

Karthik Subramanian

Software Engineer - Distributed Systems & Backend Infrastructure - Platform & Data Infrastructure - Java/Python - Cloud (GCP/AWS)

✉ karthiksha874@gmail.com ☎ +1-646-236-6150 📍 New York, NY 🌐 LinkedIn 🔗 Portfolio 🏠 Github

SUMMARY

Software engineer with 3 years building large-scale internal infrastructure at Fortune 500 scale, not operating pipelines, but designing the systems other engineers depend on. One of 5 backend engineers who built Connexio, Lowe's internal platform replacing Informatica and Talend, serving 50+ engineering teams running 19,000+ pipelines at petabyte scale.

EDUCATION

M.S. in Computer Science, *New York University*
GPA: 3.7/4.0

08/2024 – 05/2026 | New York, NY

Bachelor of Engineering (B.E.) in Information Science
New Horizon College of Engineering, Visvesvaraya Technological University

06/2017 – 06/2021 | Bangalore, India

WORK EXPERIENCE

Software Engineer, Data Platform, *Lowe's*

08/2023 – 07/2024 | Bangalore, India

- Led core subsystem development on Connexio. On a 5 person backend team, architected the RDBMS ingestion and transformation framework in Scala, supporting PostgreSQL, MySQL, Oracle, Teradata, and DB2 with configurable parallel connections and optimized query execution, serving 50+ engineering teams running 19,000+ pipelines at petabyte scale, driving approximately \$3M/year in vendor cost savings by replacing Informatica and Talend.
- Built Spark Structured Streaming framework powering 2,000+ pipelines processing approximately 1M events per second during peak traffic (Black Friday and Cyber Week); implemented lag and offset alerting, early-warning thresholds, and exponential-backoff retries to maintain reliability.
- Implemented platform guardrails including PII masking, schema compatibility enforcement, retry semantics with exponential backoff, and pipeline validation gates, reducing support escalations by 60%.
- Optimized backend services on GCP (Compute Engine, Cloud Functions, Pub/Sub), improving real-time API latency by approximately 20% through BigQuery query optimization, partitioning, and clustering strategies, reducing infrastructure costs.

Associate Software Engineer, Data Platform, *Lowe's*

08/2021 – 08/2023 | Bangalore, India

- Early engineer on Connexio, Lowe's internal platform built to replace Informatica and Talend. Joined the data platform team as a founding contributor and helped build the foundational ingestion and validation components that later scaled across the organization. Promoted to Software Engineer after 2 years.
- Built GCP migration infrastructure enabling 50+ teams to migrate 5,000+ tables from on-prem Hadoop to cloud; automated ingestion, transformation, validation, and governed egress with PII masking and schema validation as part of federated Data Mesh architecture.
- Partnered with DevOps to enforce unit and regression test coverage gates in CI, reducing troubleshooting time by approximately 60%.

LEADERSHIP & IMPACT

Partnered with the Data Org Chief of Staff to build a table ownership registry org-wide, adopted across 50+ teams, improving escalation speed and eliminating manual Slack-hunts during on-call incidents.

PROJECTS

TenantLens - Multimodal AI Agent, NYC Tenant Rights 📄, *Google GDG x Columbia Business School Hackathon*

- Designed end-to-end architecture for a 3-agent A2A pipeline with ADK orchestrator; locked inter-agent contracts and payload schema upfront to enable parallel development across a 4-person team.
- Built the landlord breach-detection agent against NYC HPD's live Open Data API, deployed the full stack to Google Cloud Run, and served as de facto tech lead - reviewing PRs and resolving cross-agent integration issues.

SteamLens, Distributed NLP Pipeline (Dask, GPU Transformers, Big Data), *Built distributed systems pipeline using Dask and GPU Transformers that groups Steam reviews by theme and auto-summarizes pros and cons; processes approximately 400k reviews in under 3 minutes with Streamlit UI.*

ProofGate 📄, *Blast-radius firewall for AI agents · Enterprise Agents Hackathon (Emergence AI x Nebius)*

- Built a runtime governance firewall for AI agents that scores the "blast radius" of a proposed action before execution, blocking catastrophic mistakes (e.g., a delete call that would wipe 10k+ production rows) that pass standard access-control checks.
- Designed a deterministic Python policy engine that owns the final ALLOW/BLOCK decision from LLM-extracted risk signals (Nebius Token Factory) - the model itself never adjudicates - backed by server-validated, tamper-proof rollback snapshots and postcondition verification against real outcomes.
- Built a live MCP gateway (OAuth 2.1 + PKCE) and autonomous agent loop; validated against adversarial inputs including forged proofs and models that misreport risk, backed by nearly 300 passing tests.

SKILLS

Languages: Python, Java, Scala, SQL, Bash, C, C++, JavaScript

Architecture & Design: Control plane design, multi-tenant architecture, schema governance, failure model design, data mesh, ETL/ELT orchestration

Data & Streaming: Apache Spark, Apache Kafka, Apache Airflow, Dask

Platform & Infra: Docker, Kubernetes, CI/CD, Jenkins, Linux, Git, REST APIs, Microservices, System Design

Databases: PostgreSQL, MySQL, MongoDB, DynamoDB, Oracle, Teradata

Cloud: Google Cloud (GCP), AWS, Azure