DHRUVRAJ SINGH RATHORE

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

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*

SKILLS

Programming: Python, SQL, T-SQL, PySpark, Shell Script, Bash
Cloud Infrastructure & Services: AWS (EC2, S3, Lambda, Glue, Redshift, EMR)
Big Data & Databases: Spark, Hadoop, HDFS, Snowflake, Relational (MySQL, PostgreSQL), NoSQL (MongoDB, Redis), Schema Design (Star/Snowflake), dbt
Tools & Visualization: Apache Airflow, Docker, Git/GitHub, CI/CD, JIRA, AWS Quicksight, Tableau

WORK EXPERIENCE

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

- Built **scalable ETL pipelines using Spark on AWS** EMR/Glue to process 50+ TB of structured and semi-structured S3 data, **reducing processing time by 30%** through optimized partitioning and resource utilization.
- **Streamlined daily validation** for 200M+ records using **Airflow DAGs**, enabling a live data-quality dashboard that **dropped QA escalations by 50%** and strengthened client confidence in data accuracy.
- **Led the redesign of legacy Snowflake and Redshift data models**, improving schema design and distribution strategies, reducing query runtime by 40%.
- **Collaborated with product and analytics teams** to implement a serverless data pipeline using AWS Lambda, Athena, and DynamoDB for on-demand S3 queries, **accelerating ad-hoc client data delivery time by 50%.**

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

- Engineered **PL/SQL and Snowflake-based data workflows for financial analytics**, improving report generation speed by 40% via query tuning and indexing.
- Developed a Python-based keyword matching algorithm to automate claim-to-deduction mapping, **increasing recovery efficiency by 60%.**
- **Collaborated with product owners** to optimize Accounts Receivable (AR) schemas and data structures in **Snowflake**, ensuring clean, analysis-ready data that **reduced ad-hoc query time for finance teams by 30%.**
- **Designed AR analytical dashboards** in Power BI and Tableau to automate reporting on AR aging, disputed invoices, and collections KPIs, **reducing manual workload by 70%.**

PROJECTS

Order Analytics Pipeline | **GitHub Link** | *Python, dbt, Snowflake, Airflow, Astro CLI*   Sept 2025 – Oct 2025
- Created an end-to-end analytics pipeline using dbt, Snowflake, and Airflow to transform raw transactional data into analytics-ready tables with automated daily scheduling and validation.
- Containerized with Docker and version-controlled via Git, enabling reproducible and seamless CI/CD integration.

TravelGenie | **GitHub Link** | *Python, REST APIs, BERT, RAG*   March 2025 – May 2025
- Developed an AI-powered travel itinerary planner that fetches real-time flight, hotel, and attraction data using parallel REST API pipelines, cutting research time from hours to minutes.
- Designed ETL and ranking workflows for data cleaning, normalization, and personalized recommendations, resulting in faster, more reliable, and better personalized budget-friendly itinerary generation.

ACHIEVEMENTS

- **Proposed and led development of a Command Center dashboard** to automate client data deliveries, cutting delivery time from an average of 5 hrs to under 1 hr, **helping the team secure 4 new client deals worth $10M** in a single quarter.
- **Delivered internal AR dashboards** for 22 client collector teams, **cutting workload by 60% ($400K/yr via 5 FTEs saved)** and adding $1.8M in recovered cash.