

V GNANA SIVA SAI

Platform / Infrastructure Engineer | Distributed Systems & Kubernetes

gnana097@gmail.com | 9578627762

gnanasivasai.dev | linkedin.com/in/gnanavaradaraju | github.com/gnana997 | npmjs.com/~gnana997

PROFESSIONAL SUMMARY

Platform and infrastructure engineer with 6+ years building distributed systems, control planes, and orchestration tooling in Go. I build platforms with identity-based access, per-user governance, and auditability, including **Periscope**, a multi-cluster Kubernetes console with impersonation-based isolation and keyless cluster access. At Pure Storage, I architect high-throughput Go services (100K+ logs/second), modernize Kubernetes infrastructure (25% AWS cost reduction), and build AI-powered platform tooling.

PROFESSIONAL EXPERIENCE

Pure Storage, Member of Technical Staff 3 (MTS-3)

June 2023 - Present | Bangalore, India

AI Platform Engineering

- Designed and built an internal platform for no-code creation and execution of LLM agents, with horizontally scalable, stateless services
- Built an AI-powered observability agent for natural-language incident debugging across logs, Kubernetes, and metrics via Model Context Protocol (MCP) servers; shipped to production, with the team reporting an estimated 10-15% reduction in MTTR

Kubernetes & Cloud Infrastructure

- Migrated Kubernetes fleet from EKS 1.24 to 1.31, modernized EC2 instance types, and consolidated infrastructure, achieving a 25% AWS cost reduction
- Operated production Kafka, Elasticsearch, and Cassandra clusters; led the Cassandra v3-to-v4 migration on production using K8ssandra with Medusa backup integration

Distributed Systems & Performance

- Designed a high-throughput traffic monitoring service in Go using goroutine-based concurrency, sustaining 100,000+ logs/second at low latency
- Built a rule-evaluation engine in Go that tokenizes queries into AST trees with parallel multi-level parsing, holding 30-50ms latency across 1,000+ rules
- Built a Salesforce auth-token pooling wrapper over jsforce in the platform library, adopted org-wide; eliminated redundant re-authentication calls flagged by Salesforce and reduced auth-overhead latency

Velotio (Client: Pure Storage), Senior Software Engineer

August 2021 - May 2023 | Pune, India

- Re-architected a serverless application to a server-based design, introducing Snowflake connection pooling for a measured 20% improvement in application performance
- Replaced heavy multi-join Snowflake queries on hot API paths (5+ tables, up to an 11-table join) with parallel data pulls joined in application code, cutting latency on those paths by roughly 70%

Tata Consultancy Services, Systems Engineer

June 2019 - July 2021 | Pune, India

- Worked on migrating on-premises infrastructure to a serverless architecture for the London Stock Exchange Group, part of an effort that reduced operational costs by 15%

SELECTED PROJECTS

Periscope: Multi-Cluster Kubernetes Console

Personal open source | Go + React/TypeScript | periscopehq.dev | github.com/gnana997/periscope

- Built a self-hosted, multi-cluster Kubernetes console shipped as a single stateless Go binary with an embedded React SPA and no external database dependency
- Implemented per-user access through Kubernetes impersonation, attributing every action to the user's IdP identity and enforcing per-cluster RBAC; eliminated static cloud credentials using IRSA / Pod Identity, with OIDC sessions held in memory only, no shared service accounts, and no persisted kubeconfs
- Streamed live resource state over Server-Sent Events with reconnect resumption and per-user stream caps so one user cannot overload a large cluster; logged every privileged action to a SQLite-backed audit store, with structured JSON to stdout and no external logging system needed
- In active development (v1.1): AWS EKS-native features (Access Entries / Pod Identity viewer, Inspector v2 CVE surfacing, Karpenter dashboard); v1.2 adds agentic operation, with an MCP layer and AI debugging agent that act through the same user-impersonation and audit path as humans

OPEN SOURCE

NPM Packages (npmjs.com/~gnana997)

- **mcp-dev-kit** (v0.2.0): testing & debugging toolkit for MCP servers
- **@gnana997/node-jsonrpc** / **node-ipc-jsonrpc** / **node-stdio-jsonrpc**: transport-agnostic JSON-RPC 2.0 clients for TypeScript/Node.js, over generic transport, Unix-socket IPC, and stdio respectively

VS Code Extensions

- **Ollama Copilot**: local-LLM-powered code completion and chat in VS Code, with zero code leaving the machine

SKILLS

Languages: Go, TypeScript, JavaScript, Java, Python

Infrastructure & Orchestration: Kubernetes, AWS (EKS, IRSA / Pod Identity), Docker, Helm, Terraform

Distributed Data & Messaging: Cassandra, Kafka, Elasticsearch, Redis, PostgreSQL, Snowflake

CERTIFICATIONS

- **Certified Kubernetes Administrator (CKA)**, Cloud Native Computing Foundation (2022)

EDUCATION

BTech, Vellore Institute of Technology

2015 - 2019 | Vellore, India