Srujay Reddy Vangoor

Location: Sacramento, CA
Phone: +1 (916) 940-5986

Email: srujayreddyv@icloud.com
LinkedIn: linkedin.com/in/srujayreddyv
GitHub: github.com/srujayreddyv

Summary

- Full Stack Software Engineer with 6+ years of experience building scalable web applications and cloud native backend systems. Strong in Python, FastAPI, React, and AWS, with expertise in system design, interoperability, and resilient distributed architectures. Known for modernizing legacy systems, designing high throughput distributed services, and delivering production GenAI systems including RAG pipelines and multi agent workflows that automate compliance and operational processes.

Skills

- **Backend & APIs**: Python (FastAPI, Django), Node.js (Express), Go (Golang), C# (ASP.NET Core), REST, GraphQL, WebSockets
- **Frontend & Mobile**: React, Next.js, TypeScript, JavaScript, React Native
- **Cloud & DevOps**: AWS, Azure, Docker, Kubernetes, Terraform, CI/CD (GitHub Actions, Azure DevOps), CloudWatch, Grafana, Prometheus
- **Data & Analytics**: PostgreSQL, MS SQL Server, DynamoDB, Redis, Snowflake, Amazon Redshift, Airflow, Kafka, Spark, Elasticsearch, Power BI, Tableau
- **AI & ML**: AWS Bedrock, RAG Pipelines, LangChain, CrewAI, Hugging Face Transformers, Vector Databases (FAISS, Pinecone, OpenSearch), PyTorch, TensorFlow, MLflow, Scikit learn

Work Experience

- **Software Engineer** — Sacramento, CA (Hybrid)
  California Department of Developmental Services (DDS) — *Feb 2025 – Present*
  - Built full stack applications serving internal and external users using FastAPI, Express, React, and NextJS on Aurora PostgreSQL, implementing RBAC, Azure AD SSO, and Azure DevOps CI/CD deployments to AWS ECS.
  - Scaled FastAPI microservices handling 20K+ daily requests on ECS and Lambda, implementing Cognito based authentication and EventBridge driven workflows to maintain 99% uptime for mission critical healthcare integrations.
  - Modernized legacy COBOL, JCL, and Db2 batch systems into Python based event driven services aligned with FHIR models, reducing nightly processing time by 60% and eliminating 200+ manual operational steps.
  - Built and deployed a React and TypeScript portal on Terraform provisioned AWS infrastructure, reducing AI Assessment entry time from 30 to 10 min through automation and UX optimization.
  - Reduced analyst effort by 85% by engineering a production LLM powered QA platform using Bedrock with document ingestion, semantic retrieval, and grounded query answering for claims and regulatory review workflows.
  - Developed a multi agent GenAI chatbot using FastAPI, NextJS, Bedrock, and CrewAI across four knowledge bases, implementing routing logic and delivering source grounded responses for healthcare and compliance use cases.
- **Software Engineer** — Sacramento, CA (Remote)
  California Department of Conservation (DOC) — *Jan 2024 – Feb 2025*
  - Built a production seismic hazard modeling platform using Python and OpenQuake with supporting services in Perl and PHP, delivering ground motion predictions for California fault zones with response times under 3 seconds.
  - Designed and automated large scale seismic data pipelines processing 10+ GB per day, improving ingestion efficiency and reducing data retrieval time by 50% to support high throughput detection workflows.
  - Strengthened security for 60+ remote monitoring stations by implementing secured API Gateway endpoints with IAM based access control and Secrets Manager driven credential management, passing compliance audits with zero findings.
  - Improved reliability of geospatial services by resolving ArcGIS server performance bottlenecks and optimizing map rendering with GeoServer and Leaflet.js, reducing downtime and increasing system stability.
- **Software Application Developer Intern** — Sacramento, CA
  Population Research Center (PRC), Sacramento State — *Jan 2023 – Jun 2024*
  - Built a staff management system using C#, Entity Framework, and SQL Server, optimizing stored procedures, indexing, and query performance to reduce administrative workload by 40%.
  - Designed and automated health data ETL pipelines using .NET, SQL Server, and PowerShell, migrating legacy Access and Oracle datasets into centralized SQL analytics stores for Power BI reporting under HIPAA compliant workflows.

- Developed interactive data entry tools and analytics dashboards for statewide public health studies, increasing data collection efficiency by 60% through automated validation and data quality checks.
- Supported high volume CATI survey operations by implementing system upgrades and PowerShell diagnostic utilities, reducing survey delays by 30% while maintaining HIPAA compliance.

Software Engineer Hyderabad, IN
Human Sciences Research Group (HSRG), IIIT-H *Jan 2021 – Aug 2022*

- Designed and built a cloud native archival and data processing platform for social media and news streams, enabling realtime analysis of over 1 million data points during the Indian Farmers Protests.
- Fine tuned multilingual BERT and IndicBERT models using Hugging Face Transformers for sentiment analysis and topic modeling, improving classification accuracy by 8% over baseline models.
- Developed secure REST APIs using C# and ASP.NET Core to support high volume data ingestion and metadata management, with React based dashboards for exploratory analysis.
- Containerized services with Docker and implemented automated build, test, and deployment pipelines using GitHub Actions, reducing release cycles by 40% and improving environment reproducibility.

Software Engineer Hyderabad, IN
Change and Continuity in Spiti Valley, ICSSR Sponsored Project *Jul 2018 – Dec 2020*

- Engineered a cross platform GIS platform using Python, Django, PostgreSQL, and PostGIS to map 280+ heritage sites across 25 villages, enabling spatial search, analytics, and interactive visualization.
- Built responsive React web and offline first React Native mobile clients with Leaflet.js, achieving 60% faster load times in low bandwidth field environments.
- Designed and deployed REST APIs with Django REST Framework to support spatial and attribute based queries with pagination and caching, reducing query latency and improving scalability.
- Developed Python based ETL and spatial data processing workflows integrating survey, census, and archival datasets into a unified geospatial repository, improving metadata accuracy by 40% and enabling near realtime updates.

Education

California State University, Sacramento Sacramento, California
Master of Science (MS) in Computer Science *Aug 2022 – Jan 2025*

International Institute of Information Technology (IIIT-H) Hyderabad, India
Bachelor of Technology (BTech) in Computer Science *Aug 2015 – July 2019*

Relevant Projects

FastChat - Real-Time Chat Application

- Built a Dockerized full stack app with FastAPI backend, React frontend using TypeScript and Material UI, and PostgreSQL, supporting WebSocket based realtime messaging, presence, and typing indicators.
- Implemented JWT authentication, rate limiting, input validation, structured logging, and CI driven integration tests with PyTest and GitHub Actions, plus performance monitoring for production reliability.

Local Retrieval Augmented Generation (RAG) System

- Built a local RAG pipeline using Hugging Face embeddings and Pinecone vector search to enable semantic retrieval across a 1,200 page corpus, achieving 93% top k retrieval relevance on evaluation queries.
- Designed an MLflow based evaluation framework to benchmark retrieval quality, optimize chunking and embedding strategies, and refine prompt grounding for production ready deployment.

TaskHub - Task Management Platform

- Architected a full stack system with a modular React NextJS SPA and Express based microservices backend secured with JWT authentication and PostgreSQL, enabling realtime collaboration across 100+ workspaces.
- Implemented Redis caching, load balancing, and blue green deployments via Azure DevOps pipelines, achieving 99% uptime and reducing API latency by 70% under concurrent traffic.

Sacverse - Augmented Reality (AR) Campus Tour App

- Designed an AR based campus navigation prototype with geospatial search, route planning logic, and low bandwidth optimized client rendering workflows for mobile environments.
- Conducted structured usability testing with 50+ pilot users, iterating on spatial overlay accuracy and interaction flows, resulting in a 40% improvement in measured usability metrics.