AI-Driven Optimization of Backend and Cloud Infrastructure in Large-Scale Financial Systems

Vivek Kumar Mishra
IEEE Senior Member
IEEE, Delaware, USA
vivekmishra@ieee.org

Abstract: *Large financial institutions operate some of the most demanding distributed systems in existence, processing millions of transactions per second under strict latency guarantees, unyielding regulatory deadlines, and increasingly constrained cloud budgets. Despite widespread adoption of cloud-native architectures and microservice-based backends, infrastructure management in these environments remains overwhelmingly manual or rule-based, resulting in chronic over-provisioning, suboptimal query performance, and reactive incident response. In this paper, we present a unified conceptual framework for applying artificial intelligence to three foundational challenges in financial infrastructure: (1) reinforcement learning and contextual bandit methods for cost-aware microservice autoscaling and job scheduling under regulatory and SLA constraints, (2) learning-based database indexing, caching, and query routing for extreme-scale transactional and analytical workloads, and (3) AI-driven incident prediction and automated remediation for complex distributed systems. For each pillar, we describe problem formulations, system architectures, model selection considerations, training strategies, and deployment safeguards tailored to the risk-sensitive nature of banking and capital markets environments. We argue that the integration of these techniques into a coherent AI-augmented operations layer can deliver substantial reductions in cloud expenditure, measurable improvements in SLA compliance, and a meaningful decrease in mean time to resolution for production incidents, capabilities that are rapidly becoming differentiators for financial institutions operating at global scale.*

Keywords: Reinforcement Learning, Cloud Infrastructure, Microservice Autoscaling, Database Optimization, AIOps, Incident Prediction, Financial Systems, Machine Learning for Systems

Copyright to IJARSCT
www.ijarsct.co.in

DOI: 10.48175/IJARSCT-31286

615

ISSN
2581-9429
IJARSCT