

Ritesh Sharma.

Senior Software Engineer / Tech Lead
Java & Distributed Systems · AI-Native Products
India · Remote (US / EU overlap)

EMAIL ritesh00sharma@gmail.com

LINKEDIN [/in/ritesh-sharma-](https://in.linkedin.com/in/ritesh-sharma-)

GITHUB [@jamnaritesh](https://github.com/jamnaritesh)

Senior software engineer and hands-on technical lead — **8+ years** building institutional-equities trading systems that clear **10M+ trades a day**. Cut a Tier-1 platform's recovery time from **90+ minutes to 7**; guide **12–15 engineers** and lead Sev-1 response. Most recently: an AI document-intelligence product, from **customer discovery to live beta**, as sole engineer.

EXPERIENCE

Consultant, Equities Trading Platform

APR 2023 – PRESENT

CLSA · Pune, India [VIA NITYO INFOTECH](#)

- Cut startup of a **Tier-1, multi-region trading platform** from **over 90 minutes to 7 minutes**, shrinking the recovery window — and the operational and regulatory exposure — of intraday restarts across regions.
- Traced the bottleneck to a single-threaded, full-scan Cassandra load; rebuilt bootstrap as **parallel reads over non-overlapping Murmur3 token ranges**, tuning queue and threshold settings to sustain the higher startup concurrency.
- Pinned component-owned Ignite caches to designated nodes using **node filters** and re-sequenced startup so non-cache-owning components join only after data population — eliminating avoidable rebalancing during cluster formation.
- **Sole engineer** for the platform's Apache Ignite upgrades from **2.3.x to 2.15.x**; diagnosed and resolved **partition-map exchange (PME) delays** and cluster instability spanning topology, compatibility and JVM configuration.
- Provide **architecture reviews and technical direction for 12–15 engineers** while remaining hands-on; lead **Sev-1 on-call** investigations across cache access patterns, queue contention, thread utilization and service latency.

Java

Spring Boot

Apache Ignite

Cassandra

JVM

SRE

Senior Software Engineer / Tech Lead

JUL 2018 - APR 2023

CLSA Technologies & Services · Pune, India

- Architected **JADE**, a regulatory-reporting microservice streaming **10M+ equity trades/day** at **<50 ms latency** on Spring Boot and Apache Ignite.
- Built the market-side **netting engine**, netting **1,600+ trades/second** and cutting downstream settlement traffic by **25%**.
- Optimized a **55-node Apache Ignite cluster** through affinity colocation and PME tuning, reducing cold-start time by **60%** — groundwork for the later full startup redesign.
- Modernized the legacy C# confirmation bus to Java/Spring and integrated **Omgeo CTM** for T+0 confirmations, cutting turnaround from **5 minutes to under 30 seconds**.
- Redesigned the **GENESIS** multi-asset booking interface, reducing manual data entry by **35%** and booking errors by **20%**.
- Instituted Jenkins-based **CI/CD** with containerized deployments, taking release time from **2 hours to 15 minutes** and production defects down **30%**.
- Led and mentored a **five-developer team**; conducted **200+ code reviews** and ran training on distributed caching and performance tuning.
- Built tooling adopted by engineering teams: a real-time front-to-back **connectivity graph** that halved troubleshooting time, and a custom **FIX parser** that sped issue triage by **40%**.

Java

Spring Boot

Omgeo CTM

FIX

Angular

Jenkins

SELECTED WORK

Document-Intelligence Engine for HVAC Estimating

STEALTH SAAS · BETA

Python · Gemini · Claude · computer vision · deterministic validation

- Designed and built the entire Python engine that turns **hundred-page engineering drawing sets** into structured, estimator-ready data: document parsing and **computer vision** feed **Gemini and Claude** reasoning, with **deterministic validation** gating the output.
- Engineered **whole-drawing-set reasoning** rather than page-by-page OCR — equipment on one sheet links to legends, notes and schedules on others, so the engine reconciles **cross-sheet equipment identity**, duplicate mentions, missed detections and conflicting evidence, with every fact **traceable to its source drawing**.
- Shaped the product through **customer discovery** with HVAC estimators; **V1 is live with beta users**, instrumented with **Langfuse** and running on AWS.

Python

Gemini

Claude

Computer Vision

Langfuse

AWS

SKILLS

LANGUAGES & FRAMEWORKS

Java (8–21) · Spring Boot · Spring Cloud · Python · SQL · Hibernate/JPA · JUnit · Bash

DISTRIBUTED SYSTEMS & DATA

Apache Ignite (affinity, PME, cluster tuning) · Cassandra (internals, token-range parallelism) · Redis · Oracle · PostgreSQL · concurrency · high availability

PRODUCTION ENGINEERING

JVM tuning · Sev-1 incident response · Micrometer · Prometheus · Docker · Jenkins · CI/CD · AWS (EC2, EKS, RDS)

EDUCATION

B.E. in Information Technology

2018

MIT College of Engineering, Pune

APPLIED AI

LLM applications (Gemini, Claude) · document intelligence · computer vision · structured extraction · deterministic validation · entity resolution · Langfuse

FINTECH DOMAIN

Institutional equities · trade lifecycle · booking · netting · settlement · regulatory reporting · T+0 confirmations · Omgeo CTM · FIX

PRODUCT DELIVERY

Customer discovery · domain research · MVP/beta delivery · architecture reviews · technical leadership

CERTIFICATIONS

Claude Certified Architect – Foundations ·

Anthropic · [Credly](#)