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**

- Developed scalable ETL pipelines using **Python**, **Airflow**, and **AWS Glue** to ingest AWS cloud logs into **Splunk**, enabling real-time monitoring, proactive alerting, and reducing system downtime by 27% across 5+ production environments.
- Automated ingestion of AWS metrics and logs using **Airflow** and custom **Python** scripts, reducing manual workload by 40% and enhancing operational visibility into log anomalies through integration with **ServiceNow** ticketing.
- Optimized 80+ complex queries across **PostgreSQL** and **MySQL** by analyzing execution plans and indexing strategies, improving query performance by 55% and ensuring SLA compliance for log analytics dashboards.
- Built ELT pipelines with **BigQuery** and **GCP Composer**, integrating healthcare partner APIs and ensuring HIPAA-compliant ingestion of 1M+ daily records while maintaining 99.9% pipeline reliability and alerting via **DataDog**.
- Designed modular orchestration DAGs using **Apache Airflow** to support ingestion and transformation logic across **GCP** and **Azure**, improving task granularity, monitoring, and recovery from failed stages by 62%.
- Engineered cost-effective pipelines using **GCP Dataflow** and **BigQuery**, reducing cloud spend by 18% while maintaining high availability for near real-time analytics in critical business dashboards.
- Developed CI/CD-integrated pipeline deployment framework using **Terraform**, **GitHub Actions**, and **GCP Cloud Build**, ensuring version-controlled production releases and reducing deployment errors by 30%.
- Leveraged **Pub/Sub** and **Cloud Functions** to enable event-driven transformations on cloud log events, eliminating batch delays and supporting business analytics with sub-minute latency in reporting workflows.
- Created over 50+ **dbt** models for standardized data transformations on **BigQuery**, incorporating test cases, documentation, and lineage tracking using dbt Cloud for audit readiness and regulatory traceability.
- Monitored cloud pipelines using **ServiceNow** integration and custom dashboards, creating automated ticket generation and resolution workflows to reduce incident resolution time by 25% and boost team SLAs.

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

- Simulated **LoRaWAN** sensor transmission using The Things Network, captured packets via MQTT, and ingested payloads into **PostgreSQL** using **Python**, improving prototype IoT monitoring framework for academic networks.
- Parsed sensor payloads using **Python** and structured data into **PostgreSQL** with time-indexed tables; wrote custom **SQL** queries for latency and signal dropout analysis, enabling performance benchmarking of academic network prototypes.

PROJECTS

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

- Built a real-time IoT pipeline using **Kafka**, **Spark**, and **Cassandra**, enabling low-latency stream processing of sensor events and storing anomaly alerts for visualization and downstream analytics.
- Exposed metrics and alert data through **FastAPI** and developed a responsive **React** frontend with MUI, allowing users to explore time-series insights via interactive visualizations and system heatmaps.

Smart Energy Load Forecasting (**Azure Data Factory** | **Databricks** | **Power BI**) — **March 2024**

- Engineered ETL pipelines using **Azure Data Factory** and transformed regional energy load data using **PySpark** in **Azure Databricks**, supporting time-series modeling and improving forecasting RMSE by 17% on test regions.
- Developed Power BI dashboards to visualize demand trends and model outputs, enabling policy teams to simulate tariff scenarios and reducing analysis cycle time from 3 days to under 1 hour using precomputed query layers.

Patient Risk Classification (**BigQuery** | **Python** | **Airflow**) — **Feb 2023**

- Designed a scoring model to classify chronic illness risk based on longitudinal patient records, preprocessing 5M+ records using **Python** and orchestrating transformation logic in **Airflow** on **GCP Composer**.
- Integrated outputs into a **BigQuery** warehouse and enabled downstream analytics for health ops teams, improving patient targeting accuracy by 26% and enhancing compliance with internal audit checkpoints.

TECHNICAL SKILLS

Languages & Scripting: Python, SQL, Shell
Data & Cloud Platforms: AWS, GCP, BigQuery, Snowflake
ETL & Pipelines: Airflow, Glue, Dataflow, dbt
Backend & Infra Tools: Terraform, GitHub Actions, Docker
Data Modeling & Visualization: dbt, Power BI, D3.js
Soft Skills: Communication, Collaboration, Root-Cause Analysis