

Renadh Chowdhury

646-220-6996 | renadhc@gmail.com | renadh12.github.io | github.com/renadh12

SUMMARY

Staff Platform Engineer with 8+ years, currently building production grade ML inference systems for real-time fraud detection and risk scoring at Visa AI Platform. Expert in ML Infrastructure, Network, Kubernetes, Distributed Systems, ML Lifecycle solutions and Developer Productivity.

TECHNICAL SKILLS

Languages: Python, Rust, Java, TypeScript, JavaScript, Groovy, SQL

Frameworks: FastAPI, Uvicorn, Node.js, Flask, Serverless, Google Cloud SDK

Network & Infra: F5 Load Balancer, DNS, Istio, VXLAN, Calico CNI, Kubernetes, cert-manager

Libraries: kube-rs, tokio, k8s-openapi, PySpark, ONNX Runtime, pandas, NumPy, Matplotlib, Serverless Offline, Ray Serve, KubeRay

Data: Apache Kafka, Redis, Spark

Cloud: AWS (Lambda, S3, EKS), Azure (AKS, Key Vault, Managed Identities, HDInsight Kafka), GCP (GKE, Compute Engine, Google CAS)

Platform: Docker, Prometheus, Grafana, Jenkins, GitHub Actions, Packer, Helm

PROFESSIONAL EXPERIENCE

Staff Machine Learning Engineer, Real-Time Inference & AI Platform

Feb. 2024 – Present

Visa Inc.

Austin, TX

- Led technical strategy and ground-up development of 2 production grade AI/ML platforms (GCP GKE, Visa MKE on-prem) powering real-time fraud detection, serving 13+ models including CyberSource (\$623B payment volume, 6.4B transactions annually), Visa Direct, and Visa A2A Payments, handling a combined total compute of 12k+ transactions per second.
- Built high-performance real-time inference system coupling ML serving framework (Python/Rust/FastAPI) and ONNX Runtime optimizations with advanced Kubernetes patterns (multi-container pods: 1 proxy:N inference)—delivering 86% resource reduction (600→80 CPU) with 60% latency improvement (50ms→20ms at 99.95% processing 2000 TPS) for 9 CyberSource models migrated from bare-metal to on-prem MKE while achieving sub 5ms (99.95%) at 500 TPS with only 36 vCPU for 1 model in GCP.
- Architected and implemented scalable Kubernetes infrastructure through network redesign (Istio Gateway/Virtual Service/DestinationRule → K8s Service → Pod) optimized for high-throughput, low-latency inference and led strategic cluster optimization achieving 53% namespace, 46% pod, 50% CPU reduction—enabling cost-effective migration from bare-metal to on-prem MKE versus expensive public cloud solutions.
- Accelerated time-to-production by 70% (6 months to 14 weeks) by developing automated CI/CD platform with Jenkins/Packer/Helm pipelines, self-service APIs/SDKs and reusable infrastructure sizing templates for model packaging and deployment workflows—enabling ML engineers, data scientists to deploy and scale models independently with zero platform bottlenecks.
- Bootstrapped enterprise ML platforms with data streaming pipelines (Kafka/Redis/Flink/Hadoop) for feature engineering and model training, observability (Prometheus/Grafana/ElasticSearch/Kibana), security (certificate management, vault, encryption), and benchmarking tooling—enabling end-to-end ML lifecycle with ITDR compliance.
- Mentored 12 engineers and led cross-functional collaboration with 10+ teams (Data Science, Infrastructure, Cybersecurity, Model Engineering) while training Production Reliability Engineering teams on multi-datacenter deployments across platforms.

Senior Software Engineer, Risk & Authentication

Sept 2022 – Feb 2024

Visa Inc.

Austin, TX

- Designed Java-based Karate framework for real-time API validations across 5 data centers for EMV 3D-Secure Authentication flows in Production, leveraging Apache Kafka and Hashicorp Vault.

Software Engineer, Developer Productivity

Dec. 2020 – Sept. 2022

FOX Corp.

New York, NY

- Built event-driven CI/CD platform with AWS Lambda/SNS/SQS and TypeScript/Python to enable on-demand deployment and improve developer productivity.

Founding Partner, Lead Engineer	Feb. 2020 – Sept. 2022
<i>Next Level Sports Management</i>	<i>New York, NY</i>
• Co-founded and developed full-stack platform from concept to production, serving 100+ athletes and universities across 3 continents.	
Software Engineer	Mar. 2019 – Aug. 2019
<i>Red Ventures</i>	<i>Austin, TX</i>
• Spearheaded end-to-end test automation for The Points Guy US website (\$50M revenue, 10M monthly visitors) using JavaScript/Cypress/Selenium with New Relic monitoring, improving system reliability and reducing validation time by 10%.	
Software Engineer	Aug. 2017 – Mar. 2019
<i>Affinity Solutions Inc</i>	<i>New York, NY</i>
• Developed Python based selenium test automation suite for financial loyalty program platform (\$2M revenue, 10M users), reducing bugs by 15% through comprehensive validation and monitoring.	

PERSONAL PROJECTS

PlatML <i>Rust, Axum, React, GKE</i> github.com/renadh12/platml	2025
• PoC ML platform with sub-minute model deployment via Rust-powered APIs and model management	
KubeRust - Rust based K8 Controller <i>kube-rs, k8s-openapi, tokio, kind, docker</i>	Dec 2024 – Present
• PoC Kubernetes controller for dynamic node scaling based on pod utilization metrics	
projectX <i>PySpark, Docker, Azure</i> github.com/renadh12/projectX	2023
• PoC End-to-end ML pipeline with feature engineering and model evaluation on AKS	

EDUCATION

Western Governor's University	Jan 2025 - Present
<i>Bachelor of Science in Software Engineering</i>	<i>Online, USA</i>
NYC Data Science Academy	Oct 2019 - Dec 2019
<i>Data Science with Python: Machine Learning, Data Analysis</i>	<i>New York, NY</i>