

# SAI CHARAN

Data Engineer

✉ saipcharan2023@gmail.com | [in linkedin.com/in/psai-3a7156385](https://www.linkedin.com/in/psai-3a7156385) | [github.com/charan3129](https://github.com/charan3129)

☎ +1 (774) 503-5595 | 📍 New Bedford, MA

## PROFESSIONAL SUMMARY

Results-driven Data Engineer with 3+ years of experience designing scalable ETL pipelines and cloud-based data platforms using **Python, SQL, PySpark, Azure, Snowflake, and AWS**. Proven ability to architect enterprise data lakes, optimize real-time and batch workflows with **Kafka** and **Airflow**, and enforce data quality standards across hybrid environments. Detail-oriented problem solver with strong leadership, cross-functional collaboration, and stakeholder communication skills, committed to delivering automated, high-performance data solutions that drive measurable business outcomes.

## TECHNICAL SKILLS

- **Cloud & Warehousing:** Azure (ADF, Databricks, Synapse, ADLS Gen2), Snowflake, AWS Redshift, AWS S3, Google BigQuery.
- **Big Data & Streaming:** Spark (PySpark), Kafka, Event Hubs, Hive, HDFS, MapReduce.
- **Languages & Tools:** Python, SQL, Spark SQL, Bash, Airflow, Docker, Azure DevOps, CI/CD Pipelines, Git, Power BI, Tableau, Jira, dbt, Terraform, Jenkins, Jupyter, Looker.
- **Data Modeling:** ETL/ELT Development, Dimensional Modeling (Star/Snowflake Schema), SCD Type I/II, Lakehouse Architecture, Partitioning Strategies, Query Optimization, Parquet, ORC.
- **Core Competencies:** Data Warehousing, Data Pipeline Development, Data Integration, Data Governance, Agile/Scrum, Problem-Solving, Stakeholder Communication, Time Management, Adaptability.

## PROFESSIONAL EXPERIENCE

**DATA ENGINEER** — Walmart, USA

*Jul 2025 – Present*

- Architected a cloud-based data lake on ADLS Gen2, integrating 10+ disparate retail systems and centralizing reporting for 50+ stakeholders.
- Engineered 20+ automated ETL pipelines using Azure Data Factory, eliminating 15 hours/week of manual data reconciliation.
- Led the migration of legacy frameworks to event-driven pipelines using Kafka and Event Hubs, increasing throughput by 40% for high-volume transactional data.
- Optimized PySpark job performance through partitioning and join tuning techniques, accelerating execution efficiency by 35%.
- Deployed forecasting and anomaly detection models (Prophet, Isolation Forest) into production workflows, reducing inventory discrepancies by 25%.
- Mentored 3 junior engineers on Airflow DAG design and Spark best practices, improving team delivery velocity by 20%.

**DATA ENGINEER (INTERNSHIP)** — Walmart, USA

*May 2024 – Dec 2024*

- Spearheaded the migration of structured retail datasets from on-premises to Snowflake, leveraging Snowpipe and micro-partitioning to reduce query latency by 30%.
- Built end-to-end ETL workflows using PySpark and Azure Data Factory, processing 2TB+ daily batch loads with 99.5% uptime.
- Established automated data validation checks and unit testing using Python, resolving data discrepancies across 3+ downstream systems.
- Coordinated with senior engineers to diagnose pipeline failures, reducing mean time to resolution by 40% and improving workflow reliability.

**DATA ENGINEER** — Maxgen Technologies, India

*Aug 2021 – Jul 2023*

- Modernized legacy ETL pipelines from on-prem SQL Server and Teradata to cloud-native Spark workflows, boosting daily job reliability by 30%.
- Processed large-scale data using HDFS and MapReduce, ingesting and transforming 5TB of daily transactional records with zero data loss.
- Designed a data quality validation framework using Python and Hive, achieving 99% data integrity across 5M+ records.
- Implemented Star and Snowflake schemas with SCD to standardize reporting layers, enabling 20% faster downstream analytics delivery.
- Reduced storage costs by 20% by adopting Parquet and ORC formats with Snappy/Gzip compression strategies.

## PROJECTS

**ENTERPRISE SALES PERFORMANCE INTELLIGENCE ENGINE** Azure Synapse · Power BI · SQL

- Developed a real-time analytics engine unifying CRM, sales, and financial data through Azure Synapse and Power BI.
- Generated KPI dashboards with predictive scoring models that improved quota attainment by 19%.

**HEALTHCARE ANALYTICS & DRUG SAFETY DASHBOARD** Azure ADLS · ADF · Snowflake · dbt · Airflow · Power BI

- Constructed an end-to-end pipeline integrating CMS Medicare and FDA API data into ADLS with Snowflake and dbt transformations.
- Automated drug safety KPI reporting and operational alerts, cutting manual effort by 80% and improving clinical visibility by 35%.

## EDUCATION

**UNIVERSITY OF MASSACHUSETTS DARTMOUTH**

*May 2025*

M.S., Data Science

Dartmouth, MA

## CERTIFICATIONS

Learning Data Analytics: 1 Foundations — LinkedIn Learning (Jul 2025) | Learning Data Analytics Part 2 — LinkedIn Learning (Aug 2025) | Data Science Methodologies — IBM (Jul 2022)