

Atif Imam

Lead / Senior Data Engineer · 8+ Years

Phone: +91 79033 04505 · mohd.atifimam@hotmail.com · linkedin.com/in/atifimam · github.com/imamAtif

Professional Summary

Senior Data Engineer with **8+ years** of experience architecting and **leading** large-scale data platforms across healthcare, hospitality, and e-commerce. Proven track record of **owning end-to-end platform decisions**, from infrastructure design and governance to engineering standards and team enablement, while delivering measurable outcomes: **30% Snowflake cost reduction**, **10x pipeline performance improvement**, and **35% GCP throughput gains**. Experienced in mentoring engineers, driving cross-functional alignment, and translating complex business requirements into scalable, governed architectures. Deep expertise in **Airflow**, **dbt**, **Snowflake**, **Spark**, and **AWS**.

Key Achievements

- **30% Snowflake cost reduction:** warehouse sizing optimisation, query profiling, and workload management across a healthcare data platform (Diaceutics)
- **10x ingestion performance improvement:** redesigned partition strategies and Spark configurations for large-scale Parquet pipelines (Atria)
- **35% GCP pipeline acceleration:** end-to-end architectural refactoring of high-volume customer data workflows (Wayfair)
- **Zero unplanned downtime** on 50+ business-critical pipelines via fault-tolerant Airflow DAG design with automated alerting and retry logic
- **3× performance awards** across career: Employee of the Quarter (2023), Bravo Award (2020), Hats-Off Award (2019)

Technical Skills

Languages & Scripting	Python, SQL, PySpark, Bash/Shell Scripting
Data Warehousing	Snowflake, BigQuery, Redshift
Orchestration & Transform	Airflow, dbt, Dagster, Meltano
Compute Engines	Apache Spark
Cloud & DevOps	AWS (EKS, S3, CloudFormation, Secrets Manager, Batch, Lambda), Docker, Kubernetes, GitLab CI/CD
Leadership	Technical mentorship, architecture ownership, engineering standards, cross-functional stakeholder management

Professional Experience

Sr. Data Engineer (Tech Lead) DataBerry Tech Pvt Ltd	Dec 2024 – Present
Client: Diaceutics (Healthcare Analytics)	Remote
• Acted as platform tech lead: owned architecture decisions for end-to-end ELT workflows using Airflow, dbt, Snowflake, and AWS; drove adoption of modular data modelling practices across the engineering team. • Reduced monthly Snowflake compute costs by 30% through warehouse sizing optimisation, query profiling, and workload management, while simultaneously cutting dashboard query latency. • Designed and enforced a data governance framework covering RBAC policies, tag-based and dynamic masking, and PII protection, ensuring regulatory compliance for sensitive healthcare data. • Built a zero-defect data quality framework incorporating schema enforcement and reconciliation logic, setting the quality benchmark for all downstream business deliverables.	
Senior Data Engineer Forary Soft	Dec 2023 – Dec 2024
Client: Wayfair (E-Commerce)	Remote
• Led architectural refactoring of GCP data pipelines, reducing processing time by 35% and significantly improving system reliability for high-volume customer data workflows.	

- **Owned design and delivery** of analytical models measuring email/SMS campaign effectiveness, enabling the marketing team to reduce customer fatigue and optimise communication frequency.
- Engineered fault-tolerant Airflow DAGs with automated alerting and retry logic, achieving near-zero unplanned downtime for 20+ business-critical pipelines.

Sr. Associate Lead / Sr. Data Engineer | RedDoorz

Aug 2021 – Aug 2023

Noida

- **Led the data platform team** as the most senior engineer on the squad. Owned architecture decisions, defined engineering standards, and mentored 3 junior engineers on best practices, version control, unit testing, and automated data validation.
- **Deployed and owned production Airflow on AWS EKS** with Secrets Manager and EFS Drivers, creating a secure, scalable orchestration environment serving **50+ DAGs** across the organisation.
- Architected end-to-end ELT/ETL pipelines using Python, and Pyspark to ingest large-scale datasets from **6+ sources** (MySQL, Postgres, Freshworks, AppsFlyer, Zendesk, CleverTap) into S3, Redshift, and Snowflake.
- **Built a competitive intelligence framework** using web scraping (Agoda, tiket.com, Traveloka) to surface real-time market pricing insights for revenue teams, a first-of-its-kind initiative at the company.
- Provisioned cloud infrastructure via AWS CloudFormation and built GitLab CI/CD pipelines, **driving the team's first adoption of infrastructure-as-code practices** and reducing manual deployment errors.
- Awarded **Employee of the Quarter (2023)** for delivering high-impact data infrastructure improvements.

Data Engineer / Associate | Axtia India Pvt Ltd

Feb 2020 – Aug 2021

Noida

- Achieved a **10x reduction** in Parquet file ingestion runtime by redesigning partition strategies and optimising Spark configurations.
- Core contributor to flagship product **DataMax**. Built complex business logic in PySpark, Redshift, and Matillion that formed the foundation of client-facing analytics.
- **Developed a custom Data Quality Check pipeline in Dagster**, setting a new internal standard for data accuracy across all deliverables.
- Awarded **Bravo Award (2020)** for outstanding contribution to product development.

Programmer Analyst | Cognizant Technology Solutions

Jan 2018 – Jan 2020

Bengaluru

- Automated high-volume manual data tasks using PySpark and Python, significantly boosting team productivity.
- Modernised legacy Hive queries into optimised PySpark scripts, reducing runtime and infrastructure resource consumption.
- Built pricing and sales data models sourced from SAP systems and contributed to a reusable data loading framework.
- Awarded **Hats-Off Award (2019)** for exceptional performance.

Education

B.Tech in Computer Science | NIIT University, Neemrana

2013 – 2017