

Anirudh Herady

602-793-8590 | aherady@asu.edu | linkedin.com/in/anirudhherady | github.com/avh17 | anirudhvportfolio.vercel.app

PROFESSIONAL SUMMARY

Full-stack engineer specializing in scalable backend systems and AI applications. Delivered production microservices for video streaming and intelligent chatbot solutions. Excel in fast-paced, high-impact technical environments.

EDUCATION

Arizona State University	Tempe, AZ
<i>M.S. Computer Science – GPA: 3.78/4.0</i>	<i>Aug. 2024 – May 2026</i>
Manipal Institute of Technology	Manipal, India
<i>B.Tech. Computer and Communications Engineering</i>	<i>Jul. 2018 – Jul. 2022</i>

TECHNICAL SKILLS

Languages: Python, JavaScript, TypeScript, SQL, C/C++, Bash

Backend Engineering: FastAPI, Node.js, Flask, RESTful APIs, GraphQL, Microservices Architecture

Cloud & DevOps: AWS (Lambda, EC2, S3, SQS, ECR), Docker, CI/CD (Github Actions), Linux

AI & Machine Learning: RAG (Retrieval-Augmented Generation), LangChain, LLMs (GPT, Gemini), Vector Embeddings

Databases: PostgreSQL, MongoDB, Redis, SQLite, SQLAlchemy ORM

Frontend: React.js, Next.js 14, Tailwind CSS, Streamlit

PROFESSIONAL EXPERIENCE

Software Intern

Jul. 2025 – Aug. 2025

Potters Tech

Remote

- Delivered a production-ready Location Intelligence Chatbot in 8 weeks that eliminated the need for analyst support, enabling 100+ non-technical users to independently query geospatial datasets across 50+ cities.
- Improved chatbot response accuracy by 60% through architecting a custom RAG pipeline that combined PostgreSQL retrieval with GPT-4 synthesis, reducing average query resolution time from 5 minutes to under 10 seconds.
- Integrated OpenStreetMap APIs and real-time location services to deliver 95% accurate POI search results across 2,000+ locations, handling 500+ daily queries with sub-second response times.

Associate Software Developer

Aug. 2022 – Sept. 2023

Valtech India

Bengaluru, India

- Built and deployed 8 FastAPI microservices with 20+ REST/GraphQL endpoints that supported a multi-tenant video streaming platform serving 10,000+ concurrent users with 99.9% uptime.
- Reduced Docker image sizes by 60% and cut deployment time from 15 minutes to 5 minutes by implementing multi-stage builds, accelerating the team's CI/CD pipeline and saving 2 hours per day across 5 developers.
- Optimized database performance by 40% through designing normalized PostgreSQL schemas and MongoDB compound indexes, reducing query latency from 800ms to 480ms for high-volume user data operations.
- Secured platform access for 3 distinct user permission levels by implementing JWT and OAuth2-based RBAC, preventing 15+ potential security vulnerabilities identified during penetration testing.
- Mentored 3 junior developers through 20+ code reviews and weekly onboarding sessions, reducing new hire ramp-up time by 30%.

PROJECTS

Distributed Face Recognition System | AWS (Lambda, EC2, SQS, S3), Python, Docker

- Designed two serverless architectures (Lambda vs EC2) for video frame processing that achieved 99%+ recognition accuracy at \approx 3s latency under 100+ concurrent requests, providing cost-performance analysis for production deployment.
- Developed a Python autoscaler that dynamically provisions 0-15 EC2 instances based on SQS queue depth, reducing infrastructure costs by 45% during off-peak hours while maintaining performance during traffic spikes.
- Ensured 99.5% system reliability during burst traffic by decoupling ingestion and processing layers with SQS, successfully handling 10x traffic surges without service degradation.

StoreIt: Cloud Storage Solution | Next.js 14, TypeScript, Appwrite

- Built a full-stack file management platform with secure authentication and storage using Appwrite, supporting 50+ concurrent users with real-time file upload/download capabilities.
- Developed 20+ reusable React components with dynamic storage analytics charts, reducing UI development time by 40% for future features and improving user engagement by 25%.