VAIBHAV AGARWAL

Lead Software Engineer | Streaming Data Processing, Distributed Systems & Cloud Infrastructure

Bangalore, India (Remote-Ready) | +91 8447552014 | vaibh.agarwal06@gmail.com | LinkedIn | Portfolio

PROFESSIONAL SUMMARY

Principal-level Backend & Distributed Systems Engineer with ~8 years of experience owning the architecture, reliability, and cost efficiency of high-throughput streaming pipelines, real-time telemetry systems, and multi-tenant data platforms. Core platform architect and major technical contributor driving end-to-end system design for a multi-tenant cloud platform serving ~2,000 enterprise customers and ingesting ~2M+ daily streaming events. Proven track record of leading streaming architecture decisions - achieving a ~100x reduction in consumer lag, cutting cloud compute and storage costs by 40–70%, driving a strong written engineering culture (design docs, root-cause analyses, runbooks), and leveraging AI-assisted development to multiply delivery velocity.

CORE TECHNICAL SKILLS

- **Languages & Frameworks:** Go (Golang - Primary), Java, Python, SQL, GraphQL, Node.js, Shell Scripting
- **Streaming & Event-Driven Systems:** Apache Kafka, Amazon MSK, Change Data Capture (CDC), Debezium (Binlog Streaming), Consumer Autoscaling, Back-Pressure & Windowed Batching, Idempotency & Deduplication, Dead-Letter Queues (DLQ), Offset & Delivery Semantics
- **Storage & Caching:** Managed NoSQL (MongoDB), Analytical Datastores (ClickHouse, Apache Druid, Elasticsearch), In-Memory Caches (Redis Pub/Sub & Caching), Relational (MySQL)
- **AWS, Containers & CI/CD:** AWS (MSK, Secrets Manager, Cloud Services), Kubernetes (Custom Controllers, Dynamic Client), KEDA, Docker, Argo, CI/CD Pipelines
- **Observability, Governance & AI Tooling:** Prometheus, Grafana, ELK Stack, Fluentbit, AWS Cloudwatch, Structured Logging, Architecture Decision Records (ADRs), Operational Runbooks, AI-Assisted Development (Claude Code)

PROFESSIONAL EXPERIENCE

LEADSQUARED - Bangalore, India

Multi-Tenant Enterprise Data Ingestion & Analytics Platform (~2,000 Enterprise Customers | ~2M Daily Streaming Events)

Lead Software Engineer (Domain Architect - Streaming Ingestion & Query Platform) | April 2025 – Present

- **~100x Streaming Lag Reduction & Back-Pressure Control:** Owned end to end re-architecture of the core Kafka to ClickHouse ingestion pipeline processing ~2M daily events across ~1,000 tenant instances, replacing synchronous single-record writes with a windowed batch-accumulation and in-memory idempotency/deduplication engine to improve consumer lag by ~100x. Enforced flush-barrier semantics to guarantee strict offset ordering and delivery semantics across interleaved schema and data streams.
- **40% Cloud Infrastructure Cost Reduction & Topology Simplification:** Designed and hardened workload-driven consumer autoscaling on AWS MSK and Kubernetes using KEDA, cutting compute spend by 40%. Drove deliberate architectural simplification by retiring a fragile custom master-worker pod model in favor of KEDA minimum-replica baselines, adding level-triggered self-healing reconciliation, scale-to-zero for idle workloads, and per-group resource tuning.
- **Resilience, Error Classification & Dead-Letter Routing:** Overhauled event-driven failure recovery by deprecating legacy delay-retry queues and architecting strongly typed error classification (segregating transient infrastructure faults from malformed payload packets) with deterministic retry policies, Dead-Letter Queue (DLQ) routing, and strict schema safety guardrails.
- **Observability Defaults & Deployment Hygiene:** Raised platform operational standards by instrumenting partition-level lag and flush-boundary metrics in Prometheus, adding visibility into the infra by adding Grafana Dashboards with Prometheus and CloudWatch metrics, optimizing structured log volume, and establishing feature-flagged canary rollouts with zero-downtime rollback procedures.
- **Written Engineering Culture & Technical Direction:** Established a rigorous written culture across the engineering org by authoring formal Architecture Decision Records (ADRs), system context specifications, QA test plans, rollout checklists, and operational runbooks.
- **Cross-Functional Partnership & Mentorship:** Acted as the central technical authority across 5 external data-producing engineering teams to govern data contracts, schema evolution, and integration SLAs. Anchored primary on-call incident response, enforced non-negotiable code review standards, and mentored junior engineers.

Senior Software Engineer | April 2024 – March 2025

- **65% Memory & 70% Storage Cost Optimization:** Diagnosed production resource bottlenecks end-to-end, isolating and fixing critical memory leaks to reduce service memory footprint by 65%, and re-architecting NoSQL/configuration storage schemas to cut cloud storage costs by 70%.
- **Distributed Query Engine & Cross-Dataset Joins:** Stabilized and scaled the analytical query compiler supporting multi-dataset Common Table Expression (CTEs) joins and dynamic pivot aggregations, resolving systemic identifier collisions and hierarchy propagation defects.
- **Production Root-Cause Analysis & Scheduling Reliability:** Authored formal written root-cause analyses before implementation to permanently resolve complex UTC timezone normalization and recurrence edge cases in the distributed scheduling engine, eliminating recurring production incidents.

Software Engineer | December 2021 – March 2024

- **Zero-Downtime Platform Migration & NoSQL Tooling:** Engineered automated CLI migration tools, data-transformation pipelines, and MongoDB aggregation/indexing scripts to transition enterprise customers from legacy reporting models to a modern multi-tenant schema with zero data loss.
- **High-Throughput APIs & Incremental ETL:** Built customer-facing analytical REST APIs, Redis cache-invalidation workflows, role-based access controls, and incremental ETL data synchronization pipelines; conducted technical hiring interviews and mentored junior developers.

HEWLETT PACKARD ENTERPRISE (HPE) - Bangalore, India

Software Engineer (July 2019 – December 2021) | Software Engineering Intern (January 2019 – July 2019)

- **Cloud Data Services & Streaming Telemetry:** Engineered backend microservices across Functions, Analytics, and GraphQL modules using Go, Java, Python, Kafka, Apache Druid, and Kubernetes, shipping the core REST API framework that transitioned HPE Storage Data Services from on-premise hardware to a cloud-native SaaS console.
- **Server Telemetry Monitoring & Fault Prediction (Java / Python):** Built a system performance telemetry monitoring and ML-driven fault-prediction engine for HPE Synergy server infrastructure, integrated cloud-init, and created automated OS provisioning tools that accelerated multi-OS artifact testing.
- **1st Prize Hackathon - Automated Support Triage ($100/Query Savings):** Designed and built an automated L1/L2 support triage bot that won 1st Prize in the HPE Storage Hackathon and projected support cost reductions of $100 per query.

EDUCATION

B.Tech in Computer Science and Engineering (Data Science Specialization) — PES University, Bangalore (2015 – 2019) | **GPA: 9.55 / 10.0**
Honors: Multi-semester recipient of the CNR Rao Merit Scholarship; Subject Matter Expert (SME) for PESU I/O, designing and teaching a hands-on Big Data with Python course.

PUBLICATIONS & RESEARCH

- **ISRO & Coffee Board of India (Feb 2019):** Co-authored and presented "Deep Learning and Statistical Models for Detection of White Stem Borer Disease in Arabica Coffee" at the ISPRS-GEOGLAM-ISRS Conference following a research internship at RRSC-ISRO.
- **Indian Institute of Technology (IIT) Mandi (Summers 2017 & 2018):** Research Intern modeling network telemetry anomalies and failure prediction using LSTM, NARX, and ARIMA statistical time-series architectures.