

Harish Mahamure

+91 7028601207 | harishmahamure.dev | hmahamure10@gmail.com || Pune, India

SUMMARY

Engineering Leader and Staff-level Full-Stack Engineer with 8+ years of experience designing and scaling cloud-native, distributed systems across healthcare, SaaS, and high-traffic digital platforms. Currently leading architecture and engineering initiatives across multiple product squads while managing a team of ~20 engineers and remaining hands-on with system design, performance optimization, database architecture, observability, and critical code contributions.

Proven track record of reducing API latency by ~83%, operating platforms handling 3M+ daily requests with 99.9% uptime, modernizing monolithic applications into domain-aligned services, designing scalable data and asynchronous processing architectures, and reducing infrastructure costs by up to 80%.

Strong expertise in distributed systems, PostgreSQL, Redis, Kubernetes, React, TypeScript, Node.js, API performance engineering, CI/CD, observability, and cloud infrastructure.

TECH STACK

Languages & Frameworks: TypeScript • JavaScript • React • Next.js • React Native • Expo • Node.js • NestJS • Rust • SQL

Data & Search: PostgreSQL • Oracle • MySQL • Redis • Elasticsearch • Typesense • ClickHouse

Cloud & Infrastructure: Azure • AWS • Kubernetes • Docker • NGINX • Cloudflare • CDN • PgBouncer

Observability: OpenTelemetry • HyperDX • ELK • Prometheus • Distributed Tracing • Structured Logging

Architecture: Distributed Systems • Microservices • Domain-Driven Design • Event-Driven Architecture • Async Processing • Multi-Tenant SaaS • Feature Flags • High Availability • API Design • Performance Engineering

EXPERIENCE

Senior Software Architect

Bajaj Finserv Health Limited / Vidal Health | November 2024 – Present | Pune, India

- Lead architecture and engineering initiatives across healthcare platforms spanning employee portals, claims journeys, benefits management, provider integrations, investigation workflows, HR platforms, and internal enterprise systems.
- Manage and mentor an engineering organization of approximately 20 engineers while remaining hands-on with system design, performance optimization, production troubleshooting, and critical technical implementations.
- Reduced critical API latency from 1500 ms to 250 ms (~83%) by decomposing monolithic database workloads into domain-aligned services, isolating read-heavy workloads, optimizing indexing, introducing Redis caching, and moving expensive processing to asynchronous workflows.
- Re-architected database and service boundaries to improve throughput, transactional isolation, horizontal scalability, fault containment, and maintainability.
- Designed PostgreSQL scaling strategies using PgBouncer, read replicas, indexing, workload isolation, Redis caching, and partitioning for high-concurrency enterprise workloads.
- Redesigned bulk-processing and data-ingestion systems using asynchronous pipelines, staged validation, retry strategies, idempotent processing, and failure-recovery mechanisms.

- Lead architecture reviews and system-design decisions across multiple squads covering service boundaries, API contracts, database architecture, caching, observability, scalability, deployment strategy, and production readiness.
- Established engineering practices around SLIs/SLOs, production dashboards, observability, runbooks, release governance, incident handling, and reliability reviews.
- Designed progressive rollout strategies using internal releases, feature flags, percentage-based rollouts, and kill switches to reduce deployment risk.
- Drive engineering standards, technical mentoring, capacity planning, architecture governance, and cross-squad technical alignment.

Co-Founder / CTO / Principal Engineer

DesignTemplate / Doographics | September 2023 – November 2024 | Pune, India

- Co-founded and engineered a cloud-native content and design platform serving users across 50+ countries.
- Architected and operated infrastructure handling 3M+ daily requests while maintaining approximately 99.9% availability.
- Designed distributed backend services, frontend platforms, search systems, caching layers, CDN architecture, asset pipelines, storage systems, and deployment infrastructure.
- Improved application performance by approximately 50% through Elasticsearch indexing, Redis caching, database/query optimization, CDN usage, and frontend performance improvements.
- Reduced infrastructure costs by approximately 80% by redesigning content delivery around CDN caching, hybrid storage architecture, optimized compute usage, and static asset distribution.
- Built automated CI/CD pipelines that improved release velocity by approximately 70%, enabling faster and more reliable production deployments.
- Designed SEO-first URL, metadata, and content architecture generating more than 200K monthly organic clicks.
- Built infrastructure capable of absorbing large traffic spikes while maintaining predictable infrastructure costs and operational reliability.
- Designed platform capabilities across authentication, subscriptions, content delivery, search, creator workflows, asset processing, payments, and multi-tenant application architecture.
- Secured approximately USD 125K in Microsoft Azure cloud credits, significantly extending infrastructure runway during platform growth.
- Owned technical roadmap, architecture decisions, infrastructure strategy, engineering hiring, mentoring, delivery planning, and production operations.

Senior Web Lead

Bajaj Finserv Health Limited | June 2021 – September 2023 | Pune, India

- Led migration from a monolithic frontend architecture to microfrontends, enabling independent deployments, clearer ownership boundaries, and modular scalability across product teams.
- Improved frontend performance and reduced load times by approximately 30% through code splitting, lazy loading, bundle optimization, dependency reduction, and improved asset-delivery strategies.
- Designed a CMS-driven dynamic page architecture, allowing business and content teams to launch and modify pages without requiring application deployments.
- Supported critical digital healthcare and transaction journeys processing more than ₹2 crore in monthly transactions, with focus on reliability, availability, and user experience.
- Built reusable React and TypeScript component libraries and established common frontend patterns across teams.
- Improved API integration architecture, error handling, resiliency, caching, and application performance for high-traffic workflows.

- Worked across frontend, backend, infrastructure, QA, product, and business teams to deliver end-to-end product initiatives.
- Mentored developers through technical design reviews, code reviews, debugging, performance analysis, and architecture discussions.

EARLY EXPERIENCE

Talent Skechers: Software Engineer | April 2021 – June 2021

Tangible Technology: Software Engineer | July 2020 – April 2021

Ninthmotion: Frontend Developer | April 2019 – June 2020

SNDH Infosolutions: Software Developer | March 2017 – April 2019

CORE EXPERTISE

Distributed Systems • Microservices Architecture • System Design • Database Decomposition • PostgreSQL Scaling • Query Optimization • Read-Heavy Workload Isolation • Redis Caching • Async Processing • Event-Driven Systems • Bulk Data Pipelines • Kubernetes • API Performance Engineering • CI/CD • Observability • Distributed Tracing • SLI/SLO Design • Reliability Engineering • Feature Flags • Progressive Rollouts • Multi-Tenant SaaS • Cloud Architecture • Engineering Leadership

SELECTED ENGINEERING IMPACT

- Reduced API latency from 1500 ms → 250 ms (~83%)
- Scaled production platforms to 3M+ daily requests
- Maintained approximately 99.9% platform uptime
- Improved application performance by approximately 50%
- Reduced frontend load times by approximately 30%
- Improved deployment/release velocity by approximately 70%
- Reduced infrastructure costs by approximately 80%
- Generated 200K+ monthly organic search clicks
- Scaled platform reach to users across 50+ countries
- Supported digital transaction journeys exceeding ₹2 crore monthly
- Secured approximately USD 125K in Azure cloud credits
- Currently lead and mentor an engineering organization of approximately 15+ engineers

SELECTED TECHNICAL WORK

AI Infrastructure & Local Inference

- Built and benchmarked local LLM inference environments using NVIDIA GPUs, CUDA, llama.cpp, vLLM, quantized models, and optimized KV-cache configurations.

Multi-Tenant SaaS Architecture

- Designed a multi-tenant platform architecture for device management, memberships, reservations, session lifecycle, dynamic pricing, and real-time device communication.

ENGINEERING LEADERSHIP

Engineering Team Leadership • Architecture Reviews • Technical Roadmaps • Mentoring • Cross-Squad Architecture • Production Readiness • Incident Management • Performance Reviews • Capacity Planning • Engineering Standards • System Design Reviews • Stakeholder Management • Developer Productivity • Technical Hiring

EDUCATION

Bachelor of Technology – Computer Science | Shivaji University, Department of Technology | July 2012 – July 2017

Open to: Remote, Global, Staff Engineer, Lead Engineer, Principal Engineer, Engineering Manager, and Senior Software Engineer, Senior Engineering Manager opportunities.