Jinay Shah

Danbury, CT | (203) 512-6561 | jinay.y.shah@outlook.com | linkedin.com/in/jinay24/ | github.com/jinayusa

SUMMARY

Backend Engineer with 3+ years of experience building high-scale systems across finance, healthcare, and public policy. Delivered measurable impact from cutting API latency by 60% to deploying ML-based identity systems in production. At ASU's Decision Theater, I drive real-time backend optimization for platforms used by 2,000+ stakeholders. Specializing in debugging production schema issues, optimizing APIs for scale.

SKILLS

Languages: Python, Java, SQL, JavaScript, TypeScript
Frameworks: Flask, FastAPI, Spring Boot, Django, Node.js
Cloud & DevOps: AWS (EC2, Lambda, S3), Docker, Kubernetes, Jenkins, GitHub Actions
Data & Infra: PostgreSQL, MongoDB, Redis, Kafka, Pandas, TensorFlow
Engineering Patterns: REST APIs, WebSockets, Caching, Rate Limiting, Circuit Breakers, CI/CD
Certificates: AWS Academy Data Engineering, Algorithmic Toolbox, AWS Fundamentals: Going Cloud-Native

EXPERIENCE

Decision Theater Network (ASU) | Software Developer — Tempe, USA | June 2024 – Present

- Re-architected Minerva's backend to enable **interactive** and **machine-readable economic relationships** (**PIVI**); **APIs** now power **interactive trade visualizations** across countries and sectors.
- Diagnosed and resolved schema drift in CuRVE's **analytics API** that silently broke **JSON exports**, pushing a **live patch within 24 hours** to preserve **stakeholder delivery timelines**.
- Designed CuRVE's **real-time data layer (Flask + Redis + SQLAlchemy**), slashing latency from **250ms to 100ms**, and scaling to **1,400+ RPS** with auto-invalidation **cache logic**.
- Tuned high-cost SQL endpoints using **EXPLAIN**, **optimized joins**, and applied **targeted indexing**, resulting in a **2x performance boost** for high-traffic dashboards.
- Automated deployment pipelines with **GitHub Actions + rollback support**, boosting release frequency from **1 to 2–3 per week** without **production downtime**.

KiVee Softech | Senior Software Developer — Surat, India | January 2022 – July 2023

- Designed **voice authentication APIs** using **MFCC + TensorFlow**, improving **auth speed by 70% (800ms to 240ms)** and reducing **false accept/reject rates by 30%**.
- Restored ML model accuracy post-deployment by detecting voice sample drift and retraining on domain-specific noise profiles. Added OTP fallback implemented via Twilio + Redis expiry store to handle inference failures.
- Introduced **Kafka queues** to **asynchronously process non-critical flows**, significantly increasing **throughput** and **halving request congestion**.
- Built internal retry queues and **circuit breakers** in **distributed microservices**, reducing **failure spikes under high load by 35%** and improving **request success consistency**.
- Mentored 5+ junior devs and enforced **structured PR reviews** that halved **post-merge bug rates** and increased **team story point throughput by 30%** within two quarters.

Kintu Designs | Software Developer — Surat, India | February 2021 – January 2022

- Developed a personalized food recommendation system using **NLP (SpaCy) + user clustering**; **scaled to 300K+ daily queries** within **2 weeks of viral launch**.
- Tuned backend to maintain **sub-200ms response latency** during peak traffic via **API optimization**, async batching, and **clustering cache layers**.
- Closed the loop with real-world user health data, resulting in improved diet match rates and a **20% boost in user-reported satisfaction**.

PROJECTS

Image Recognition as a Service — AWS, Python, Node.js, React.js

- Architected an image classification backend with GPU-accelerated image classification using AWS EC2 (G4dn) with Lambda orchestration, reducing inference latency by 40% (300ms to 180ms) for large payloads.
- Built a live operations dashboard in React to monitor performance, failure rates, and classifier accuracy.

RiskPulse: Real-Time Financial Fraud Analytics | GitHub — Python, Flask, Kafka, MongoDB

- Engineered a fraud detection pipeline processing 5,000+ events/sec with sub-second latency, using Kafka + multithreaded Flask workers.
- Blended logistic regression with rules for fast, explainable fraud detection; enabled live model hot swapping.

EDUCATION

Arizona State University — Tempe, USA
Master of Science in Information Technology | GPA: **4.0** — Aug 2023 – May 2025

Uka Tarsadia University — Bardoli, India
Bachelor of Engineering in Computer Engineering | GPA: **3.63** — July 2018 – May 2021