

SAIYASHWANTH VOTTIKOTI

Data Engineer | AWS, Spark, ETL/ELT, Snowflake

+91 77299 33973 | yashwanthv9339@gmail.com | Hyderabad, India | LinkedIn | Portfolio

SUMMARY

Data Engineer with 3+ years of experience building scalable ETL/ELT pipelines and data processing solutions using Python, SQL, PySpark, Apache Spark, AWS, and Apache Airflow. Skilled in batch and real-time data processing, data lakes, data warehousing, and data validation. Developed and optimized pipelines processing 100M+ daily events and 2TB+ of data, supporting analytics, reporting, risk, compliance, billing, and Customer 360 use cases.

WORK EXPERIENCE

Data Engineer (Contract)

Wells Fargo | Hybrid – San Francisco, CA

Feb 2025 – Present

- Built batch ETL pipelines integrating loans, deposits & card data (10M+ records/day) into a unified AWS lake, cutting reconciliation effort by ~30%.
- Designed Customer 360 datasets covering 5M+ customer profiles, powering credit risk, exposure & compliance reporting across 3 business units.
- Implemented 15+ automated reconciliation/validation checks, boosting audit accuracy to 99.8% and reducing regulatory reporting errors by ~25%.
- Optimized large-scale Spark joins & aggregations on 2TB+/day datasets, cutting job runtime 30-40% through tuning and partitioning.
- Automated 20+ data workflows via Apache Airflow, ensuring 100% on-time regulatory delivery and saving ~15 hrs/week in manual scheduling.
- Delivered 10+ analytics-ready datasets (\$B+ exposure) with risk/fraud/compliance teams using Snowflake & SQL AST-based lineage parsers.

Technologies: AWS S3, AWS Glue, Apache Spark, Apache Airflow, SQL, ETL/ELT, Snowflake, SQL Abstract Syntax Tree

ETL Developer (Full-Time)

Reliance Jio Infocomm Ltd | Hybrid – Hyderabad, India

Dec 2021 – Jun 2023

- Built Spark batch ETL jobs processing 50M+ CDR, SMS & internet usage records daily, feeding billing systems for 10M+ subscribers.
- Designed daily aggregation pipelines computing usage metrics for 10M+ subscribers, cutting aggregation runtime by ~25% via partition tuning.
- Applied watermarking & checkpointing to handle late-arriving, out-of-order events across 5+ sources, reducing data loss/duplication by ~40%.
- Optimized Spark transformations & joins on 1TB+ daily telecom datasets, achieving 99%+ SLA compliance for downstream billing deadlines.
- Managed 15+ Apache Airflow workflows automating & monitoring daily batch jobs, cutting manual intervention by ~20 hours/month.
- Monitored production pipelines handling peak loads of 100M+ daily events, resolving failures within SLA & maintaining 99.9% uptime.

Technologies: Apache Spark, Apache Airflow, SQL, PySpark, ETL/ELT Data Pipelines

PROJECTS

End-to-End Batch ETL Pipeline on AWS with Airflow

- Built a batch ETL pipeline extracting 1M+ records/day from a public API into an AWS S3 data lake using PySpark on EMR.
- Modeled a Star-Schema warehouse in Redshift, orchestrated with Airflow (99%+ DAG success), and visualized in Tableau.

Technologies: Apache Spark, Apache EMR, Apache Airflow, AWS S3, AWS Redshift

Real-Time Streaming Pipeline — Kafka & Spark Streaming

- Built a real-time Kafka pipeline streaming 100K+ clickstream events/hour into Spark Structured Streaming with sub-second latency.
- Applied watermarking and 5-min windowed aggregations, persisting results to S3 and Redshift for near-real-time analytics.

Technologies: Apache Kafka, Spark Structured Streaming, AWS S3, AWS Redshift, Real-Time Data Processing

SKILLS

Programming Languages: SQL, Python

Big Data & Processing: Apache Spark, PySpark, Spark Structured Streaming

Cloud & Data Platforms: AWS S3, AWS Glue, AWS EMR, AWS Redshift, Azure Data Factory, Azure Databricks, Azure Data Lake Storage, Google Cloud Storage, BigQuery, Dataproc, Snowflake

Data Engineering: ETL/ELT Pipelines, Batch Processing, Real-Time Data Processing, Data Warehousing, Star Schema, Data Validation, DBT

Orchestration & Streaming: Apache Airflow, Apache Kafka, Kubernetes

Visualization: Tableau, Power BI

EDUCATION

Master of Science in Management, Data Analytics

Aug 2023 – Apr 2025

Indiana Wesleyan University, Marion, Indiana

Bachelor of Technology in Mechanical Engineering

Aug 2017 – Jul 2021

Sri Indu College of Engineering & Technology, Hyderabad

CERTIFICATIONS

Academy Accreditation – Databricks Fundamentals, Databricks

Digital Skills – Artificial Intelligence, FutureLearn