Pranaav Srinivasan

Github | LinkedIn | Portfolio | pranaavsri7@gmail.com | +91 7358628186

Summary

Data Engineering–oriented AI undergraduate with hands-on experience building scalable batch and real-time data pipelines. Strong background in Python, SQL, Spark, Kafka, and AWS for ETL, data modeling, and analytics-ready systems. Passionate about designing reliable data platforms that support machine learning, accessibility-driven products, and large-scale decision systems. , click here.

Education

2023 – 2027 B.Tech in Artificial Intelligence and Data Science, Amrita Vishwa Vidyapeetham
2021 – 2023 Class 12 (CBSE), Devi Academy Higher Secondary School

Technical Skills

Programming Languages	Python, Java, C, C++, Scala, Rust, SQL, HTML
Data Engineering	Apache Spark, Kafka, Hadoop, ETL Pipelines, Batch & Streaming Processing, Data Modeling, Apache Airflow, Snowflake ,SQS , Kinesis
ML / AI Frameworks	PyTorch, TensorFlow, HuggingFace, Scikit-learn, Pandas, NumPy, MLflow,LangChain,RAG, LangGraph,Pydantic,Ray,SpringBoot, PyArrow
Databases	PostgreSQL, MongoDB, Redis, Snowflake, Apache Hive, ChromaDB
Cloud & DevOps	AWS (S3, EC2, IAM), Docker, Jenkins, Linux, Terraform, Ansible, Grafana, GitHub Actions, GitHub, Prometheus, ELK Stack, MLflow
Tools & Platforms	Git, Jupyter, Power BI, Flutter, FastAPI, REST API, Dask, Nginx

Work Experience

Swarm Lens, USA Remote — AI Engineer November 2025 – March 2026

Designed and built scalable big-data pipelines for high-volume ingestion, transformation, and validation to support enterprise risk mitigation and data migration.Ensured data quality, fault tolerance, and SLA compliance across batch and real-time workloads. Contributed to production deployments, CI/CD automation, and operational monitoring, applying MLOps and AIOps practices to improve pipeline reliability and support downstream analytics on cloud-based data platforms.

Projects

- **Real-Time FinTech Analytics & Risk Monitoring Platform**
  Built end-to-end streaming pipelines using Apache Kafka and PySpark Structured Streaming with validation, enrichment, and deduplication at scale. Engineered ELT transformation workflows in Snowflake with raw and curated data layers in AWS S3 following medallion architecture. Orchestrated hybrid batch and streaming workflows via Airflow with DAG-level monitoring, alerting, and retry logic ensuring fault tolerance, SLA adherence, and data integrity.

- **Multilingual RAG Chatbot with Data Ingestion & Moderation Pipelines**
  Designed scalable batch and real-time pipelines for multilingual document ingestion, preprocessing, and moderation as the data backbone for a RAG system. Implemented an embedding generation pipeline integrated with hate speech detection as a data quality filtering stage, ensuring only validated data reached the vector store. Optimized retrieval-path data flows for low-latency inference and deployed infrastructure with cloud-ready configurations on AWS with end-to-end observability via CloudWatch.