

ANIRUDH MAKULURI

602-880-5334 | anirudh.makuluri@gmail.com | linkedin.com/in/anirudh-makuluri | makuluri.com

SUMMARY

Full-stack software engineer with early-stage startup experience building developer tools, AI agents, and cloud infrastructure. Scaled a real-time collaboration platform to 12K+ users and created SmartDeploy, an AI-native cloud deployment platform that analyzes repositories, generates deployment blueprints, and executes AWS deployments. Strongest in backend systems, cloud workflows, real-time applications, and AI-assisted developer tooling.

EDUCATION

M.S Computer Science <i>Arizona State University, Tempe, AZ (ASU)</i>	August 2024 – May 2026 GPA: 4.00/4.00
B.E. Electronics & Instrumentation <i>Birla Institute of Technology and Science Pilani (BITS Pilani)</i>	August 2018 – May 2022

PROFESSIONAL EXPERIENCE

HomeHudl — Software Engineering Intern Implemented a frontend API proxy that routes client requests through the application server, hiding backend service URLs from browser-exposed code and improving environment isolation across deployments.	July 2026 – Present
Shardings — Full-Stack Software Engineer (Early Engineering Team) - Scaled real-time collaboration workflows to 12K+ users by refactoring meeting state from instance-local memory to Redis-backed shared state, enabling stateless Node.js servers to recover room context after reconnects, refreshes, and late joins. - Reduced idle compute cost by introducing autoscaling for production GCP e2-medium app servers, replacing always-on capacity with a 1–3 instance policy while preserving availability during peak traffic. - Built a retry-safe asynchronous notification pipeline using Cloud Tasks, Cloud Run, and Nodemailer to deliver approved-user notifications, newsletters, and AI-generated meeting recaps, handling 300+ emails/day. - Migrated dashboard and meeting workflows from legacy vanilla JavaScript to React by building 15 reusable React + Tailwind components, improving UI consistency and accelerating feature delivery across web collaboration flows. - Accelerated critical mobile bug-fix rollouts by implementing CodePush OTA updates for Android and iOS, reducing JavaScript-layer fix delivery from app-store review cycles to under 5 minutes.	June 2022 – June 2024
Shardings — Software Engineer Intern - Streamlined full payment gateway with Razorpay and Stripe integrations supporting 3 pricing tiers (Basic, Standard, Pro); designed secure webhook handling, idempotency, and error recovery, handling 100+ transactions per month with 99% success rate. - Introduced Google Analytics to track 8+ user interaction and error events (room creation, joins, failures), improving observability into frontend behavior and enabling faster debugging of user-facing issues.	July 2021 – March 2022

PROJECTS

SmartDeploy — AI-Native Cloud Deployment Platform - Improved deployment transparency by building a Next.js deployment control plane that converts repository scans into reviewable blueprints, letting users inspect services, routing, build artifacts, environment setup, live logs, run history, and runtime health before deploying. - Decoupled deployment execution from the main WebSocket control plane by routing deploy runs through SQS FIFO, Lambda, and one-off ECS Fargate workers, isolating long-running build/deploy jobs while streaming live status across 140+ tracked runs. - Introduced an observable deployment agent with tools for querying deployment state, runtime health, logs, historical runs, and deployment documentation; tracked latency, tool calls, and outputs to evaluate reliability and refine prompts/tool definitions.	July 2025 – July 2026
Accio — Document Intelligence & RAG System - Built a system (FastAPI, Streamlit) to extract structured entities and relationships from 2.8K+ pages of complex unstructured PDFs into a Neo4j knowledge graph, enabling queryable insights across regulatory data. - Designed an evaluation pipeline (precision, recall, F1) to measure extraction quality, achieving 77.5% F1 and improving retrieval accuracy for downstream AI workflows and decision-making systems.	September 2025 – December 2025

OPEN SOURCE & AWARDS

- Open Source (MOSS)** — Contributed multiple merged pull requests to **MOSS**, an open-source retrieval layer for AI systems, spanning Go SDK development, documentation improvements, and deployment-related bug fixes.
- 1st Place (Auth0 AI Agents) & 2nd Place (Google Track)** — Innovation Hacks 2.0, ASU (600+ participants), April 2026. Developed *TravelMate*, an AI travel companion with a real-time voice interface for itinerary assistance.

TECHNICAL SKILLS

Languages: TypeScript/Javascript, Python, Go, Java, SQL
Backend/Systems: Node.js/Express, FastAPI, Spring Boot, REST APIs, WebSockets, Socket.IO, Microservices, Redis
Cloud/DevOps: AWS (ECS, Lambda, SQS, CodeBuild, ECR, ALB, S3, CloudWatch), GCP, Docker, Terraform, GitHub Actions CI/CD
Databases: PostgreSQL/Supabase, Firestore, MongoDB, DynamoDB, Neo4j
AI/LLM: RAG, embeddings, vector search, LangChain, LangGraph, AWS Bedrock, MCP, TensorFlow, PyTorch
Frontend/Mobile: React, React Native, Next.js, Zustand, TanStack Query, Redux, Tailwind CSS
Certifications: Oracle Generative AI Professional, Machine Learning Specialization (Stanford/DeepLearning.AI)