

DEEPESH RATHOD

Founding Engineer • Tech Lead • Backend & Cloud Architecture

+91 79872 15728 • deepeshd03938@gmail.com • github.com/deepesh-01 • Bengaluru, India

SUMMARY

Founding Engineer and Tech Lead with 4+ years architecting and scaling production systems from zero to multi-tenant scale. At Zoca, defined the foundational AWS ECS/Fargate architecture that supported customer growth from 90 to 1,500 (1,000+ paying, \$1M+ ARR), designed tenant isolation and audit infrastructure for future SOC2/HIPAA readiness, and built an AI-native engineering toolchain that cut RCA time from hours to minutes. Balances deep architectural ownership with technical leadership — runs the hiring loop, mentors as Senior Buddy, and shapes engineering culture as deliberately as systems.

DEFINING WORK

- **Stack-defining decisions:** chose AWS ECS + Fargate over EC2, consolidated 3 servers and 50+ Lambdas into 6 autoscaling services + 10 workers + 10 Lambdas, and orchestrated zero-failure migration of all legacy users — the calls that supported scale from 90 to 1,500 customers and \$1M+ ARR.
- **AI-native engineering platform:** built a Claude Code + custom Metabase MCP + AWS CLI toolchain that compresses outage and customer-issue RCAs from 2–3 hours of dedicated engineering bandwidth to ~5 minutes; complemented by an auto-updating doc pipeline and a Mac-based local infrastructure clone where AI agents write tests and validate user journeys end-to-end.
- **Force-multiplier leadership:** helped scale engineering from 2–3 to 15–18, ran 20+ technical interviews, served as Senior Buddy for 4+ engineers (mentees shipping bug fixes in week 1 and end-to-end features within their first month) — while owning audit infrastructure and AWS ownership after senior departures.

EXPERIENCE

Zoca — Founding Engineer / Tech Lead | Jan 2025 – Present

- Architected the foundational stack for a multi-tenant SaaS that scaled from **90 to 1,500 customers (1,000+ paying, \$1M+ ARR)** during my tenure; chose AWS ECS + Fargate over EC2 and consolidated 3 servers + 50–60 Lambdas into **6 autoscaling services + 10 workers + 10 Lambdas**, drastically reducing operational surface area while preserving pod-driven parallel delivery.
- Designed tenant isolation and database-layer audit logging — every internal access is gated through a unified RBAC system and recorded as an immutable change history, building the foundation for **SOC2/HIPAA readiness** across all tenant operations.
- Led **zero-failure migration** of all legacy users to the new architecture — no data loss, no production incidents, no rollbacks.
- Built an **AI-native engineering toolchain** (Claude Code + custom Metabase MCP + AWS CLI) that compresses outage and customer-issue RCAs from **2–3 hours of dedicated engineering bandwidth to ~5 minutes** — a 25–35x speedup that compounds across every incident; complemented by an auto-updating doc pipeline and a Mac-based local infra clone enabling AI agents to write tests and validate end-to-end flows without staging dependencies.
- Integrated Retell / Voice AI for an early experimental cohort and built the observability layer — surfacing latency, error, and behavioral signals the moment each call ends.
- Led the **Scheduling & Payments Pod** — transitioned from hands-on builder to technical guide, owning end-to-end delivery while keeping RFC processes, code review standards, and architectural guidelines aligned with business goals.
- Helped scale engineering from **2–3 to 15–18 engineers** — ran **20+ technical interviews**, served as Senior Buddy for 4+ engineers; mentees consistently fixing bugs in week 1 and shipping end-to-end features within their first month.

Zoca (formerly Chrono) — Founding Engineer | Jun 2023 – Dec 2024

- Spearheaded migration of the mobile-app backend from a legacy Python monolith to Node.js services; established team-wide developer experience standards and shared utility tooling that scaled cleanly with the team and accelerated onboarding for incoming engineers.

- Engineered a distributed media pipeline handling **1,000+ user uploads/day across 900+ GB of S3 storage**, with content-aware duplicate detection (0.1–5% reclaimed depending on user upload volume) and intelligent destination management.
- **Owned the entire AWS Cloud Infrastructure & DevOps** after senior leaders departed — absorbed full infrastructure ownership during a critical leadership transition, driving cost, reliability, and security decisions while keeping product delivery uninterrupted.
- Architected a hyper-local keyword demand system with 5-mile-radius geo-accuracy by integrating US Census data — enabling SMB-targeted SEO insights at scale.
- Designed and shipped v1 of the Scheduling product end-to-end: business logic, reminder workflows, and the underlying database schema.

Zoca (formerly Chrome) — Software Engineer Intern | Sep 2022 – May 2023

- Built citation and NAP (Name/Address/Phone) consistency automation — delivered a 1-month scoped initiative in 1 week, freeing the founding team for higher-leverage work.
- Implemented lemmatization-based keyword matching to improve search relevance and recall across content surfaces.
- Built a dynamic website generation pipeline combining Jinja templates with image classification, automating per-business landing page creation.

Arbunize Digital (VideoWiki) — Backend Developer Intern | Jan 2022 – Apr 2022

- Built backend APIs in Python/Django using FFmpeg for the VideoWiki video editor.
- Engineered a real-time progress tracking system for long-running FFmpeg tasks, improving editor UX for end users.

Yosecal Tech LLP (Finmo) — SDE Intern | Oct 2021 – Dec 2021

- Built NestJS servers integrating AWS SES, CloudTrail, SQS, and Slack API — powering management-event logging and transactional email rendering.

WEEKEND ARCHITECTURAL EXPLORATION

- **DevEx Automation: AI-Powered PR Review** — Slack + Git + Claude pipeline running LLM-powered sentiment analysis and technical review on PRs; **reviewed 30 PRs in the first 3 days vs. ~15 baseline manually** — 2x review throughput with improved consistency.
- **WeLog** — collaborative real-time journaling for partners; built the full-stack architecture end-to-end (auth, signup, encrypted private data, real-time shared entries with per-user data isolation) — currently in UAT.
- **MarketSignal AI** — real-time trading automation engine ingesting Angel One market data on a 15-min cron, evaluating conditions, and delivering Telegram alerts — full pipeline running on local infrastructure with no third-party orchestration.
- **Private Cloud on Raspberry Pi** — self-hosted server stack via Cloudflare Tunnels (no public IPs, no port forwarding); **2+ months continuous uptime, 3–5s cold boot**; engineered a password-less camera-to-laptop streaming flow over a unified LAN.

TECHNICAL SKILLS

Languages: TypeScript, JavaScript, Python, C++, Java, C

Backend & Frontend: Node.js, NestJS, Express, Django, Django REST Framework • React, Redux

Cloud & Infrastructure: AWS (ECS, Fargate, EC2, SES, CloudTrail, SQS, IAM), Docker, CI/CD pipelines

Databases: PostgreSQL, MongoDB

AI / Automation: Claude Code, custom MCP servers, Retell / Voice AI integration, AI agent observability, lemmatization, image classification, Selenium

Architecture & Leadership: Multi-tenant SaaS, microservices, monorepo, zero-downtime migrations, tenant isolation & RBAC, hiring & mentorship

EDUCATION

B.Tech, Information Technology — Shri Shankaracharya Technical Campus, SSGI (CSVTU) • GPA 7+/10 • 2019 – 2023