

MAHENDRA SINGH CHOUDHARY

(+91)7487038824 · cmahendra9999@gmail.com · [linkedin.com/in/mahendra2890/](https://www.linkedin.com/in/mahendra2890/) · github.com/mahendra2890

Senior Backend Engineer with 4 years in architecting reliable, highly available, scalable distributed systems, AI integration, and multi-cloud infrastructure optimization. **I turn complex problems into simple solutions.**

PROFESSIONAL EXPERIENCE

SaxeCap | Senior Software Engineer | San Francisco, California, US - Remote 2026 – Present

- Built an AI workflow agent streamlining postcall processing for 10 million calls a year

DelightChat | Senior Backend Engineer | Arden, Delaware, US - Remote 2024 – 2026

- Architected and built custom Celery **job scheduler**, reducing server memory usage by **75%+** and avoiding a multi-node t4g.large→t4g.xlarge upgrade (saved **~15%** of infrastructure cost)
- Ideated and developed** from scratch an **infrastructure monitoring and alerting system** for 100s of AWS/Grafana/Dynatrace metrics, with preemptive **automated** slack alerts preventing downtime and improving **service reliability** - reduced major outage/slowdown events to near zero
- Designed and developed language-agnostic AI capabilities generating **+\$50k** ARR, including a RAG-based customer support agent processing 15k queries/month and text transformation tools 50k+ queries/month
- Delivered payment gateway integrations and COD-to-prepaid conversion automation, ensuring **\$20M+** in at-risk orders were converted to prepaid for DelightChat users
- Engineered R2-first storage with seamless S3 failover, **cutting infra costs 20%** (~\$10k/year) and sustaining almost 100% uptime during outages
- Reduced peak **RDS CPU load by over 70%** through efficient database connection management
- Overhauled authentication** system and enhanced **security** with optimized flows, rate limits, and IP based frequent-access limits, reducing credential compromise incidents to zero
- Tech Stack:** Python, Django, DRF, AWS (CloudWatch, Lambda, S3, EC2, RDS, SNS, SQS), GraphQL, REST APIs, Redis, Celery, Jenkins, Dynatrace, Shell Scripting, RAG, AI Agents, Agentic AI, Cursor, OpenAI, Cloudflare(R2, Sippy, on demand migration between S3 and R2), Pusher, Containerisation

DocuEdge - Deloitte USI | Software Development Engineer | Bengaluru 2022 – 2024

- Built backend and orchestration layer from scratch for **multi-cloud, headless** document classification service, achieving **82%+ success rate** and **67% improvement** over client's previous system
- Built an E2E lossless data migration pipeline, archiving **100+ TB** from SFTP to Azure Data Lake.
- Developed an automated testing tool for decision models, **reducing testing time by 99%+**
- Tech Stack:** Python, Django, Drools engine, Java, Spring Boot, Pandas, GCP, AWS, Azure, Oracle ATP

IIT Jodhpur | Undergraduate Student 2018 – 2022

- Graph Optimization: Hybrid PageRank achieving 12X speedup using **parallel processing**
- Secure chat app, CLI health center tool and interactive maze pathfinder with algorithm visualization
- Tech Stack: Python, CUDA, C++, Shell, GPUs, Parallel Programming, React, Js, Docker, OpenMP

SKILLS AND CORE COMPETENCIES

- Backend Development:** Python, Django, DRF, PostgreSQL, Redis, Celery, API Design, Microservices
- Cloud & Infrastructure:** AWS, Cloudfare, GCP, Azure, Docker, Linux, Git, Jenkins, CI/CD, Grafana, Mixpanel
- AI & Automation:** RAG, Agentic AI, Workflow AI Agents, Cursor, LLMs, Recommendation engines
- Specializations:** System Architecture, Performance, AI Integration, Monitoring & Alerting, Security, Orchestration, Cloud Cost optimisation, Payment Gateways, Parallel Programming
- Cross-Domain:** Fintech, Healthcare, E-commerce, SaaS | Startup to Enterprise | IC to Team Lead

EDUCATION

Indian Institute of Technology, Jodhpur - B.Tech in Computer Science and Engineering 2018 - 2022