

# Kashif Wahaj

Senior Software Engineer

LinkedIn: [linkedin.com/in/kashifwahaj](https://www.linkedin.com/in/kashifwahaj)

Email: [kashifwahaj@gmail.com](mailto:kashifwahaj@gmail.com)

Mobile: +91-9389534027

---

## SUMMARY

Senior Full-Stack Engineer with 6+ years of experience designing and scaling distributed web platforms, multi-tenant architectures, and performance-critical systems. Strong background in event-driven systems, Node.js, Kafka, microservices, and cloud-native systems on AWS. Experienced in designing high-performance APIs, implementing LLM-powered RAG pipelines, and leading architecture initiatives for production systems handling large-scale workloads.

---

## EXPERIENCE

- Etech Globals** Gandhinagar (Remote)  
*Senior Software Engineer (Full-time)* Sept 2022 - Present
  - Architected a **multi-tenant CMS platform** supporting region-based content isolation using database-level tenant segregation and middleware-driven tenant resolution, currently supporting **15+ tenants**
  - Designed a modular **section-based rendering engine** enabling schema-driven dynamic content blocks with tenant-scoped custom CSS sandboxing and automated static deployment pipeline.
  - Architected **RBAC + ABAC security model with MFA and SSO integration** supporting multi-organization access control.
  - Designed a **cloud-native quality evaluation platform** for analyzing large-scale agent-customer interactions using speech-to-text pipelines and LLM-based sentiment/compliance scoring. Solved latency issues via batching and async Kafka pipelines.
  - Planned & developed asynchronous event-driven architecture using Kafka, enabling horizontally scalable processing of **8,000+ daily interactions**.
  - Optimized performance under peak concurrency, **reducing processing latency by 40%** through **gRPC-based inter-service communication**.
  - Built a reusable front-end **design system using Radix and Tailwind aligned with WCAG 2.2 AA** standards and conducted internal architecture training sessions, mentoring engineers on distributed systems and full-stack best practices.
  - Led backend API architecture using AdonisJS/Express and integrated **LLM-powered services with Retrieval-Augmented Generation (RAG) pipelines** for automated evaluation of agent-customer interactions, implementing prompt engineering strategies and production-grade logging, monitoring, and failure recovery mechanisms.
- Tata Consultancy Services** Noida  
*Software Developer (Full-time)* July 2019 - Sep 2022
  - Led front-end architecture modernization from legacy jQuery system to React + Redux-based architecture.
  - Developed lazy loading, virtualized lists and code splitting, **reducing bundle size by 32%** and improving Lighthouse performance **score from 65 to 90+**
  - Developed multi-environment CI/CD pipeline using GitHub Actions and AWS EC2 for controlled release.
  - Improved API performance and bulk authentication workflows, **reducing response latency by 60%**.
  - Contributed to cross-functional architecture discussions for scalable SPA deployments.

---

## SKILLS

- Languages:** JavaScript, TypeScript, Java, Python (basic)
- Backend:** Node.js, Express, AdonisJS, Spring, gRPC, Kafka, RabbitMQ, Redis, REST APIs, Laravel
- Front-end:** React, Redux, Tailwind CSS, Virtualized rendering, Cypress, Jest
- Databases:** MySQL, MongoDB, Cassandra, SQLite
- Cloud & DevOps:** AWS (EC2, S3, RDS), Docker, CI/CD, Nginx, Linux
- Distributed Systems & Observability:** Event-driven architecture, Prometheus, Grafana, Loki, Jaeger
- AI & Tools:** LLM Integration, RAG Pipelines, Multi-agent workflows, n8n, WebRTC, WebSockets

---

## SELECTED ENGINEERING PROJECTS

- Workflow Orchestration Engine:** Built event-driven workflow engine with pluggable execution nodes supporting DAG-based execution. Powered asynchronous pipelines for speech-to-text (STT) and downstream processing workflows, enabling scalable handling of audio-to-text transformations. Designed retry policies, idempotent job execution, failure recovery, background processing, and dead-letter queues using message queues.
- Distributed Chat System (Peer Coordination):** Engineered an end-to-end encrypted **WebRTC-based** communication platform with peer coordination logic, approval-based access control, and real-time signaling using WebSockets. Designed architecture to handle connection state management and fault tolerance, along with embeddable web components enabling seamless integration across multiple client applications.

---

## EDUCATION

- University Institute of Engineering and Technology** CSJM University, Kanpur  
*Bachelor of Technology - Information Technology; GPA: 8.5* 2015 - 2019