PRANAV REDDY GADDAM

San Jose, CA | +1 (925) 605-9293 | reddy.pranav.gaddam@gmail.com | LinkedIn |

Profile

Software engineer specializing in AI platform development and distributed systems, with hands-on experience implementing LLM-based solutions and high-scale data processing systems. Proven track record building production-ready applications using Python, JavaScript, and cloud infrastructures.

Experience

Data Engineer | *VE Projects Pvt Ltd* **Aug 2023 – Jul 2024**

- Implemented a database migration system from **Oracle** to **Amazon Redshift**, optimizing storage and compute resources to achieve $12,000 annual cost savings and **12%** improvement in system performance.
- Built and optimized real-time data processing application handling **15M+** daily records from **30+** sources using **Apache Kafka**, **PySpark**, cloud infrastructure, reducing manual intervention by **29%** monthly.
- Collaborated with **product** and **business** teams to develop automated quality validation systems using **SQL**, **Python**, and **Apache Airflow**, improving data accuracy by **60%** and accelerating BI report generation by **3x**.

Projects

QuizForge : Multi-Agent Learning Platform **Jun 2025 - Jul 2025**

- Built AI-powered platform using **OpenAI GPT-4,** Next.js 14, **TypeScript**, TailwindCSS, and FastAPI, architected to handle **500+** concurrent quiz sessions with real-time analytics processing.
- Integrated **Cerebras** hosted LLMs and **MongoDB** with async drivers to process **10,000+** quiz questions per minute through optimized indexing pipelines.
- Designed production-grade architecture with clean HTML rendering, async I/O optimization, and comprehensive unit testing using **Jest** and **pytest**, ensuring **99%** uptime during stress testing.

GitBridge: Turning GitHub Repos into Diagrams, Podcasts, and Conversations **Jul 2025 – Aug 2025**

- Engineered an AI platform processing **3GB+** repositories with diagram generation in **<6s** using **Mermaid.js** and podcast creation in **<3 minutes** with real-time streaming, using **ElevenLabs**.
- Optimized conversational AI using **Qwen-32B** LLM with **sub-200ms** response latency, enabling real-time developer interactions with complex codebases and commit histories.
- Developed a full-stack system using **FastAPI** and **Vite.js** + **TailwindCSS**, deployed via Docker on AWS **EC2**, concurrent processing capabilities.

Technical Skills and Certifications

- **Languages** : Python, SQL, JavaScript, TypeScript, HTML
- **Tools** : Docker, Kubernetes, MongoDB, PostgreSQL, Git, Jest, pytest, Jenkins
- **Technologies/Frameworks** : React, Next.js, FastAPI, Django, AWS, GraphQL, RESTful APIs, Object-oriented programming, SQLAlchemy
- **Achievements** : **Winner**, SpartUp Spring Hackathon 2025, hosted by **Cisco** and **SJSU**
- **Certifications** : AWS Certified Cloud Practitioner, AWS Certified AI Practitioner
- **Patents and Publications**: Patent on ML Techniques for hate speech detection (Patent No. 202341050260A); Published Feature Extraction for student classification at ICIRTCS-23 (Paper ID: ICIRTCS-23-033)
- **Extracurriculars** : SMEC Technology Awareness Month Director of Operations

Education

San Jose State University | M.S., Data Analytics | ***CGPA 3.8/4.0*** **Aug 2024 - May 2026**

- **Coursework:** Big Data, Machine Learning, Data Warehouse & Pipeline , Distributed Systems

St. Martin's Engineering College | B. Tech in Computer Science and Engineering **Aug 2019 - May 2023**

- **Coursework:** Database Management Systems, Distributed Systems, Data Structures and Algorithms