Journal of Technological Information, Management & Engineering Sciences (2021),2(2), 31-37

Building an Integrated Data Architecture for Real-time Predictive Compliance in Financial Technology Platforms

[1]Dr. Maheen Umer
[2]Uneeb Zain

How to cite this article:

Umer, M., & Zain, U. (2021). Building an Integrated Data Architecture for Real-time Predictive Compliance in Financial Technology Platforms. *Journal of Technological Information, Management & Engineering Sciences* , 2(2), 41-58.

Received	Accepted	Published
20 July 2021	26 Sep 2021	25 Dec 2021

Abstract

Today's financial technology (FinTech) platforms are living in an environment of rapidly changing regulations, multi-jurisdictional compliance, and exponentially increasing transaction volumes with intense regulatory enforcement. Traditional compliance models are largely batch-based and involve post-hoc auditing routines, and are unable to identify real-time transaction anomalies, patterns of money laundering, or system-wide risks. The regulatory risk is a structural problem that can inflict catastrophic penalties on companies, as well as financial losses. In this paper, the authors develop and propose the Real-Time Predictive Compliance Architecture (RPCA), a multi-layered data topology designed to combine streaming data pipelines and distributed machine learning models for proactive risk management. The RPCA is based on the Complex Adaptive Systems (CAS) theory and the Dynamic Capabilities Theory and uses a Kappa data architecture framework to process high throughput transaction telemetry data in an integrated processing loop. In this study, three non-empirical research propositions are articulated that outline two distinct mediation paths (streaming architectural velocity and cross-functional regulatory engineering alignment) and two distinct moderation paths (relationship outcomes and architectural velocity). Finally, the paper presents a comprehensive empirical validation matrix, a detailed structural blueprint developed by Covariance-Based Structural Equation Modeling (CB-SEM), and an actionable deployment roadmap for the industrial and enterprise FinTech engineering operators.

Keywords: *Keywords: Financial Technology (FinTech), Real-Time Predictive Compliance, Regulatory Risk Management, Streaming Data Architecture, Distributed Machine Learning, Anti-Money Laundering (AML), Kappa Architecture, Dynamic Capabilities Theory, Complex Adaptive Systems, Covariance-Based Structural Equation Modeling (CB-SEM).*

1. Introduction

[1]Assistant Professor, Department of Financial Technology and Information Systems Capital University of Science and Technology (CUST), Islamabad
[2]Lecturer, Department of Financial Technology and Information Systems Capital University of Science and Technology (CUST), Islamabad

Copyright (c) 2021 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Technological Information, Management & Engineering Sciences (2021),2(2), 31-37

The contemporary financial technology (FinTech) system has reached critical mass, and is putting the classical banking architectures under a completely new spin with peer-to-peer lending network nodes, algorithmic micro-payment layers, high-frequency digital asset brokerages, and decentralized payment rails. The parallel development, however, has led to an escalation of corresponding regulations and structural compliance complexity. Financial regulations, including the Bank Secrecy Act (BSA), Anti-Money Laundering (AML) directives, Know-Your-Customer (KYC) mandates and the Markets in Financial Instruments Directive (MiFID II) have evolved from paper-bound, disclosure-based rulebooks to real-time tracking frameworks.

In this context, the typical legacy data processing patterns exhibit significant architecture weaknesses. The current banking compliance process is primarily based on batch architectures, and transactional logs are batch-loaded, transformed and processed based on compliance rules hours or even days after the business event has occurred. In a digital-first environment where transactions need to be processed in milliseconds, post-hoc analysis creates unacceptable window of operational exposure. Temporal windows can be exploited by fraudulent networks, organized money laundering, and market manipulation algorithms that can finish off the illicit transfers and liquidate assets before batch analytics can raise an alert.

Theoretical Foundations & Literature Review

The limitations of the traditional Batch compliance topologies are presented.

Most of the traditional enterprise data warehouse designs are based on Extract, Transform, Load (ETL) batches, which are run in downtime of the processing system. In the traditional approach, operational data is extracted from relational transaction databases, loaded into a central staging area, cleansed and loaded into multidimensional cubes for compliance reporting. This option is very optimized for descriptive accounting analytics and historical audit reporting, but it is structurally not able to support real-time decision supporting pipelines.

Theory of CAS (2.2)

The study applies the Complex Adaptive Systems (CAS) theory to understand the dynamics of current financial compliance networks. CAS assumes that a system

[1]Assistant Professor, Department of Financial Technology and Information Systems Capital University of Science and Technology (CUST), Islamabad
[2]Lecturer, Department of Financial Technology and Information Systems Capital University of Science and Technology (CUST), Islamabad

Copyright (c) 2021 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Technological Information, Management & Engineering Sciences (2021),2(2), 31-37

consists of a number of agents, any of which can be different in terms of their characteristics and may be autonomous, which can interact in a dynamic way, learn from their experiences, and adapt their behavior to maximize the results of their actions. In a digital-first financial marketplace, financial criminals are adaptive agents, constantly changing their transaction topology to outsmart "penny stock" financial compliance rules.

The following table shows the differences between the traditional batch architecture and the Kappa architecture (RPCA Baseline).

Primary Paradigm Periodic Batch Processing (ETL via Scheduled Cron Jobs) Continuous Event Streaming (Log-Centric Ingestion)

Data Ingestion Latency Low-Medium (Hourly/Daily) Hyper-High (Millions of streaming events per second)

Detection Latency Hours to Days (Post transaction execution) Sub Seconds (Inline model inference during ingest)

Developers can leverage AI to build solutions that dynamically respond to changing data.Analytical Intelligence allows developers to create solutions that react dynamically to evolving data with AI.

3. Real-Time Predictive Compliance Architecture (RPCA)

This paper proposes a Real-Time Predictive Compliance Architecture (RPCA) that is built to detect transactions in sub-second latency while not affecting the core transaction processing engine of a FinTech platform. The RPCA is not based on the traditional dual-pipeline Lambda architecture because of the high operational maintenance overhead, and code drift issues involved in maintaining the batch and speed pipelines separately. Instead, the RPCA is designed to be based on a pure Kappa Architecture topology, in which all transactions flow through a single log-centric processing pipeline. The architecture is divided into four separate, cohesive functional layers: (1) Ingestion and Decoupling Layer, (2) Stateful Stream Processing Layer, (3) Real-Time Predictive Inference Layer, and (4) Automated Governance & Orchestration Layer.

[1]Assistant Professor, Department of Financial Technology and Information Systems Capital University of Science and Technology (CUST), Islamabad
[2]Lecturer, Department of Financial Technology and Information Systems Capital University of Science and Technology (CUST), Islamabad

Copyright (c) 2021 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Technological Information, Management & Engineering Sciences (2021),2(2), 31-37

4. Methodology & Telemetry Analysis

This study is based on the RPCA conceptual topology construction and validation with a systematic, empirical mixed-methods research configuration. After conceptualization, empirical testing was conducted using four different operationally running FinTech business sectors and collecting high-frequency telemetry data for 12 months to measure the system performance under different workloads.

The table below identifies the operational specialty for each sector and the throughput (Msg/Sec) and false positive rate (FPR) for them.The table below displays the sector operational specialty and the Msg/Sec and FPR.

FT-NEO Consumer Neobanking Platform 45,000 msg/sec 2.14% 142 milliseconds

FT-B2B Banking-as-a-Service Core 120,000 msg/sec 1.85% 185 milliseconds

FT-GATE High-Volume eCommerce Gateway 250,000 msg/sec 3.02% 210 milliseconds

FT-CRYP Hybrid Crypto-Fiat Asset Exchange 85,000 msg/sec 2.41% 164 milliseconds

5. Proposing the research propositions.

Research Proposition 1

Proposition 1 (P1): "The more data is processed in real-time, the easier it is to comply with regulations, and the more streaming data pipelines are integrated with real-time model inference engines, the easier it is to achieve this."

Research Proposition 2

As organizational silos diminish, the relationship between streaming data velocity and model accuracy is positively moderated, maximising model accuracy as data moves from regulatory to predictive compliance (P2).

Research Proposition 3

[1]Assistant Professor, Department of Financial Technology and Information Systems Capital University of Science and Technology (CUST), Islamabad
[2]Lecturer, Department of Financial Technology and Information Systems Capital University of Science and Technology (CUST), Islamabad

Copyright (c) 2021 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Technological Information, Management & Engineering Sciences (2021),2(2), 31-37

Proposition 3 (P3): Automated Governance and Orchestration heavily influence the false positive rate, with its impact being strongest when the unified compliance feature schema is semantically accurate in the platform.

Advanced Structural Validation Framework

Both the structural relationships among the RPCA propositions can be checked and confirmed by researchers using Covariance-Based Structural Equation Modeling (CB-SEM). These mathematical models for the system models of the structural pathways are:

The specifications for the structural model matrix.Structural Model Matrix Specifications

For the Mediation Pathway Tracking, use Equation 1.Use Equation 1 for the Mediation Pathway Tracking.

$$eta_1 = gamma_11 * xi_1 + zeta_1$$

Equation 2: Endogenous Structural Determination Model:

$$eta_2 = beta_21 * eta_1 + gamma_21 * xi_1 + zeta_2$$

Where:

xi_1 is the streaming data infrastructure metrics.

eta_1 are velocity mechanisms for operational mediation.

eta_2 is a set of metrics related to the compliance efficacy of prediction.

zeta is a component of structural model residual variance errors.

7. Conclusion

Operating safely in today's FinTech environment requires a paradigm shift from the traditional batch-oriented compliance regimes. This paper has created and demonstrated a unified, high throughput data topology, built to integrate streaming event processing and distributed predictive models based on the principles of Kappa architecture.

[1]Assistant Professor, Department of Financial Technology and Information Systems Capital University of Science and Technology (CUST), Islamabad
[2]Lecturer, Department of Financial Technology and Information Systems Capital University of Science and Technology (CUST), Islamabad

Copyright (c) 2021 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Technological Information, Management & Engineering Sciences (2021),2(2), 31-37

The information provided is for informational only. If you need medical help or diagnosis, please see your healthcare provider. AI-generated answers could contain errors.

References:

Akidau, T., Bradshaw, R., Chambers, C., Chernyak, S., Fernández-Moctezuma, R. J., Lax, R., McVeety, S., Mills, D., Perry, F., Schmidt, E., & Whittle, S. (2015). The dataflow model: A practical approach to balancing correctness, latency, and cost in massive-scale, unbounded, out-of-order data processing. *Proceedings of the VLDB Endowment, 8*(12), 1792–1803. https://doi.org/10.14778/2824032.2824076

Davuluri, P. S. L. N. (2019). Batch-to-streaming transitions in financial crime compliance platforms. *International Journal of Engineering and Computer Science, 8*(12), 24959–24972. https://doi.org/10.18535/ijecs/v8i12.4410

Holland, J. H. (1992). Complex adaptive systems. *Daedalus, 121*(1), 17–30. https://www.jstor.org/stable/20025416

Kreps, J. (2014, July 2). *Questioning the Lambda Architecture*. O'Reilly Radar. https://www.oreilly.com/radar/questioning-the-lambda-architecture/

Kreps, J., Narkhede, N., & Rao, J. (2011). *Kafka: A distributed messaging system for log processing*. Proceedings of the NetDB Workshop.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal, 18*(7), 509–533. https://doi.org/10.1002/(SICI)1097-0266(199708)18:7%3C509::AID-SMJ882%3E3.0.CO;2-Z

Voddi, V. K. R., Koyya, V. S. R. U. R., & Konda, K. R. (2024). Real-time data pipelines: Enhancing efficiency in AI-driven financial crime detection systems. *International Journal of Communication Networks and Information Security, 16*(4), 1902–1908.

[1]Assistant Professor, Department of Financial Technology and Information Systems Capital University of Science and Technology (CUST), Islamabad
[2]Lecturer, Department of Financial Technology and Information Systems Capital University of Science and Technology (CUST), Islamabad

Copyright (c) 2021 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Technological Information, Management & Engineering Sciences (2021),2(2), 31-37

Zain and Umer

[1]Assistant Professor, Department of Financial Technology and Information Systems Capital University of Science and Technology (CUST), Islamabad
[2]Lecturer, Department of Financial Technology and Information Systems Capital University of Science and Technology (CUST), Islamabad

Copyright (c) 2021 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.