SANJAI TRICHINOPOLY SHANMUGAM

Phone: +1 6025655912 | tssanjai98@gmail.com | https://linkedin.com/in/tssanjai

EDUCATION

ARIZONA STATE UNIVERSITY (Tempe AZ) **Aug 2023 - May 2025**
Masters in Computer Science **GPA: 3.75/4.00**

DAYANANDA SAGAR COLLEGE OF ENGINEERING (Bangalore, INDIA) **Aug 2016 - May 2020**
Bachelor of Engineering Electronics and Instrumentation **GPA: 8.8/10.0**

WORK EXPERIENCE

FLOW Austin, TX
Full Stack Engineer Intern **Aug 2024 – Dec 2024**

- Built a scalable SaaS dashboard using **React** and **Tailwind CSS**, delivering 10+ pages (user, billing, subscription) aligned with UX guidelines. Collaborated with UI/UX, QA, and data teams in **Agile sprints** via **JIRA** to ensure timely and integrated feature delivery.
- Developed secure backend services with **Django**, **PostgreSQL**, and **Stripe**, including **JWT authentication**, full subscription workflows, and reusable **REST APIs** for product management. Integrated APIs with frontend for seamless **CRUD operations**.

ACCENTURE Bangalore, INDIA
Data Engineer **Nov 2020 – Jul 2023**

- Designed and maintained scalable ELT pipelines using **Airflow**, **AWS Glue**, and **Python**, automating ingestion of real-time cloud logs into **Splunk**, resulting in 30% reduction in manual intervention during log monitoring.
- Optimized performance of complex analytical queries across **PostgreSQL** and **MySQL**, leading to 25% faster report generation and improved database efficiency for data analytics teams using advanced **SQL** techniques.
- Developed dashboards and pipeline monitors using **Grafana** integrated with **Dockerized** Airflow logs; enabled alerting on SLA breaches and reduced ETL debugging time by 30% across business-critical batch workflows.
- Built fault-tolerant ingestion templates in **Apache Airflow** for structured, semi-structured, and **ClickHouse** targets; implemented backfill, failure alerts, and dependency tracking to support CI-integrated deployments.
- Engineered ingestion pipelines in **Azure Data Factory** to load datasets into **Databricks** for advanced analytics; reduced data lag by 35% through delta-based processing and modular ingestion logic using **PySpark**.
- Implemented secure API data ingestion to consolidate external reporting sources into **Azure SQL DB**, streamlining third-party data syncs and enabling a centralized data view for executive dashboards.
- Designed and deployed robust ETL frameworks on **Databricks** with **PySpark**, processing 2TB+ monthly traffic data for commercial strategy teams to enhance customer segmentation and lifecycle modeling decisions.
- Built schema-validated ELT models in **dbt** and scheduled them with **Airflow**, supporting curated datasets used in Azure-hosted dashboards; implemented freshness tests and versioned artifacts for production MBR consumption.
- Packaged ETL workloads in **Docker** and deployed to **Azure Container Instances**, integrating with **Airflow** for orchestration and GitHub Actions for CI/CD; enabled repeatable, version-controlled deployment pipelines.
- Collaborated with platform engineers to develop scalable workflows using **AWS Step Functions** and **Lambda**, enabling on-demand provisioning of analytics sandboxes and speeding up test deployments by 50%.

ERNET Bangalore, INDIA
Data Engineer Intern **June 2019 – Aug 2019**

- Built a simulation framework for **LoRaWAN** sensor devices via **The Things Network**, generating synthetic network telemetry used to test routing algorithms and packet loss rates across variable payload sizes and gateway distances.
- Developed preprocessing scripts using **Python** and stored decoded payloads in **PostgreSQL**, enabling downstream analytics and improving retrieval times by indexing message metadata for time-series visualizations in dashboards.

PROJECTS

Real-Time IoT Data Processing (**Python** | **FastAPI** | **Kafka** | **Spark** | **Cassandra** | **React**) **Oct 2024**

- Designed and implemented an IoT data pipeline using **Kafka** for real-time stream processing, generating **alerts** based on sensor thresholds, and stored metrics with alert statistics in **Cassandra** for **analytics**.
- Used **Spark** and **Pandas** for data processing and built RESTful APIs with **FastAPI** to expose metrics and alerts, while creating dynamic visualizations on a **React** front-end using **MUI** and **Mermaid** charts.

Smart Grid Failure Forecasting (Databricks | Spark | MLflow) **March 2024**

- Developed a machine learning pipeline using **Spark** in **Databricks** to predict smart meter failures based on signal patterns and anomaly detection, achieving 82% F1 score across 5M records with class imbalance managed via SMOTE.
- Tracked experiment metadata and parameters using **MLflow**, optimizing hyperparameters for XGBoost models, and logged model artifacts and metrics for reproducible deployment within the cloud platform.

Ad Spend Attribution Engine (Airflow | BigQuery | Looker) **Feb 2023**

- Built ETL workflows in **Airflow** to extract multi-channel marketing data into **BigQuery**, integrating campaign attribution logic to model ROAS and generate cohort-based insights for marketing leaders and sales stakeholders.
- Created LookML models in **Looker** to support interactive dashboards and recurring MBRs, reducing manual campaign analysis time by 70% and helping the growth team optimize spend allocation across product verticals.

TECHNICAL SKILLS

Languages & Scripting: Python, SQL, Bash
Data & Cloud Platforms: AWS, Azure, PostgreSQL
ETL & Pipelines: Airflow, dbt, Glue
Backend & Infra Tools: Docker, Lambda, Step Functions
Data Modeling & Visualization: Databricks, D3.js, Looker
Soft Skills: Collaboration, Problem Solving, Documentation