DHRUVRAJ SINGH RATHORE

Austin, Texas | (737)-206-1179 | dhruvrajrathore2011@gmail.com | Portfolio | LinkedIn | GitHub

Data Engineer with 3 years' experience skilled in Python, PySpark, and AWS, with expertise in building end-to-end ETL pipelines, Airflow orchestration, and data modeling that improve reliability, speed, and scalability across enterprise systems.

EDUCATION

Texas A&M University　August 2024 – December 2025
Master of Science in Data Science, College of Engineering, GPA: 4.0　*College Station, TX, United States*

SRM Institute of Science and Technology　July 2018 – May 2022
Bachelor of Technology in Computer Science, GPA: 3.8　*Chennai, Tamil Nadu, India*

WORK EXPERIENCE

AI Engineer Intern　June 2025 – Present
TechSur Solutions　*Reston, Virginia, United States*

- Developed an **intelligent automation pipeline** that generates and validates application code from structured requirements, **reducing manual development effort by 70%** and improving delivery consistency.
- Implemented **automated error detection and correction workflows** to identify build failures, fix code issues, and re-verify deployments, **increasing first-attempt build success from 30% to 85%.**
- Integrated **CI/CD pipelines** with APIs and internal tools to create an end-to-end workflow from feature definition to deployment, **improving release speed by 40%** and reducing manual handoffs.

Data Engineer　December 2022 – June 2024
Draup Business Solutions　*Bangalore, Karnataka, India*

- Designed **scalable PySpark ETL pipelines on AWS EMR** to process 50 TB+ workforce data, **improving runtime by 30% and ensuring 99% data accuracy** through optimized partitioning and strict schema validation.
- Automated **daily validation for 200M+ records using Airflow DAGs**, powering a live data-quality dashboard that cut QA escalations by 50% and strengthened data trust across the clients.
- Modeled **Redshift OLAP schemas** for 20M+ job and location records, optimizing joins and aggregations to **deliver 40% faster analytical queries for business stakeholders.**
- Implemented a **serverless data pipeline using AWS Lambda + DynamoDB** for on-demand S3 queries, reducing ad-hoc client **data delivery time by 30%.**

Associate Data Analyst　August 2021 – November 2022
HighRadius Corporation　*Hyderabad, Telangana, India*

- **Optimized SQL and Snowflake workflows** for AR data processing using indexing and window functions, cutting runtime by 40% and improving reporting speed.
- Developed a Python-based keyword matching algorithm to automate claim-to-deduction mapping, **increasing recovery efficiency by 60%.**
- **Collaborated with consultants and product owners** to restructure AR data models in Snowflake, ensuring clean, analysis-ready data that **reduced ad-hoc query time for finance teams by 30%.**
- Designed **AR dashboards in Power BI and Tableau** that reduced collectors' workload by 60% and **improved cash recovery by 20%**, adopted by 20+ client teams.

PROJECTS

Order Analytics Pipeline | *GitHub* | *Python, DBT, SnowFlake, Airflow, Astro CLI*　Sept 2025 – Oct 2025

- Built an end-to-end pipeline using **dbt, Snowflake, and Airflow** to transform raw transactional data into analytics-ready tables with daily automated scheduling
- Implemented **automated data validation tests** to ensure accuracy and consistency across staging and reporting layers.
- Containerized the workflow with **Docker and version-controlled it via Git** for reproducible deployments and seamless automation.

SKILLS

Programming: Python, SQL, T-SQL, Pandas, Shell Script, Bash
Cloud Infrastructures/Services: AWS EC2, S3, Lambda, Glue, Redshift, Data Lakes, Kafka, Azure Databricks
Big Data & Databases: Relational Databases (MySQL, PostgreSQL), NoSQL, Redis, MongoDB, Schema Design (Logical/Physical), Data Modeling (Snowflake/Star Schemas), Data Mining, Spark, Hadoop, HDFS, DBT
Tools & Visualizations: Version Control(Git, GitHub), CI/CD, JIRA, Apache Airflow, Docker, AWS Quicksight