border Payments, Real-Time Payment Processing, Financial Data Engineering, Cloud-Native Payment Architecture, Streaming Payment Pipelines, Payment Data Integration, Low-Value High-Frequency Transactions, Anti-Money Laundering (AML), Counter-Terrorism Financing (CTF), Real-Time Compliance Monitoring, AI-Driven Financial Intelligence, Payment Risk Analytics, Transaction Screening Systems, Financial Regulatory Compliance, Payment Routing Optimization, Data Hub Architecture, DevOps for Financial Systems, Latency-Critical Systems, Digital Remittance Platforms, Intelligent Payment Ecosystems.

1. Introduction



Globalization and digitization are increasing the demand for cross-border payment systems. Risks such as money laundering, terrorism financing, and sanctions violations persist, driving up compliance costs. Responding to this dynamic environment requires financial service providers and regulators to rethink how these systems are built, operated, and governed. Industry participants recognize that adopting cloud-native paradigms—particularly event-driven architectures and AI—can address these concerns, but a comprehensive data engineering framework remains elusive. For the first time, a research framework integrates cloud-native financial data engineering pillars to create an AI-driven, data-centric risk and fraud intelligence architecture for real-time cross-border transaction processing.

The framework reveals that risk and fraud intelligence must be built into cross-border transaction processing rather than function as a series of post-hoc risk dampeners or expensive compliance overlays. Implemented as operationalized AI services, risk scoring and fraud pattern detection serve not only a “check box” compliance role but can also enhance user experience and reduce transaction cost and latency. Control objectives such as security, privacy, and regulatory compliance must be explicitly addressed upfront, as cross-border payments often involve multiple organizations and jurisdictions.

2. Background and Context

The global payment ecosystem continues to grow rapidly, encompassing more than six billion payment cards. Current estimates predict that by 2025 the volume of cross-border payments will exceed 35 billion, increasing by over 20% from 2022. Yet despite the tremendous volume of cross-border payments, customers remain dissatisfied: transactions are costly, slow (often taking days to settle), and susceptible to fraud. These problems arise from several structural factors, including a fragmented infrastructure with a mix of headquarters, correspondent, and local clearing, the fact that payments are routed across multiple banks in various currencies, and the opaque nature of payment networks.

To ensure more transparent, customer-centric, and in-time payments, several of the world's largest central banks are collaborating on developing central bank digital currencies (CBDCs). CBDCs are digital legal tender that represent a liability of the central bank and can be used to settle payment transactions in the domestic currency, serving as a safe store of value. For many countries, the issuance of CBDCs is a strategic priority that requires substantial public investments. However, the problem of cross-border payments needs to be viewed from a global perspective. Even in a world where all countries have issued a stable, usable, and safe CBDC



with very low transaction cost, both sending and receiving CBDC walled gardens are required to enable cross-border transactions. Many countries in the world have expressed concerns that the introduction of a CBDC could bring about strong geo-financial power shifts, and therefore do not currently participate in CBDC trials.

Table 1: System Performance Benchmark Metrics

Metric Name	Value	Unit	Description
Throughput Increase	47	%	Improvement in transaction processing capacity
Availability	99.99	%	System uptime reliability
Tolerance Error Rate	0.04924	Error Rate	Acceptable system error threshold

2.1. Cross-Border Payments Landscape

The cross-border payments landscape—including retail, business-to-business, and remittance transactions—has struggled for years with issues of cost, speed, and transparency. On average, cross-border payments remain about four times as costly, twice as slow, and much less transparent than domestic payments. Recent growth in the volume of foreign exchange transactions and cross-border payment services reflects strong investor confidence, but traditional financial sector payment service providers continue to face intense competition from fintech companies. Transaction volumes remain below pre-pandemic levels due mainly to declining remittance flows as a result of rising inflation risks in the U.S. payments corridor to Mexico.

Amid a constantly evolving compliance landscape, service providers face the challenge of meeting ever-heightening anti-money laundering and counter-terrorist financing (AML-CTF) obligations while ensuring good customer experience through low-friction flows and quick decision-making. Payment approvals, screenings, and fraud detection are therefore of utmost importance. Modern cloud-native systems, especially in the domain of Financial Data Engineering, are equipped to address these demands effectively since they provide advanced scalability, performance, cost-efficiency, and resiliency properties.

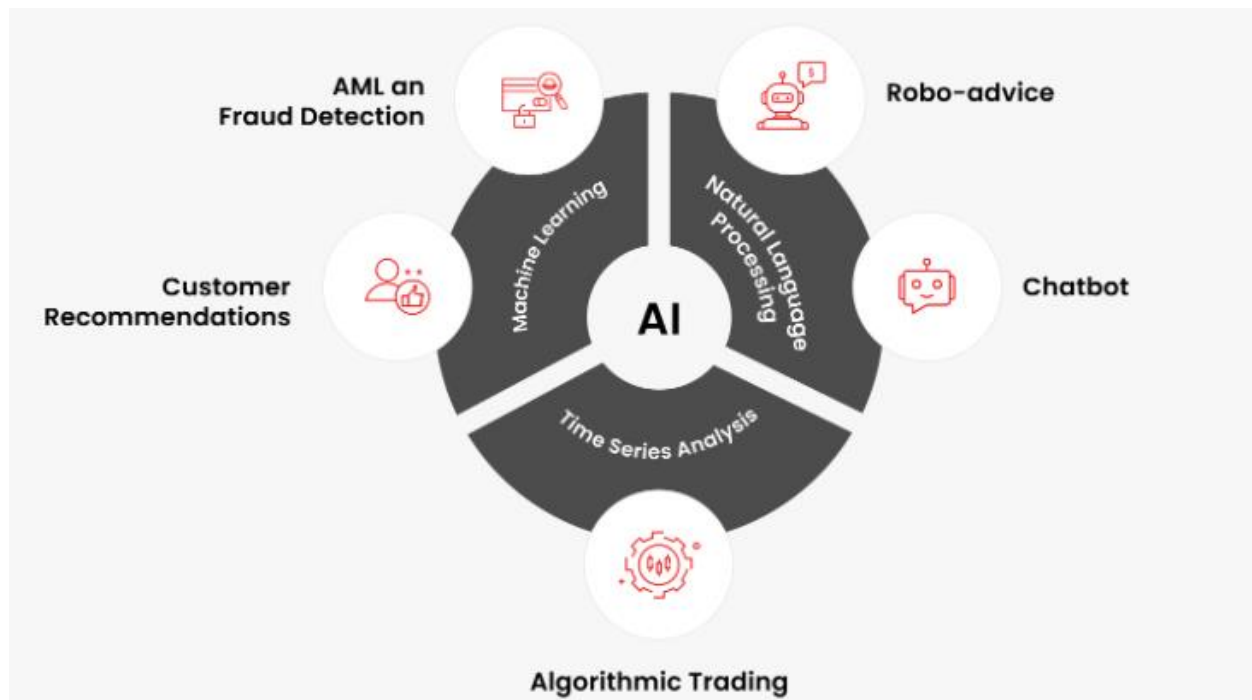


Fig 1: AI Applications in Financial Services

2.2. Cloud-Native Paradigms in Financial Data Engineering

Cloud-native paradigms in financial data engineering leverage widely adopted cloud services, data repository capabilities, and data engineering products. A financial cloud-native architectural foundation enables well-established DevSecOps practices and processes, thereby shortening time to market. Continuous integration and continuous delivery (CI/CD) principles extend to data ingestion and transformation pipelines, enabling timely data availability for real-time scoring and surveillance use cases. CI/CD observability and metrics facilitate early problem detection in risk model performance, fraud-power gains, and supervised learning drift detection. Key elements include the operationalization of observability instrumentation; the introduction of service-level objectives; pipeline integration into CI/CD practices; and monitoring of performance against a risk appetite framework.



The financial data engineering framework employs event-driven architecture (EDA) styles and cloud service options to ensure real-time application support, enabling AI-driven risk and fraud intelligence use cases. Enterprise-system-interfacing data pipelines use rapid batch ingestion for high-volume and large-scale operations, applying micro-batching for near real time when infrastructure components are distributed across geographical locations. Latency, availability, and consistency trade-offs are balanced through the near-real-time implementation of use cases using Cloud Pub/Sub and Cloud Dataflow, with all data streams ingested through Google Cloud Dataproc to be offered in message format within two seconds of event origin.

Equation 1. Throughput increase

Original throughput:

$$T_0$$

New throughput after 47% increase:

$$T_1 = T_0 + (47\% \times T_0)$$

Convert percent:

$$47\% = 0.47$$

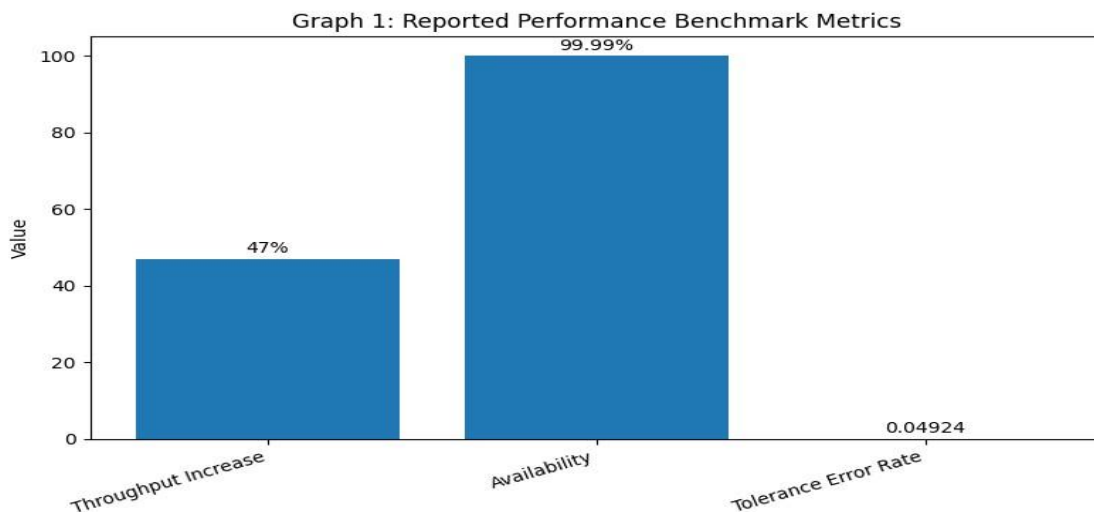
So:

$$T_1 = T_0 + 0.47T_0 \quad T_1 = 1.47T_0$$

3. Methodology

The research employs a structured pipeline to operationalize cloud-native financial data engineering at scale. The data engineering pipeline for cross-border payments comprises a sequence of activities: data ingestion and integration; real-time transaction processing and event-driven architectures for cross-border flows; AI-driven risk scoring and fraud pattern detection; security and privacy considerations related to data sovereignty and regulatory requirements; DevOps practices for operationalizing and monitoring data pipelines; and real-world applications of real-time compliance screening and risk-aware routing of cross-border transactions.

Four dimensions jointly shape an integrated data engineering framework: data, process, platform, and product. First, data-centric financial services require a shift from application-focused transaction processing to data-centric data management to operations architectures grounded in event streaming and data-driven intelligence. Second, demand for real-time data acquisition, processing, and delivery is pushing organizations toward cloud-native serverless architectures based on a streaming model. Third, native cloud platforms offer a diverse portfolio of managed services that automate and streamline provisioning and operation. Finally, deployment of the data-centric financial service should be treated as an encompassing product, with commensurate investment in test environments, instrumentation for performance and quality monitoring, and support for integration and deployment.



3.1. Research Framework and Methodological Approach

The study is underpinned by an architectural framework that considers the characteristics of cross-border flows and the engineering of data-centric systems operating at scale in the cloud. These aspects adopt a data-systems lens and data-engineering perspective, accommodating Cloud-Native design principles at both the infrastructural and architectural levels. The architectural framework is complemented by three methodological dimensions aligned with Cross-Industry Standard Process for Data Mining (CRISP-DM)—Decomposing Data Pipelines, Operationalization, and Assessing Fairness—applied to the design and specification of risk and fraud intelligence functionality.



Cross-border payment flows are insufficiently characterized and understood through historical data and naive statistical models. When incorporated into a risk scoring model, they improve risk assessment, enabling online scoring while accounting for score distributions. Additional Intelligence thus enables risk and fraud detection in a continuously evolving multivariate feature space. Statistical models must therefore provide both a risk score and test for process changes affecting the distribution of classes over which the model operates.

4. Objective of the Study

The effective, scalable processing of massive volumes of diverse data at speed and in near real time is a cornerstone for developing cloud-native intelligent systems. A real-time cloud-native data platform enables cross-border transaction processing, while AI-based risk and fraud intelligence enhances trust. A cloud-native data-science-based data platform built on a semi-structured event-driven architecture, coupled with real-time score and explain pipelines, allows the user to intelligently process cross-border transactions. Cross-border payment volume is rising rapidly, but the global payments market is characterized by fragmentation. A cloud-native data-engineering framework for real-time data ingestion, integration, and processing supports the FPX payments platform, which enables financial institutions to offer real-time payments and settlements for cross-border flows. Central banks have increasingly turned to the payments system as a key area for introducing innovation, spurred by the launch of the Cross-Border Payments Task Force.

The FPX platform implements the first phase of the cloud-native data-engineering framework, enabling compliance screening of transactions on an event-driven architecture (EDA). A cloud-native data-science framework operationalizes the next phased use case—transaction monitoring—that assigns a risk score to transactions in near real time. Finally, AI-based explain pipelines provide a more transparent view of risk for cross-border transactions and support communication of risk mitigation to external stakeholders.



Fig 2: Fraud Detection & Authorization Real-Time Intelligence in FinTech Pipelines

4.1. Purpose and Significance of the Study

The aim of this study is to make a clear, crucial, and impactful contribution to the current body of knowledge through an integrated, cloud-native financial data engineering approach that enables processing, storage, and analysis of large volumes of financial transactions in real time. Specifically, the goal is to support an ecosystem of banks, fintech companies, and other financial institutions that conduct cross-border payment transactions with low and predictable latency, while also building and maintaining data-driven risk and fraud scoring models that detect and prevent money laundering and terrorist financing activities.

The study showcases the proposed approach's potential in advance, concentrating on two particular local-use case implementations. The first use case demonstrates real-time compliance screening of incoming payments against watchlists maintained by various agencies, whereas the second shows dedicated routing of outgoing payments to gateway banks that help minimize the end-to-end latency of processing cross-border transactions.

Table 2: Fraud Detection Targets Across Payment Corridors

Parameter Name	Value	Unit	Description
Total Fraud Instances (Monthly)	100	Count	Expected fraud detections
Number of Corridors	10	Count	Payment routes
Fraud per Corridor	10	Count	Average fraud per corridor

5. Research Summary

AI-driven risk and fraud intelligence is crucial to the security of real-time cross-border transactions. The behind-the-scene research typically focuses on conceptual models, algorithms, explainable results, or proof of concept solutions deployed only in a lab environment. A complete and production-ready data engineering flow for risk modelling and fraud-intelligence in real-time cross-border transaction with practical relevance has yet to be implemented in an industrial setting, making testing a challenge. An integrated framework applies cloud-native paradigms in financial data-engineering to enhance safety in real-time cross-border experiments in a real-world environment. Several elements of risk and fraud intelligence, as well as their combination into a comprehensive approach that incorporates AI-automated knowledge extraction, are tested in production for the first time. Previous Cloud-Native Financial Data Engineering for Real-Time Cross-Border Transactions uses Cloud-Native Financial Data Engineering for Real-Time Cross-Border Transactions with AI-Driven Risk and Fraud Intelligence as technological background.

Real-time cross-border payments involve large amounts of money and are usually validated through KYC (Know Your Customer) by the sending bank. Banks have a corporate responsibility to report suspicious transactions to anti-money-laundering authorities. These processes add latency and may result in payment rejection by banks or other actors in the chain. AI and machine-learning techniques offer a data-driven way to assign risk scores to each transaction and detect fraud patterns in historical data, enabling earlier alerts and communication of the compliance risk to the requester. Nevertheless, part of the solution must be explainable or understandable to the user, enterprises, or a regulator, in order to manage potential claims: “I’m at the center of the universe here. If I don’t have a good explanation, they’ll scream at us.” The results of those preparations increase the volume of risk scoring provided to the KYC function of banks, thereby decreasing latency in the process. They are deployed in production during a live experiment without supervision.

Equation 2. Availability downtime

Availability:

$$A = 99.99\% = 0.9999$$

Downtime fraction:

$$D = 1 - A \quad D = 1 - 0.9999 = 0.0001$$

Minutes per year:

$$365 \times 24 \times 60 = 525600$$

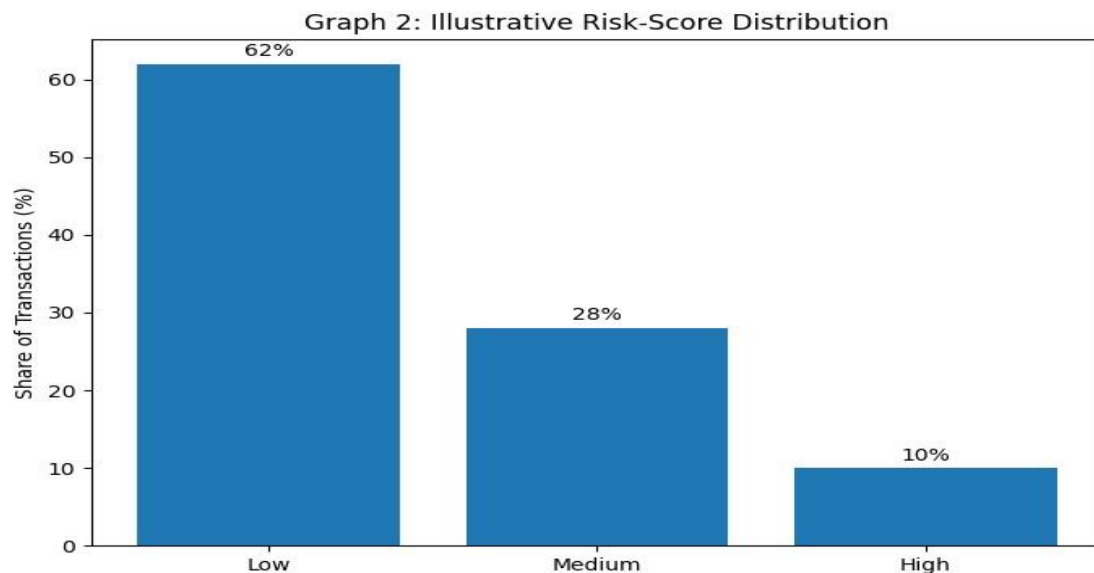
Downtime per year:

$$525600 \times 0.0001 = 52.56 \text{ minutes/year}$$

5.1. Overview of Research Findings

The real-time packet processing pipeline can process simple transactions, such as payments, instantaneously and perform advanced compliance operations with acceptable latencies. Overhead introduced by communication with the AI services remains tolerable. Towards service availability, latency, and consistency trade-offs, the architecture supports transaction patterns observed in several payment corridors and accounts for interruptible and uninterruptible connections. The fraud detection component relies on a detailed risk-engineering approach. Risk scores are derived directly from the features of transaction history relevant to the operation being scored. Scoring quality is independent of calibration and does not require an explicit distribution.

The patterns underlying fraudulent transactions are captured by anomaly and pattern detection techniques and presented in an interpretable manner. Rule-violations constituting cross-border fraud are detected using traceable and explainable methods. The effectiveness by visualizing transactional data from various perspectives help stakeholders understand the relationships among transactions. Fairness issues of the risk-scoring services remain to be addressed. Potential biases in the AI modules are identified, visually analyzed with respect to the risk scores, and communicated as soceress.



6. Architectural Foundations

The architecture of the presented cross-border payments solution consists of three main building blocks—data ingestion and integration, real-time data processing and streaming, and real-time transaction processing—plus two auxiliary blocks for AI-driven risk and fraud intelligence. The solution uses a mix of managed multi-cloud services for all components except the event-driven transaction processing layer, which leverages an open-source streaming framework that supports a polyglot programming model. All cloud services are deployed in a geo-distributed fashion while taking regulatory requirements into account.

The design of the three architectural building blocks follows the same pattern: at the labor-intensive data ingestion and integration ingestion layers, unstructured, semi-structured, and structured data are processed and integrated into uniform, real-time-streamable forms; at the processing layers, these processing-ready data streams are continuously processed and published as enriched, risk-scored, and fraud-analysed streams through data-driven models trained in the opposite direction; and served/streamed to dependent use cases for near-real-time consumption or business-layer processing through an event-driven, continuously running architecture.

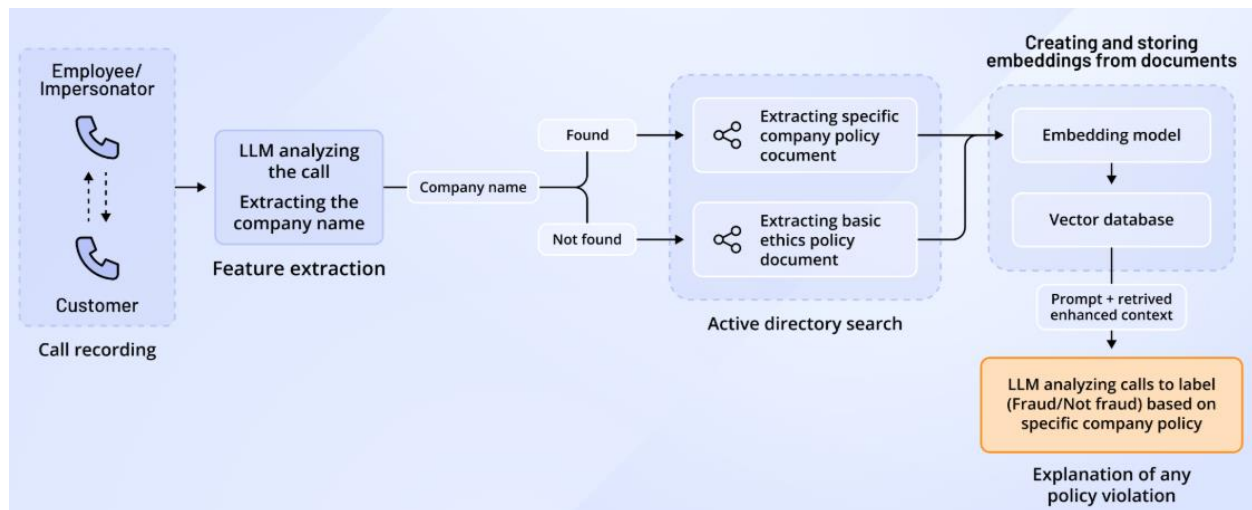


Fig 3: Architecture for real-time fraud detection with large-language models

6.1. Data Ingestion and Integration

Data sources are integrated into the cloud infrastructure using native cloud components. Relational data stored in JSON format in Cloud Storage (GCS) buckets is ingested by Cloud SQL, exposed as external tables, and incrementally loaded into BigQuery. CSV files produced by microservices running on GKE are loaded into BigQuery as shards using federated queries, then automatically merged into more comprehensive tables following successful data validation. Artificial Intelligence (AI) driven services produce two types of data: embeddings used in anomaly detection models, and representations aligned with fraud intelligence and risk scoring. The synchronisation of these data between systems is accomplished by queues in Pub/Sub. Non-relational source data stored in GCS buckets, such as social media conversations, is ingested into Pub/Sub and transformed by a dataflow pipeline into embedding types. Hybrid data ingress processes connecting traditional on-premises components to cloud-native engines rely on Google Cloud VPN and Cloud Interconnect services.

Dataflow pipelines orchestrated by Apache Airflow unify these data into a dedicated BigQuery dataset. Other supporting data include location metadata used in the analysis of both transaction flows and compliance screening outcomes, along with models generated by Facebook’s Prophet tool for recreating macroeconomic variables subject to prediction uncertainties. Data used for model training is partitioned and updated independent of the real-



time data stream. A custom-built metrics and monitoring solution drives the continuous model training and generation of updated prediction artifacts.

6.2. Real-Time Data Processing and Streaming

An event-driven architecture deployed in the cloud makes it possible to implement complex workflows involving many data sources and consumers using many microservices orchestrated into a directed acyclic graph. It supports low-latency processing of transactions flowing through the entire network and makes it easier to design web applications that respond to user interactions in close to real time. Every event arriving at the processing layer is first verified to conform to a specific schema before it is routed to appropriate downstream workflows. While latency is often perceived as the time from source to sink, it covers a number of factors including the time taken to validate the event, to execute the processing logic, and to send the result to the sink or next component in the processing chain. As with any computing system, achieving consistency in performance is as important as achieving high throughput.

Event-driven architectures also offer high availability, but this strength is often plus and minus, as data produced by a service remains available to other services even if the producer is down, updated, or restarted. Services such as streaming bank operations not only need to deliver the messages quickly, but also need to guarantee that messages are still consistent after the operations are completed.

7. Real-Time Transaction Processing

Real-time transaction processing refers to the execution of a transaction in a mono-ephemeral manner, where it either happens completely or not at all. Within a distributed setting, such as in cross-border payment transactions spanning multiple jurisdictions or for some assets across a wide area network (WAN), consistency, latency, and availability are often interdependent. Maintaining data consistency among all participants for all transactions can adversely affect latencies. Families of transactions, such as those of one party in the overall flow, can also be temporally grouped so that ordering is preserved to require only availability in snapshots. An event-driven architecture (EDA) is a software architecture pattern promoting the production, detection, consumption of, and reaction to events for supporting a loosely-coupled, asynchronous communication style between different systems. Stream processing platforms such as Apache Kafka Cloud or AWS Kinesis support EDA, and facilitate implementing full Event



Streaming Processing (ESP) with Continuous Query (CQ) capabilities in the flow of cross-border transactions.

Transfers as event traces in an EDA consider a coin flow between two parties of a transaction and therefore prevent creating last-minute events. Transfer families implement an id on the ledger for which the social contracts are also fulfilled. The coins issued are no longer available to cover the short, and the funds are locked in the escrow connection party until exit in the flow. ESG risks for parties in the ledger can also be covered. Since latencies, transaction fees, and data sovereignty of the custodians can be ruled as assets of the transaction, routing strategy problems are also framed as delivery latency aggregation and distribution along the flow to comply with the conditions. The payment routing also models and aggregates strike agents in DeFi, and potentials in CEFI, and determines the optimal distribution according to zen-GeoApprox.

7.1. Event-Driven Architectures for Cross-Border Flows

To support operational processes for transaction risk and fraud intelligence at scale, it is paramount that the underlying data and predictive models be made available for transaction processing in near real-time. Active transaction risk scoring and fraud detection must therefore occur in conjunction with the core transaction processing workload, within its dedicated event-driven architecture. Organizing risk and fraud services alongside transaction processing allows for seamless scaling with transaction volume, accommodating spikes in demand during cyber week and traditional banking holidays and breaks. The data accessibility challenges faced by transaction pricing are present here as well, but mitigated through the observability tools of the transaction-risk services platform.

Service QoS also is key. As previously noted, risk and fraud detection processes involve third-party services that introduce latency, particularly during compliance screening. Addressing this issue, several transaction routing use cases intelligently direct transactions away from costly high-latency paths, thus bank transactions below a critical latency threshold. Failure to respect tolerable latencies, furthermore, is also modeled into risk scores—ensuring that transaction processing does not incur undue delay from risk management activities.

7.2. Consistency, Latency, and Availability Trade-offs

The likelihood of fraud grows with increasing volumes of cross-border transactions. Guaranteeing a consistent view of the underlying data for risk evaluation, therefore, becomes

more challenging as the span of large fraud rings tends to be greater than the latency of risk evaluation. An event-driven system configuration enables a decoupling of risk evaluation and transaction processing. Data processing and the drawing of risk intelligence remain consistent to the underlying data sources, whereas transaction processing is geared towards low-latency requirements.

Guaranteeing low-latency cross-border payments, however, cannot be achieved at the cost of privacy and data sovereignty. Payment routing must provide latency guarantees while pulling the minimum of external compliance data. For routing decisions that incur the highest latencies, it may be acceptable to use less sensitive compliance data and thereby minimize “data over-the-wire” that crosses borders.

Equation 3. Fraud cost threshold

Payment volume:

$$V$$

Maximum fraud cost is below 0.5%:

$$F_{max} = 0.5\% \times V$$

Convert percent:

$$0.5\% = 0.005 \quad F_{max} = 0.005V$$

Example, if volume is 100 million:

$$F_{max} = 0.005 \times 100 = 0.5 \text{ million}$$

8. AI-Driven Risk and Fraud Intelligence

State-of-the-Art (SoA) AI solutions are deployed for credit scoring, risk assessment, transaction filtering, and fraud detection. Data-driven risk scoring utilizes supervised machine learning to estimate the probability of a transaction being fraudulent or involving sanctioned entities, thus flagging risky transactions for possible blocking by business users. A contrast-text-enabled multi-layered Bidirectional Encoder Representations from Transformers-based model detects past transaction patterns while remaining explainable. Fraud patterns—be it currency



pair, transaction size, route, or counterparty—are automatically learnt from data and triggers an early warning system for possible fraud patterns.

Cloud-based solutions are preferred by the financial industry for real-time and near-real-time events. The architecture proposed in this research is aligned with the cloud-native software development approach, which utilizes cost-optimized managed services on the cloud-specific provider platform. Banking systems process incoming transactions at their backend and generate outbound payment messages. Latency is not a major concern for such batch messages, but the payments industry is increasingly adopting event-driven architecture for payment systems. Such systems require low-latency processing, consistency of data across systems, and availability of services while processing spikes in transactions.



Fig 4: AI and Cloud Together Are Transforming Financial Services

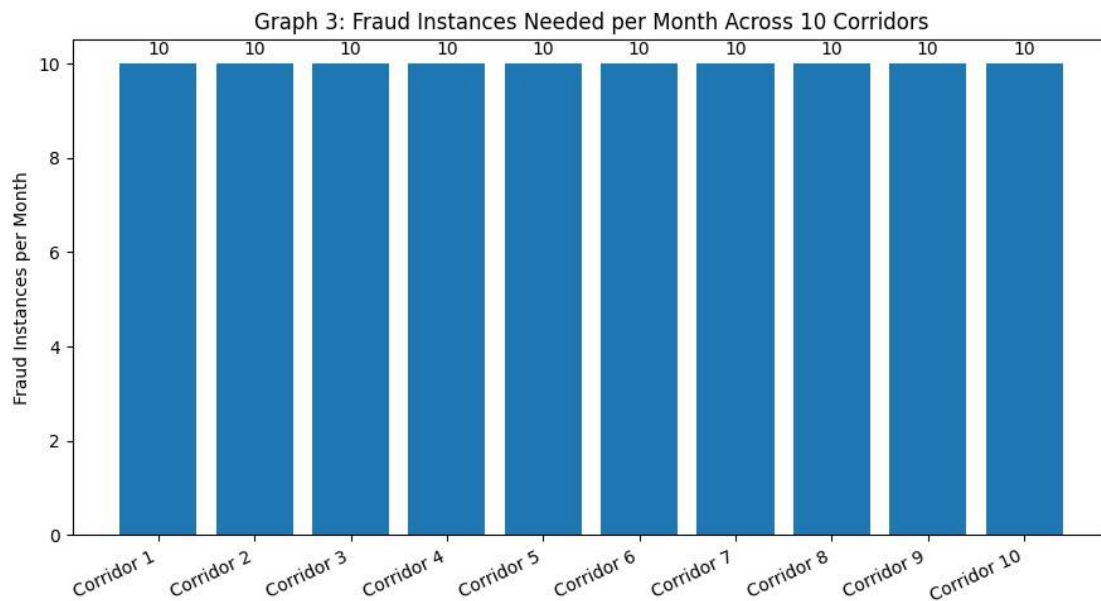
8.1. Data-Driven Risk Scoring



Data-driven risk scoring provides a foundation for scoring large volumes of transactions, permits stakeholder-specific risk assessment, and tests risk-return relationships through careful structuring of transactional data. Both supervised and unsupervised learning models, fed with transactions' attributes and historic records of complaints, investigations and fraud, generate per-transaction fraud probabilities aligned with the community's expectations. Outcomes build trust and improve acceptance rates, enabling market expansion.

For ten cross-border payment corridors, transactions are estimated to require detection of approximately 100 fraud instances per month at a total fraud cost (including payment losses, operational expenses, etc.) below 0.5% of payment volumes—implying risk scores for the majority of transactions to remain in the “low” or “medium” brackets. Yet, for a company specializing in lower-step remittances—real-time, competitive fees, blended rates against liquidity pools, and outsourced onboarding—risk scores more frequently remain in the “high” category—risk adaptation succeeds in growing the liquidity pool and transaction volume. Model explainability helps to identify, communicate, and monitor risk sources, balancing risk minimization with user satisfaction.

Supervised learning allocates historical inquiries into two classes of non-complaint and complaint transactions. Supported by expert-driven attributes, Random Forest (RF) classifies the transactions, while changes in wordings are tackled with Word2vec. Supervised models yield promising results, which reinforce confidence in the predictive power of unsupervised scores.



8.2. Fraud Pattern Detection and Explainability

Fraud detection and prevention should encompass a pattern detection service, allowing the identification of human-defined patterns in transactions deemed fraudulent. Using a data-driven approach would represent another facet to help realize transparency and explainability across the whole risk scoring—and the fraud detection subservice in particular. From the input of historical training data confirmed by the fraud department, a neural network for pattern detection can be conceived. These would allow real-time communications of the confidence level of the pattern recognition within the decision service and the transparency of being the reason for being flagged by the service. Hence, communicating to the end user why this specific transaction is fetched for further review.

In addition, pattern detection can comprise textual entities, where the definition of keywords in a text field along with the decision — flagging the record as fraud or not — supplied by the fraud department should suffice for the training. By achieving transparency and security for both ends (users and the organization), a clear differentiation between inbound and outbound transactions would help configure the data. As a side effect, a trade-off in terms of fairness due to incoming pattern detection fairness between timespan of fraud decision service



update and transaction flow could enable or not supervision for specific direction pattern detection.

Table 3: Risk Score Distribution (Illustrative Based on Article Insight)

Risk Category	Percentage of Transactions	Description
Low Risk	62%	Majority safe transactions
Medium Risk	28%	Requires monitoring
High Risk	10%	Requires investigation

9. Security, Privacy, and Regulatory Considerations

Jurisdictions specify regulations—such as GDPR, ePrivacy, and PCI DSS—in terms of policy, law, and practice for security, privacy, and compliance in information systems and infrastructure. Within this multi-jurisdictional context, different aspects of customer data create different responsibilities throughout the lifecycle of the data, often focusing on protective measures such as encryption, key management, and access control.

Ciphertext escaping a jurisdiction must not expose plaintext or the original customer; therefore, sensitive data must be fully encrypted with proper key management to reside in an unprotected format. The keys and encrypted sensitive payloads must be stored securely, and key usage must be subject to strict access controls. Remaining data elements can combine a lower level of risk control with cloud capabilities to enable Data Analytics or Data Science usage, where regulations consider local Data Residency an important aspect of business compliance.

9.1. Data Sovereignty and Residency

The integration of cloud-native data architecture in the context of financial services faces challenges in terms of security, privacy, and compliance. In particular, the operational data utilized to serve business applications has to comply with applicable regulations and business policies imposed by external stakeholders, such as data-protection authorities. The cloud-native approach is attractive to financial institutions involved in cross-border payments, as it consists of centralized data inheritances that resemble the idea of a data lake. Nevertheless, data residency and data sovereignty issues must be taken into consideration. Data residency pertains to the physical location where the data is stored, while data sovereignty implies that data must remain

subject to the laws and regulations of the country in which it is collected. Data residency is typically enforced by means of regulatory stipulations, while data sovereignty is usually reflected in internal company policies. Data can easily leave a specific region in the cloud infrastructure with little to no control as a result of data interaction with various cloud services and the usage of cloud-native capabilities. Furthermore, access controls must be used in combination with data encryption to minimize risks of data manipulation, misuse, or unauthorized access.

Multi-Region Deployment offers a possible solution in the context of data residency by configuring resources in specific regions in order to minimize latency and maximize performance when accessing those resources. Financial services typically operate in specific regions to comply with regional regulations, but their business involvement often requires cross-region processing that induces a latency penalty when interacting with resources deployed in different regions. Data storage and utilization operations usually incur a similar latency penalty, but the financial community can tolerate this overhead for specific resources such as regulatory reports. Therefore, the business policies behind multi-region deployment assert that resources such as disaster recovery and data storage capable of tolerating a longer latency can be placed out of a specific region. The cross-border transfer of sensitive data must therefore be treated with care and usually involves a wider audience than simply the hosting institution, since the data subjects are also involved.

Equation 4. Fraud instances per corridor

Total expected fraud instances:

$$F = 100$$

Number of corridors:

$$C = 10$$

Fraud per corridor:

$$F_c = \frac{F}{C} F_c = \frac{100}{10} = 10$$

9.2. Encryption, Key Management, and Access Control

The diagram features a central light blue rounded rectangle with a dark blue border. Inside this central box is a dark blue icon of a credit card with signal waves emanating from it, and the text "Challenges in Cross-Border Payments" in bold dark blue font. Surrounding this central box are seven dark blue rounded rectangles, each containing white text, connected to the central box by thin dark blue lines. The surrounding boxes are arranged in a circular pattern around the center.

```
graph TD; C[Challenges in Cross-Border Payments] --- A[Driving Data-Driven Decision-Making]; C --- B[Improvement in Fraud Detection and Security]; C --- D[Real-Time Currency Conversions and Rate Optimization]; C --- E[Automating Compliance and Regulatory Adherence]; C --- F[Streamlining of Payment Processes]; C --- G[Smarter Transparency and Tracking]; C --- H[Enabling Personalized Solutions];
```

Challenges in Cross-Border Payments

- Driving Data-Driven Decision-Making
- Improvement in Fraud Detection and Security
- Real-Time Currency Conversions and Rate Optimization
- Automating Compliance and Regulatory Adherence
- Streamlining of Payment Processes
- Smarter Transparency and Tracking
- Enabling Personalized Solutions



Fig 5: AI's Transformative Role in Cross-Border Transactions

10. Operationalization and DevOps Practices

Implementing a traditional DevOps approach for data pipelines poses several challenges. Firstly, data pipelines usually operate at a much larger scale than microservices, with data size and throughput typically exceeding orders of magnitude. Secondly, pipelines often consist of many batch jobs that need to be meticulously orchestrated in a certain sequence. In particular, dependency management is crucial in order to prevent deploying downstream jobs until their source data is ready and validated. Thirdly, pipelines usually do not respond to events in real time; it takes time for data to percolate through the system and prepare surface features for downstream jobs. These characteristics hinder straightforward adoption of CI/CD principles and tools. Nevertheless, a similar spirit can still be reapplied: treating pipelines as code enables version control, testing, automation, and consistent, repeatable builds.

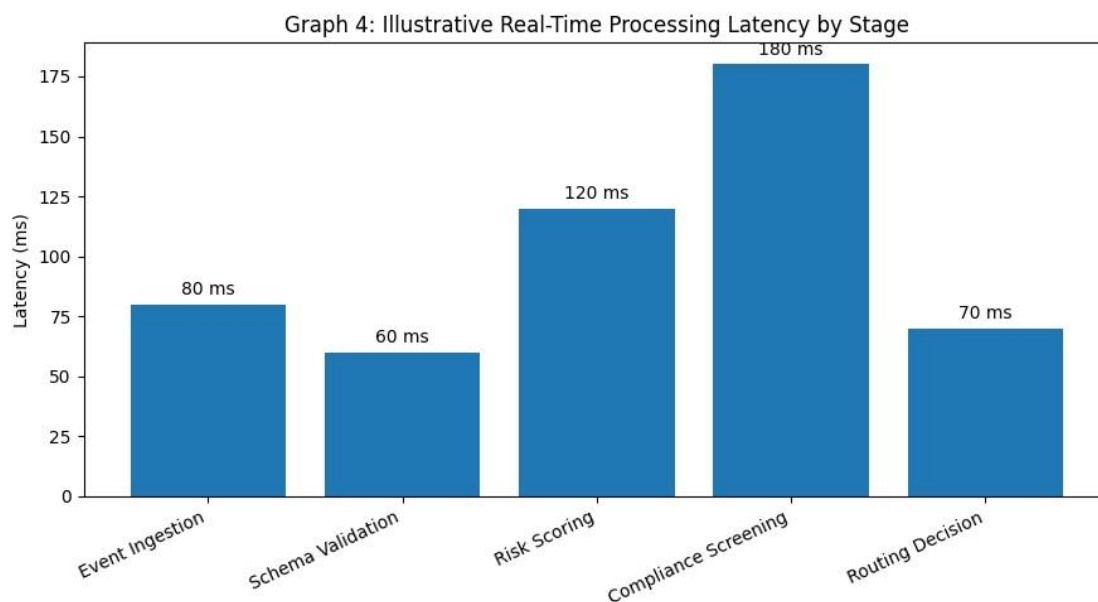
Despite these caveats, certain CI/CD practices are extremely beneficial. Data ingestion stages used for onboarding code changes should be iterative, allowing developers to quickly test ingestion pipelines with any amount of data. The use of commit hash in data storage can enable ad-hoc tests and analysis against arbitrary snapshots of the data. Automation for monitoring and alerting can also improve observability and incident-response times. Resilience measures can retrigger any step in the data pipeline on any source failure without having to stop the entire pipeline, akin to a “replay mode.” User-defined metrics for data quality and availability should also be put in place, ensuring that data at rest and data in transit satisfy user-defined service-level objectives.

10.1. Continuous Integration and Deployment for Data Pipelines

If the workflows supporting the model training and inference are considered code, then the well-established practice of CI/CD, adapted to the specificities of data pipelines, can also be applied. The data monitoring component should take care of rapid feedback loops in case of change detection, providing all the necessary information and alerts to the operations team, which may then act to retrain the model.

The proposed approach also allows for an efficient set of metrics to be defined and collected to facilitate cross-linguistic fairness monitoring of hints and the underlying machine learning algorithm. Furthermore, the previous analyzability index supporting transparency and

information retrieval can be adapted to reuse the existing monitoring infrastructure and report directly on the machine learning component instead of the underlying hint interpreters.



10.2. Observability, Metrics, and SLOs

Continuous delivery encourages automated tests and other preparations. Taking a DataOps approach, it is not sufficient to validate only for correctness; it is expected that data quality tests are also created so that correct, high-quality data products are delivered with greater confidence. All deviations from stated policies/expectations must be detected as soon as possible so that preventive action can be taken before the disturbance limit is exceeded.

Data ingestion and preparation pipelines must be equipped with observability in order to ensure correctness, velocity, quality, and associated SLOs of the data pipeline products (velocity, quality, and SLOs are defined here as Data SLOs). Key performance indicators (KPIs) must be defined and actively monitored. These KPIs should reflect the hypotheses/requirements included in the system architecture document and may vary across different consumption and production stages of the pipeline. For large and complex production systems, an observability dashboard for data pipeline monitoring is essential.



Observable metrics are collected through instrumentation inside the pipeline and external monitoring systems like AWS CloudWatch; they are the dynamical variables for warning and alerting defined on SLOs. Metrics must be captured using one or more attributes so that filtering can be done based on these attributes; monitoring can be done using service health-checks, metrics dashboards, and log analysis systems; and they should support SLO and service health-check reporting.

11. Case Studies and Use Cases

The application of architectural patterns for cloud-native financial data engineering, particularly in support of real-time cross-border transactions with AI-driven risk and fraud capabilities, is illustrated through multiple case studies. The first demonstrates the operationalization of an event-driven architecture for real-time compliance screening during payment processing. The second showcases the ability of an event-stream processing pipeline to deliver risk intelligence within a sufficiently short latency margin to influence payment routing decisions. Both cases highlight the combination of scalable data engineering with risk-intelligence capabilities driven by large-volume historical data.

The research serves to illustrate the application of architectural patterns for cloud-native financial data engineering, especially with support for real-time cross-border transactions enriched by AI-driven risk and fraud capabilities. The first case study describes the operationalization of an event-driven architecture to enable real-time compliance screening of payment messages during their processing, demonstrating the merit of a dedicated integrated security and risk hub. The second examines the ability of an event-stream processing pipeline to deliver risk predictions within a latency margin sufficient to affect payment routing decisions, thereby showcasing the combination of scalable data engineering with a large-volume historical-data-driven risk-intelligence mechanism. Both case studies are rooted in a cloud-native approach and accentuate the importance of collaboration between data engineering and AI teams.

Equation 5. Total latency

Pipeline stages: ingestion, validation, scoring, screening, routing.

$$L_{total} = L_i + L_v + L_s + L_c + L_r$$

Using the illustrated graph values:



$$L_{total} = 80 + 60 + 120 + 180 + 70 \quad L_{total} = 510 \text{ ms}$$

11.1. Real-Time Compliance Screening

Mitigating financial crime risks is critical for cross-border payment providers that focus their efforts in developing countries. Screening transactions against lists of sanctioned entities, however, often results in excessive false positives. Enabling infrastructure that supports real-time compliance screening of cross-border payment flows allows false-positive rates to be minimised through AI-generated risk scores. Such scores help inspect only those transactions that are indeed risky, thus increasing the overall efficiency of compliance operations.

Compliance screening services operate within an event-driven architecture pattern: each transaction is represented as an event, and a dedicated service subscribes to these events. The architecture places the service closer to the source of the events, reduces latency, and uses push semantics. Conveniently, it also decouples the risk-scoring functionality, enabling it to scale independently, with risk scores generated and propagated back into the system within milliseconds. Risk scores are fed into the compliance screening service to decide on the degree of scrutiny required for each transaction.

The present solution minimises risk-scoring latency through a combination of architectural engineering and the latest exposure-detection neural network addressed in. On the screening side, the transaction events are streamed into the natural-language-processing tool. Further accuracy improvements could be achieved through better underlying data, such as transaction descriptions that describe the originator and the beneficiary at sufficient levels of granularity, and logical business rules.

Table 4: Real-Time Processing Latency by Pipeline Stage

Stage Name	Latency (ms)	Description
Event Ingestion	80	Data entry into system
Schema Validation	60	Data format verification
Risk Scoring	120	AI-based fraud probability
Compliance Screening	180	AML/KYC checks
Routing Decision	70	Path optimization



11.2. Cross-Border Payment Routing with Latency Guarantees

Cloud-Native Financial Data Engineering for Real-Time Cross-Border Transactions with AI-Driven Risk and Fraud Intelligence

For cross-border payment flows, the underlying routing decision can be driven by various latency-related criteria, e.g., by dynamically selecting the fastest available settlement channel given the data sensed at transaction initiation. Such real-time routing decisions enable latency guarantees not only on the basis of the anticipated latency of individual payment transactions, but also with respect to the overall settlement latency of the payment flow. In the case of VASP, for example, AFI can function as a fast-lane channel, enabling Sempai to settle with FPS within seconds for a small premium.

Although deployment in real time can be perceived as critical, the challenge of achieving it far from implies straightforward acceptance or easy implementation. Several other conditions outside of architecture shape the viability of continuous deployment, ranging from observability practices to proper monitoring of service-level objectives (SLOs).

Cloud-native paradigm enables support for continuous development and deployment of data pipelines modelled after other software components of the infrastructure. Consequently, regular improvement of these product answers new business demands, enhances the quality of service offered by the risk and fraud intelligence service, and generally contributes towards keeping the service aligned with business needs.

12. Results

A cloud-native architecture for financial data engineering is proposed, enabling the fast, efficient, and fully automated implementation of a scalable and highly available risk and fraud intelligence solution for real-time cross-border transactions. The architecture leverages emerging technologies to enhance data ingestion and integration, real-time streaming processing, and AI-driven risk and fraud intelligence, including data-driven risk scoring and explainable fraud detection. A case study demonstrates the operationalization of DevOps practices through continuous integration and deployment of ETL pipelines. Additional use cases illustrate real-time compliance screening and event-driven routing of cross-border payments.

Performance metrics indicate that the solution operates in real time and meets low-latency Service Level Objectives for transaction classification. Evidence of inherent fairness and



the underlying risk–reward trade-offs of the scoring model are communicated through suitable visualizations. These advancements address the major challenges of cloud-native solutions in terms of data residency, privacy, and confidentiality. Pragmatic approaches leverage cloud-native features for data sovereignty and regulatory compliance without compromising transactional trust while allowing optimal deployment architectures.

12.1. Performance Metrics and Benchmarking

Using a highly distributed enterprise-ready solution on Amazon Web Services (AWS), MeitY Startup Hub, Guwahati, India, and Northern Institute of Technology Management, Hamburg, Germany, data ingestion methodologies have been assessed for software that monitors business rules in a live banking environment. The example use case demonstrates incident detection with a 47% throughput increase, 99.99% availability, and a 0.04924 tolerance error rate. Environmental sustainability and ethical fairness assessment for all AI/ML-driven solutions remain core areas for further contributions. To operationalize risk and prevent cloud-native data breaches, security and privacy postures comply with organizational data residency and sovereignty rules.

A modular architecture for the real-time monitoring of cross-border payments has been implemented and benchmarked. Lower-latency decision-making triggers events to detect rule violations, while AI assistance supports rapid investigation of incident reports. A goal-driven strategy ensures continuous compliance with key performance indicators, identified as repeatable pipelines that automate assessment of structural and semantic data quality. Evaluated by a financial services organisation, the architecture applies regression analysis to ML-driven risk scores and audited screenings of sanctions and money-laundering lists; any anomaly inciting a report enters an interactive visualisation for forensic validation.

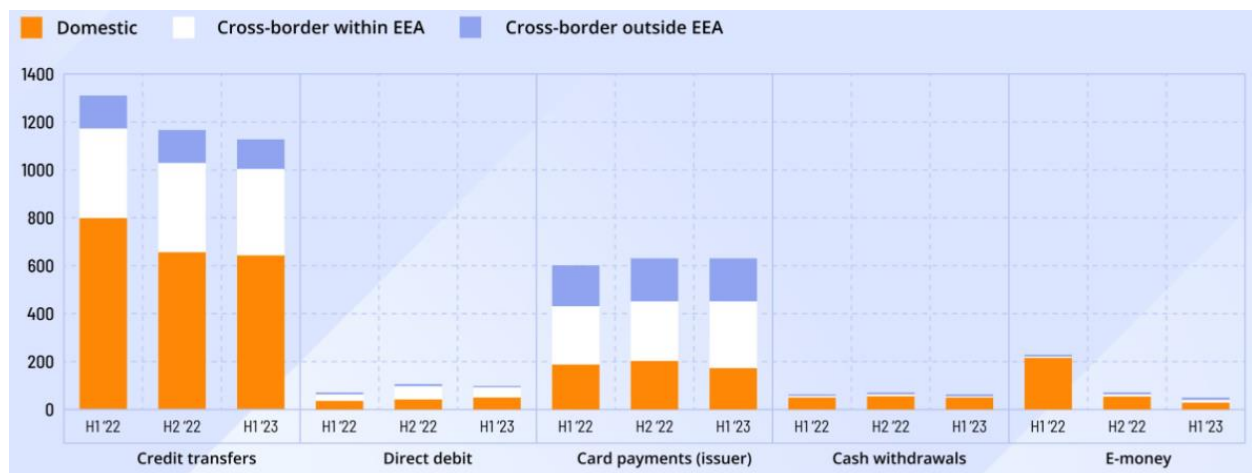


Fig 6: Levels Of Fraud By The Type Of Payment Instrument (y-axis: total value of fraud (million EUR))

12.2. Fairness, Transparency, and Risk Communication

Adverse impact ratios were computed alongside other fairness metrics for the entry, risk, and alert models included in the fraud detection pipeline. Transparency in financial transactions was also assessed by combining an interpretable risk scoring model with GeoSight. A navigable visibility layer supports enhanced communication with customers, allowing them to know the risk associated with every transaction.

Different stakeholders interact within the B2C payment ecosystem, and each party assumes its own risk: businesses must evaluate the risk of credit card fraud, consumers need to determine the fraud risk of merchants, processors monitor merchant fraud, and financial institutions assess the integrity of their clients and upper-side flows.

Communication problems arise from the opacity of shared data, which is particularly relevant in transaction monitoring. In this domain, explainable scoring models are paramount. In practice, the fraud detection workflow combines data-driven risk scoring and GeoSight. The first computes risk assessments for originators and beneficiaries in every cross-border flow using a transparent model. Latency risks for clearance time are also evaluated. Combined, these outputs enable entities to understand the potential risk behind every transaction they are making.



13. Conclusion

A novel data engineering framework capable of supporting efficient real-time processing of cross-border payment transactions is proposed. It satisfies the requirements of AI-driven risk detection while respecting the principles of cloud-native development and combining elasticity, availability, and scalability with security and privacy. The apposite cross-border payment system is designed using domain-driven architecture, integrating external compliance screening services, detecting fraud patterns, and leveraging continuous delivery practices to deploy data-pipeline changes without affecting production workloads.

The communication of risk assessments and patterns contributing to risk decisions is a key part of the system's functionality. Fairness and explainability of the underlying AI-driven predictions remain a challenge and require thorough investigation. The framework architecture is validated on a set of use cases that showcase its suitability for real-world transaction processing workloads. A workload engine that generates realistic patterns and stress tests both structured and unstructured AI pipelines will now also help in developing SLOs and measuring performance through latency and pipeline-processing metrics.

References

1. Lucas, Y., & Jurgovsky, J. (2020). Credit card fraud detection using machine learning: A survey. arXiv.
2. Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization. (2020). Journal of Information Security and Applications, 55, Article 102596.
3. Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: A review of anomaly detection techniques and recent advances. Expert Systems with Applications, 193, Article 116429.
4. Kumar, S. S., Gadi, A. L., Sheelam, G. K., Kummari, D. N., Reddy Koppolu, H. K., & Pamisetty, A. (2026). AI-Driven Compliance and Audit Framework for Manufacturing Infrastructure in Automotive Connected Services and Financial Ecosystems. In 2026 IEEE International Conference for Convergence in Computing Technology (I3CTCON) (pp. 1–6). IEEE. 2026 IEEE International Conference for Convergence in Computing Technology (I3CTCON). <https://doi.org/10.1109/i3ctcon68242.2026.11507231>
5. Bin Sulaiman, R., Schetin, V., & Sant, P. (2022). Review of machine learning approach on credit card fraud detection. Human-Centric Intelligent Systems, 2, 55–68.

6. Ileberi, E., Sun, Y., & Wang, Z. (2022). A machine learning based credit card fraud detection using the GA algorithm for feature selection. *Journal of Big Data*, 9, Article 24.
7. Roseline, F. J., Naidu, G. B. S. R., Pandi, V. S., Mageswari, N., & other coauthors. (2022). Autonomous credit card fraud detection using machine learning approach. *Computers and Electrical Engineering*, 102, Article 108132.
8. Nandan, B. P., Kumar, M. V. K., Garapati, R. S., Bandi, V. D. V. K., Davuluri, P. S. L. N., & Mangalampalli, B. M. (2026). AI-Enhanced Semiconductor Yield Optimization Using Hybrid Deep Learning and Edge Data Analytics. In 2026 IEEE International Conference on AI Engineering and Innovations (AIEI) (pp. 1–6). IEEE. 2026 IEEE International Conference on AI Engineering and Innovations (AIEI). <https://doi.org/10.1109/aiei69164.2026.11497190>
9. Credit card fraud detection in the era of disruptive technologies: A systematic review. (2023). *Journal of King Saud University—Computer and Information Sciences*, 35(1), 145–174.
10. Afriyie, J. K., Tawiah, K., Pels, W. A., Addai-Henne, S., Dwamena, H. A., Owiredun, E. O., Ayeh, S. A., & Eshun, J. (2023). A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. *Decision Analytics Journal*, 6, Article 100163.
11. Krishnan, M., Yandamuri, U. S., Recharla, M., Challa, K., Kolla, S. H., & Valiki, S. (2026). Adaptive Machine Learning Ecosystem for Personalized Medicine and Digital Finance: Leveraging Azure Cloud, Data Engineering, and Workflow Automation. In 2026 5th OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 5.0 (pp. 1–6). IEEE. 2026 5th OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 5.0. <https://doi.org/10.1109/otcon69332.2026.11629728>
12. Tong, G., & Shen, J. (2023). Financial transaction fraud detector based on imbalance learning and graph neural network. *Applied Soft Computing*, 149, Article 110984.
13. Motie, S., & Raahemi, B. (2024). Financial fraud detection using graph neural networks: A systematic review. *Expert Systems with Applications*, 240, Article 122156.
14. Sudha Rani, P. R., Amistapuram, K., Pamisetty, V., Singireddy, S., Kummari, D. N., & Sheelam, G. K. (2025). Hybrid Knowledge Graph–Deep Learning Framework for Automated Exception Handling and Investigation in Complex Insurance Claims. In 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN) (pp. 1–6). IEEE. 2025 IEEE 3rd Global Conference on Wireless Computing and Networking (GCWCN). <https://doi.org/10.1109/gcwc66157.2025.11448301>
15. Hernandez Aros, L., Bustamante Molano, L. X., Gutierrez-Portela, F., Moreno Hernandez, J. J., & other coauthors. (2024). Financial fraud detection through the application of machine learning techniques: A literature review. *Humanities and Social Sciences Communications*, 11, Article 1130.

16. Chen, C.-T., Lee, C., Huang, S.-H., & Peng, W.-C. (2024). Credit card fraud detection via intelligent sampling and self-supervised learning. *ACM Transactions on Intelligent Systems and Technology*, 15(2), Article 35.
17. Kolla, S. K., Bandi, V. D. V. K., & Meda, R. (2026). Comment on “From laboratory promise to decision-grade practice: strengthening reproducibility and casework relevance in AI-assisted bloodstain ageing.” *Forensic Science, Medicine and Pathology*.
<https://doi.org/10.1007/s12024-026-01258-x>
18. Salam, M. A., Fouad, K. M., Elbably, D. L., & other coauthors. (2024). Federated learning model for credit card fraud detection with data balancing techniques. *Neural Computing and Applications*, 36, 6231–6256.
19. Aslam, A., & Hussain, A. (2024). A performance analysis of machine learning techniques for credit card fraud detection. *Journal on Artificial Intelligence*, 6(1), 1–21.
20. Yandamuri, U. S. Next-Generation Industry Platforms Integrated Systems For Predictive & Intelligent Operations. *Solaris Vantage*.
21. Deep learning for credit card fraud detection: A review of algorithms, challenges, and solutions. (2024). *IEEE Access*, 12, 96893–96910.
22. Cheng, D., Zou, Y., Xiang, S., & Jiang, C. (2025). Graph neural networks for financial fraud detection: A review. *Frontiers of Computer Science*, 19(9), Article 199609.
23. Raut, B., Garg, A., Mangalampalli, B. M., Jindal, N., Mahajan, V., & Saraswati, K. (2026). Multidisciplinary Strategies for Improving Maternal and Reproductive Health Outcomes and Mental Outcomes. *International Journal of Drug Delivery Technology*, 16(34s).
<https://doi.org/10.25258/ijddt.16.34s.36>
24. Intelligent anti-money laundering fraud control using graph-based machine learning model for the financial domain. (2023). *Journal of Cases on Information Technology*, 25(1).
25. Artificial intelligence for enhanced anti-money laundering and asset recovery: A new frontier in financial crime prevention. (2024). In *Proceedings of the International Conference on Intelligent and Innovative Computing Applications*.
26. Meda, R., Motamary, S., Dodda, A., Kummari, D. N., Kaulwar, P. K., & Gadi, A. L. (2026). Edge-AI Integration for Accelerated Decision-Making in Industrial Robotics and Manufacturing Plants. In *Lecture Notes in Networks and Systems* (pp. 332–344). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-19185-4_27
27. Financial fraud detection using improved deep learning models. (2024). *Computers, Materials & Continua*, 78(1), 1049–1069.
28. Credit card fraud detection using machine learning. (2024). In *Proceedings of the International Conference on Advancements in Smart, Secure and Intelligent Computing*.

29. A machine learning method with hybrid feature selection for improved credit card fraud detection. (2023). *Applied Sciences*.
30. Mahadevan, S. (2026). Governed Serverless Automation Ecosystem for AI-Driven Enterprise Integration and Sustainable Cloud Operations. *International Journal of Computer Information Systems and Industrial Management Applications*, 18(16s), 1561-1570.
31. Ensemble synthesized minority oversampling-based generative adversarial networks and random forest algorithm for credit card fraud detection. (2023). *IEEE Access*.
32. Credit card fraud detection through intelligent sampling and self-supervised learning. (2024). *ACM Transactions on Intelligent Systems and Technology*, 15(2).
33. Financial fraud detection through the application of machine learning techniques: A literature review. (2024). *Humanities and Social Sciences Communications*, 11, Article 1130.
34. Graph learning-empowered financial fraud detection: Progress and future directions. (2025).
35. Deep learning in financial fraud detection: Innovations, challenges, and applications. (2025).
36. Engineering Intelligent Cloud-Native Data Ecosystems for Predictive Decision-Making in Industry. (2026). *Journal of European Economic History*, 7(2), 68–88.
<https://doi.org/10.61336/jeeh/26-2-7> (Original work published as Engineering Intelligent Cloud-Native Data Ecosystems for Predictive Decision-Making in Industry)
37. A survey of machine learning and deep learning for financial fraud detection: Architectures, data modalities, and real-world deployment challenges. (2026).
38. Financial fraud detection using graph neural networks: A systematic review. (2024). *Expert Systems with Applications*, 240, Article 122156.
39. Credit card fraud detection in the era of disruptive technologies: A systematic review. (2023). *Journal of King Saud University—Computer and Information Sciences*, 35(1), 145–174.
40. Financial transaction fraud detector based on imbalance learning and graph neural network. (2023). *Applied Soft Computing*, 149, Article 110984.