VAIBHAV JAIN

Senior Software Engineer

+919811180996 | vaibhavjain101098@gmail.com | Delhi, India | LinkedIn

SUMMARY

Senior Full Stack Engineer with 4+ years building high availability SaaS and AI enabled systems. Strong backend and system design focus with hands on experience in distributed systems, API design, CDC/data pipelines and performance optimization. Proven record of improving UX (LCP/page load), reducing cloud and LLM costs, and scaling platforms to >1M registered users while maintaining high uptime.

SKILLS & INTERESTS

JavaScript (ES6+), TypeScript, React, Next.js, NestJS, Express, Node.js, MongoDB, Firestore, ClickHouse, BigQuery, PostgreSQL, SQL, GCP, AWS, Docker, REST APIS, CI/CD, CDC / Data Pipelines, Microservices, OpenAI / LangChain, Performance Optimization

EXPERIENCE

Careerflow.ai   Jul '24 — Present
Senior Software Engineer   Delhi, India (Remote)

- Led architecture and performance work for a cloud native SaaS platform serving >1M registered users (DAU ~7.5K, MAU ~110K, peak concurrency ~1.2-1.5K)
- Reduced average page load time ~55% (4.0-4.5s → 1.8-2.0s) and LCP ~56% (3.2s → 1.4s) via API tuning, Redis caching, DB query optimization, lazy loading and code splitting
- Reduced cloud infrastructure costs ~50% and OpenAI/LLM spend ~40% over 6 months through autoscaling, resource right sizing, prompt optimization, batching and token management
- Designed and implemented CDC pipelines (Firestore → ClickHouse) to power real-time analytics dashboards for B2B stakeholders
- Built LLM-powered features (OpenAI, LangChain) handling 5K+ daily LLM requests (~150K+/month), optimizing invocation patterns to balance UX and cost
- Developed and scaled a Chrome extension integrated with 50+ job boards; improved workflow efficiency ~70% and maintained 99.8% authentication reliability
- Owned observability, alerting and incident response (Sentry, Amplitude, GA4, Slack); reduced incident detection time ~60% and supported 99.9% uptime

Jollee   Jun '22 — Jul '24
Software Development Engineer (Backend)   Gurugram, India

. Owned backend microservices processing 200K+ API requests/day; achieved p99 latency ≤200ms via caching, denormalization and parallel processing Reduced critical API response times up to 83% using Redis caching and asynchronous job processing

- Owned backend microservices processing 200K+ API requests/day; achieved p99 latency ≤200ms via caching, denormalization and parallel processing
- Reduced critical API response times up to 83% using Redis caching and asynchronous job processing
- Integrated 5+ logistics partners and Unicommerce OMS to provide real-time order tracking and reduce manual operations
- Implemented Algolia search, cutting query latency ~50% and improving discoverability
- Led cloud migration across AWS and Azure, implemented autoscaling and resilience patterns to reach 99.95% uptime and reduce operational costs ~20%

Gradeprime   Jan '22 — Jun '22
Full Stack Intern | Gradeprime   Delhi, India

- Built RESTful APIs and managed MongoDB to maintain core service availability
- Optimized MongoDB queries and handlers to reduce latency from ~180ms to ~60ms and increase throughput 3x on critical routes
- Implemented Razorpay payments (99.9% success) and OTP/email authentication to strengthen account security
- Increased test coverage from ~30% to ~75%, reducing production bugs ~60% and speeding release readiness

EDUCATION

Master of Computer Applications (MCA), Bharati Vidyapeeth's Institute of Computer Applications and Management   Apr '19 — Jun '22
                                                                                                                    New Delhi, India

Bachelor of Computer Applications (BCA), Bharati Vidyapeeth's University Institute of Management and Research   Apr '16 — Apr '19
                                                                                                                    New Delhi, India