

Syed Mazahir Mehdi Zaidi

syedmazahirmehdizaidi11@gmail.com — +91-7619930856 — LinkedIn — GitHub

Summary

Data Engineer with hands-on experience in building scalable pipelines, cloud integrations, and BI solutions. Skilled in Snowflake, SQL, AWS, Azure, dbt, and Airflow with strong Python expertise. Proven ability to deliver secure, compliant, and analytics-ready data systems.

Skills

- **Data Engineering:** Snowflake, dbt, Airflow, ETL/ELT, Data Modeling, Apache Spark
- **Cloud Platforms:** AWS (S3, Glue, IAM), Microsoft Azure, GCP
- **Visualization & BI:** Power BI, Tableau (DAX, KPI dashboards)
- **Programming:** Python, SQL, C++, JavaScript

Work Experience

Data Engineer – IBM India, Kochi, Kerala

Feb '25–Present

- Designed Snowflake schemas and ingested large datasets from SQL Server/Oracle into GCP Snowflake.
- Automated ingestion jobs on GCP-Snowflake pipelines ensuring scalable data loads.
- Tokenized PHI/financial data and generated regulatory-compliant extracts.
- Applied transformations in Snowflake to prepare analytics-ready datasets.
- **Key Achievements:** Delivered pipelines processing 10M+ records weekly, improving data availability by 35%.

Data Engineer Apprentice – IBM India, Remote

Feb '24–Jan '25

- Implemented Snowflake, dbt, and MySQL-based data pipelines.
- Built ETL workflows using Python and SQL.
- Assisted senior engineers in deploying cloud data integrations.
- **Achievement:** Deployed prototype ETL pipeline reducing manual reporting by 20%.

Education

B.Tech. Computer Science – Amity University, Lucknow

Aug '20–Jun '24

CGPA: 8.84

Projects

Customer Data Pipeline & Analytics Platform

 Source

- Built end-to-end pipeline for customer, transaction, inventory, feedback & clickstream data.
- Developed AWS Glue ETL to transform S3 raw data into curated Parquet tables.
- Loaded curated data into Snowflake staging, designed fact/dimension tables, and delivered analytics-ready datasets.
- Developed Power BI dashboards (Customer 360, Product KPIs, Engagement, Feedback metrics).
- **Tools & Tech:** AWS (S3, Glue, EventBridge, Lake Formation, Athena, CloudWatch), Snowflake, Power BI, SQL, Python

Medallion Data Lakehouse Pipeline using PySpark and dbt on Databricks

 Source

- Designed a Medallion (Bronze–Silver–Gold) data lakehouse architecture for reliable incremental processing.
- Developed CDC-based transformations with Delta Lake ensuring data accuracy and lineage.
- Automated modeling and SCD tracking via dbt for analytics-ready datasets.
- **Tech:** Databricks, PySpark, Delta Lake, dbt Cloud, Unity Catalog, Streaming

Certifications

- Snowflake SnowPro Core Certification
- Microsoft Certified: Azure Fundamentals (AZ-900)

May '25–May '27

Jul '25