

Jayanth Yanamandala

North Bergen, New Jersey | jayanthkyanamandala@gmail.com | (201) 912-6826 | [GitHub](#) | [LinkedIn](#) | [Website](#)

PROFESSIONAL SUMMARY

Software engineer with 3+ years of experience building enterprise data platforms, distributed services, and production AI systems in fintech. Works across backend infrastructure and user-facing applications, with recent experience in agentic systems, access control, deployment automation, observability, and testing.

SKILLS

Languages: Python, TypeScript, JavaScript, SQL, Bash

Backend: FastAPI, Flask, Django, REST, GraphQL, OpenAPI, Swagger, asyncio, Microservices, OAuth 2.0, JWT, RBAC, Rate Limiting

Frontend: Vue.js, Vuetify, React 19, Next.js, Tailwind CSS, AG Grid

AI/ML & Agents: RAG, AI Agents, LangGraph, LiteLLM, MCP, pgvector, Milvus, Embeddings, RAGAS, AWS Bedrock, OpenAI, Anthropic, Gemini APIs, Claude Code, Cursor, GitHub Copilot, Multi-Agent Systems

Data & Cloud: PostgreSQL, Snowflake, Redis, Kafka, Pub/Sub, Snowpipe, dbt, AWS: ECS, Lambda, S3, SQS, ECR, CloudWatch, EKS; Kubernetes, KEDA, Docker, Terraform

Testing & Observability: Pytest, Vitest, Playwright, GitHub Actions, Grafana, Prometheus, OpenTelemetry

WORK EXPERIENCE

Software Engineer, Data Platform & AI

Feb 2025 - Present

KKR

New York, NY

- Built a distributed job-execution platform supporting 6,000+ production pipelines and 273K+ weekly runs, using SQS dispatch, KEDA-autoscaled fault-isolated workers, health checks, and Grafana monitoring.
- Built the platform's usage-analytics service and pluggable instrumentation layer, capturing UI and API session events through SQS to S3 to Snowpipe to Snowflake to power adoption reporting across applications.
- Built the core of the firm's data-governance and discovery platform as 1 of 3 engineers using Vue/Vuetify, Flask/GraphQL, PostgreSQL, Snowflake, and AG Grid; reached 1,800+ provisioned and 400+ monthly active users across four business groups.
- Implemented JWT authentication, RBAC, and SSO integration across data-platform services, and automated Snowflake role provisioning to reduce access turnaround from 2 business days to under 5 minutes.
- Provisioned AWS infrastructure with Terraform and extended GitHub Actions pipelines to gate deployments on linting, builds, integration tests, and Playwright end-to-end tests for critical user flows.
- Increased frontend test coverage from 0% to 80% with Vitest and established Vitest as the default unit-testing framework for new React 19 repositories.
- Built an AI gateway unifying AWS Bedrock, OpenAI, Anthropic, and Gemini behind one interface so firm data remained within the environment; later migrated it to LiteLLM to reduce provider-integration maintenance.
- Shipped citation-backed policy and HR assistants evaluated with RAGAS against an SME-reviewed dataset, then evolved them into a tool-using agent integrating web search, datarooms, connectors, MCP servers, Outlook, Excel, and PowerPoint; increased firm adoption from roughly 30% to 60%.
- Built a dataroom ingestion layer connecting Box, SharePoint, S3, and direct uploads to vector knowledge bases, supporting individual datarooms up to 5,000 files or 100GB and team datarooms with millions of documents.

Software Engineer, Risk Engineering

Aug 2024 - Feb 2025

Visa

New York, NY

- Reduced critical-path tail latency 30% in a real-time risk-decisioning pipeline using Python, Kafka, asyncio fan-out, and request batching; shipped 6 transaction-enrichment features and added circuit breakers, timeouts, and standardized error handling

- Built rule versioning and rollback for a tier-1 fraud service evaluating 200 configurable rules per transaction, allowing Risk Strategy teams to deploy or revert changes without engineering releases and reducing turnaround from several days to under 30 minutes.
- Delivered a self-service backtesting platform for validating rule changes against 90 days of production history, removing engineering from the workflow and catching multiple high-impact misconfigurations before deployment.
- Added Prometheus metrics, Grafana dashboards, and OpenTelemetry tracing for API latency, throughput, worker health, Kafka consumer lag, error rates, and end-to-end multi-service request flows.
- Published OpenAPI and Swagger documentation for internal APIs and implemented rate limiting and request throttling to protect shared services during traffic spikes.

Software Engineer

Jan 2022 - Dec 2022

Cognizant

India

- Built and maintained 25+ Python, Django, and PostgreSQL REST APIs for a Fortune 500 bank's loan-origination platform, reducing response latency 38% through query-plan optimization, N+1 elimination, and ORM caching.
- Replaced a 12-hour nightly batch with a Kafka-backed reconciliation service, improving data freshness from T+1 to under 5 minutes and reducing reconciliation tickets from 10 per week to 2 through automated mismatch detection and alerting.
- Added Redis caching for frequently accessed metadata, achieving an 87% hit rate on hot read paths and eliminating repeated database queries during peak loan-processing workflows.

PROJECTS

Clausa — Agentic Data-Observability Platform

- Built a LangGraph agent that investigates data-quality anomalies through structured tool calls and returns evidence-grounded root-cause hypotheses using profiling statistics, lineage, and recent schema or pipeline changes.
- Implemented anomaly detection using static thresholds and rolling z-scores, with dbt manifest parsing to resolve upstream lineage and trace failures to source tables.
- Built the platform with FastAPI, Next.js, LiteLLM, PostgreSQL, and pgvector; deployed it on EKS with KEDA autoscaling and Terraform-provisioned infrastructure.
- Added an LLM evaluation suite with auto-derived labels to CI as a release gate, blocking deployments when root-cause accuracy regresses.

EDUCATION

Stevens Institute of Technology

M.S., Data Science, GPA: 3.8/4.0

Jan 2023 - May 2024

Jawaharlal Nehru Technological University Anantapur

B.Tech, Computer Science & Engineering

Jan 2018 - Jan 2022