

Siddhesh Save

+1 (315) 696-1289 | siddhesh.save05@gmail.com | LinkedIn | GitHub | Portfolio

SUMMARY

Software Engineer with 5+ years of experience building scalable microservices, distributed and cloud-native architectures, and high-performance backend solutions. Skilled in Python, Java, React, TypeScript, Spring Boot, and Hibernate, with proven success across financial services, IoT, and insurance domains. Adept at delivering Generative AI pipelines, LLM-based NLP systems, and ML-powered fraud detection to improve business outcomes. Strong expertise in Microsoft Azure, AWS, and Google Cloud Platform (GCP), with deep knowledge of Docker, Kubernetes, Terraform, and CI/CD automation to deliver secure, reliable, and production-ready systems.

TECHNICAL SKILLS

Programming Languages: Python, Java, JavaScript, TypeScript, Go, C++, C#, SQL, Bash

AI & ML: Generative AI, LLMs, Multi-Agent Systems, Model Deployment & Evaluation, NLP, TensorFlow, PyTorch, Scikit-learn, Pandas

Frameworks & Libraries: Spring Boot, Hibernate/JPA, Node.js, Express.js, FastAPI, Django, React, Next.js, Redux, Tailwind CSS

Cloud Platforms: Microsoft Azure, AWS (EC2, S3, RDS, Lambda, CloudWatch, EKS, IoT Core), Google Cloud Platform (GCP)

DevOps & CI/CD: Docker, Kubernetes, Terraform, GitHub Actions, Jenkins, Azure DevOps, AWS CDK, AWS CodePipeline, SonarQube

Databases & Storage: PostgreSQL, Redis, MongoDB, Amazon RDS, Amazon S3, ELK Stack

Testing & Automation: JUnit, PyTest, Mockito, Postman, API Testing, Selenium, Automated UI Testing

Tools & Project Management: Git, Jira, Linux, RESTful APIs, Bloomberg & Reuters APIs, Agile/Scrum, CloudWatch Monitoring

PROFESSIONAL EXPERIENCE

Software Engineer, AI Systems | State Street, USA

Jan 2025 – Present

- Developed and deployed GenAI-driven analytics pipelines as part of AI-backed portfolio systems, leveraging LLMs, NLP, TensorFlow, and PyTorch to improve investment risk analysis accuracy by 35%.
- Engineered and optimized scalable microservices supporting AI-backed portfolio systems with Node.js, React, and TypeScript, reducing portfolio reporting delays by 45% and enhancing system responsiveness.
- Designed, containerized, and orchestrated secure backend services on Microsoft Azure using Azure Kubernetes Service (AKS), Azure App Services, Azure Storage, and Terraform, improving scalability, fault tolerance, and reliability of financial data pipelines.
- Integrated PostgreSQL databases with real-time Bloomberg and Reuters APIs, cutting recalculation latency by 70% and providing faster, actionable investor insights.
- Automated and standardized CI/CD pipelines using Jenkins, Azure DevOps, and GitHub Actions, accelerating deployments by 50% while ensuring security and quality compliance.

Software Engineer | Hexaware Technologies, India

May 2021 – Jul 2023

- Designed and deployed ML-based fraud detection and claims optimization models using Python, Scikit-learn, and NLP, reducing false positives by 30% and accelerating claim settlements by 65%.
- Implemented Java 17 Spring Boot microservices with Hibernate/JPA and delivered intuitive, responsive React dashboards, enhancing usability and efficiency in trading and insurance operations.
- Secured and optimized APIs for financial transactions and fraud detection using Azure API Management and Event Grid, reducing transaction processing time by 70%.
- Managed PostgreSQL, MongoDB, and Redis for structured, unstructured, and in-memory data, enabling faster analytics, seamless queries, and responsive UIs.
- Containerized and monitored deployments with Azure DevOps, Docker, AKS, and SonarQube, achieving 99%+ uptime and improving code quality and system resilience.

Software Engineer | Cyient, India

May 2019 - Apr 2021

- Architected IoT-based asset monitoring microservices using Java, Spring Boot, and Hibernate/JPA, delivering interactive React dashboards for predictive maintenance and operational visibility.
- Engineered and fine-tuned ML-based anomaly detection models with Python and Scikit-learn, reducing equipment downtime by 25% through proactive monitoring.
- Built and secured REST APIs integrated with AWS EKS, S3, and CloudWatch, enabling real-time IoT data streaming and monitoring for industrial clients.
- Streamlined and automated CI/CD pipelines using Docker and Jenkins, cutting release cycles by 40% and improving deployment reliability and traceability.
- Enhanced and accelerated API performance through asynchronous processing and query tuning, reducing latency and improving real-time dashboard responsiveness.

EDUCATION

Master of Science in Computer Science | Syracuse University

May 2025

Bachelor of Technology in Computer Engineering | NMIMS University

May 2021

PROJECTS

Legal Clarity Agent | Python, Google ADK, Generative AI, Multi-Agent Systems, Agent Orchestration

- Orchestrated a privacy-first, multi-agent legal analysis system using Google ADK and Generative AI, processing 20+ clauses per document with 6 specialized agents for clause classification, risk assessment, and plain-English explanations via parallel workflows.

IoT-Enabled Driver Safety & Crash Detection System | Python, AWS Lambda, AWS S3, Twilio, Wokwi, GPS, Accelerometer

- Created an IoT-based driver safety system integrating accelerometer, GPS, AWS Lambda, and Twilio SMS alerts, reducing emergency response time by 40% through real-time crash detection and automation.

Flight Arrival Prediction | Python, Flask, Machine Learning, REST API, Dashboard Visualization

- Programmed and deployed an ML-based flight delay prediction model with a Flask REST API, improving prediction accuracy by 20% and enabling real-time visualization validated on 150+ live datasets.