Aditya Sugandhi

+1 448 500 6857 | Tallahassee, FL | adityasugandhi.dev.ai@gmail.com | GitHub | adityasugandhi.com

SUMMARY

Software Engineer with 5+ years building **AI data platforms** and **large-scale data infrastructure**. Expert in **Python, Java**, and **distributed systems** with production experience in **data pipeline automation**, **dataset lifecycle management**, and **AI/ML workflow orchestration**. Delivered platforms processing **50M+ events/day** with **99.99% availability**.

WORK EXPERIENCE

Software Engineer II — Nov 2024 – Present
Florida State University — Utility Digital Twin Platform — Tallahassee, FL

- Architected **high-throughput data ingestion pipeline** using **Java** and **Apache Kafka**, processing **real-time telemetry from 500K+ IoT sensors** (meters, HVAC, cooling coils) with **5M+ daily events** for digital twin modeling and predictive maintenance.
- Designed **event-driven microservices** using **Spring Boot** and **Kubernetes**, handling **time-series data ingestion** with **99.99% availability**, supporting **asset health monitoring** and real-time compliance reporting across **172+ networks**.
- Built **compliance monitoring service** using **OPA (Open Policy Agent)** scanning **10K+ asset configurations** daily against **IEEE standards** and **ISO 55000**, generating automated compliance violation reports with **98.5% compliance rate**.
- Implemented **data governance tooling** for **PII detection and handling**, **data lineage tracking**, and **audit logging**; created APIs enabling **dataset discovery** and **secure sharing** across **10K+ utility engineers** globally.
- Deployed **Redis caching** for distributed asset state and optimized **PostgreSQL horizontal sharding** for **500M+ telemetry records**, reducing query latency by **45%** for time-series analytics dashboards.

Machine Learning Engineer — Jan 2023 – Aug 2024
Prof. Olmo Zavala Romero — FSU — Tallahassee, FL

- Built **AI data pipelines** processing **1TB+ datasets** using **Python** and **Spark** for ML model training; implemented **data validation** and **quality monitoring** with **observability dashboards** (Grafana, MLflow).
- Developed **automated data lifecycle workflows** on **Kubernetes** for **dataset ingestion, transformation, and versioning**; optimized **retry patterns** and **idempotent processing** achieving **50% pipeline runtime reduction**.
- Created **Python SDKs** for **training data preparation** and **synthetic data generation**; built **evaluation harnesses** for model quality assessment with **vector embeddings** and **semantic search** capabilities.

Software Engineer I — Oct 2020 – Jul 2022
Aspire Systems — Chennai, India

- Led **full-stack development** using **Java/Spring** backend; architected **data pipelines** processing **500K+ events/day** with **Kafka** streaming and **PostgreSQL**, achieving **99.5% availability** for enterprise clients.
- Implemented **data governance layer** with **entitlement management**, **access controls**, and **audit logging**; built **RESTful APIs** for **dataset registration** and **metadata discovery** across distributed systems.
- Built **observability infrastructure** with **Prometheus/Grafana** for pipeline monitoring; implemented **distributed tracing** reducing **MTTR by 60%** through automated alerting and debugging interfaces.

Associate Software Engineer — Jun 2019 – Sep 2020
IMPETUS Technologies — Noida, India

- Developed **Python** backend services for **data ingestion pipelines**; built **RESTful APIs** handling **100K+ daily requests** with **Redis caching** and **data validation** for enterprise data platforms.
- Implemented **ETL workflows** using **Python** and **SQL**; created internal tools for **data quality monitoring** and **schema validation** supporting data engineering team productivity.

PROJECTS

- **AI Data Pipeline Platform** — Built **end-to-end data lifecycle system** for AI training data using **Python/FastAPI**. Implemented **automated ingestion, validation, PII detection**, and **lineage tracking**. Integrated **LLMs (Llama 3.3)** for **synthetic data generation** and **RAG pipelines** with vector stores. Tech: **Python, Kafka, PostgreSQL, ChromaDB**.
- **Multi-Agent Orchestration System** — Architected **21-node AI workflow** using **LangGraph** for automated research pipelines. Built **tool connectors** for 5+ external APIs with **conditional routing**, **state management**, and **observability monitoring**. Tech: **Python, LangGraph, FastAPI**.
- **Real-Time Telemetry Analytics** — Designed **time-series data pipeline** using **InfluxDB** and **Kafka Streams** processing **50M+ events/day**. Implemented **anomaly detection** for asset health with **P99 latency ¡100ms**. Tech: **Java, Kafka, InfluxDB, Grafana**.

TECHNICAL SKILLS

- **Languages:** Python, Java, TypeScript/JavaScript, SQL, C#, Go (familiar)
- **Data Infrastructure:** Apache Kafka, Spark, InfluxDB (time-series), PostgreSQL, Redis, MongoDB, Vector Stores
- **Cloud & Distributed Systems:** AWS (EKS, Lambda, S3), Azure (familiar), Kubernetes, Docker, Microservices
- **Data Lifecycle:** ETL Pipelines, Data Ingestion/Validation, PII Handling, Data Lineage, Governance & Compliance
- **AI/ML:** LLMs (OpenAI, Anthropic), RAG Pipelines, Training Data Preparation, Embedding Models, MLflow

- **Practices:** Event-Driven Architecture, gRPC/REST APIs, CI/CD, Observability (Prometheus/Grafana), Agile/Scrum

EDUCATION

Florida State University
M.S. in Computer Science   Tallahassee, FL

- **Relevant Coursework:** Distributed Systems, Advanced Database Systems, Machine Learning, Data Communications

SRM Institute of Science & Technology
B.Tech. in Computer Science   Chennai, India

PUBLICATIONS & CERTIFICATIONS

- Review Classification & False Feedback Detection (IJAST) | AWS Solutions Architect (Udemy, 2024)