

Naveed Mohiuddin

Data Engineer | AWS & Azure Certified | ✉ naveedmohiuddin0311@gmail.com | 📞 +1 773-394-9716

🌐 [linkedin.com/in/naveed-mohiuddin](https://www.linkedin.com/in/naveed-mohiuddin) | 🐙 github.com/NaveedMohiuddin | 🌐 naveedmohiuddin.vercel.app | 📍 United States

PROFESSIONAL SUMMARY

Experienced Data Engineer building cloud-native data platforms across AWS and Azure. Skilled in PySpark, Glue, Databricks, Airflow, and Data Factory for distributed ETL, dimensional modeling, and analytics workloads. AWS Certified Solutions Architect and Data Engineer focused on pipeline reliability, cost optimization, and production-grade data systems.

EXPERIENCE

Data Engineer

Jul 2025 – Present

Benda Infotech

Remote, US

- Designed serverless ETL pipelines using AWS S3, Lambda, and Glue processing ~1M records daily with 99% reliability
- Developed PySpark transformation jobs in Glue handling schema validation, deduplication, and type casting across 10+ datasets, improving batch runtime through DPU tuning and join optimization
- Designed analytics-ready dimensional datasets and Redshift load workflows supporting 3 downstream reporting use cases and improving data accessibility for business analysis
- Orchestrated workflows using Airflow DAGs with SLA monitoring and task-level retries, supporting 20+ ETL jobs
- Optimized Athena queries via partition pruning and Parquet conversion, lowering execution costs by 30–40%
- Implemented CI/CD pipelines with schema-drift alerting and record-count validation, collaborating with QA to ensure data consistency before production releases
- Tech:** AWS S3, Glue, Lambda, Redshift, Athena, EventBridge, PySpark, Airflow, SQL, Python, Parquet, SNS

Software Engineer (Data Engineering)

Feb 2022 – Jul 2023

Applied Information Sciences (Client: GEICO)

Hyderabad, India (Remote US)

- Built Azure Data Factory pipelines ingesting 100s of GBs of policy, claims, and CRM data into ADLS Gen2 using scheduled and incremental loads
- Developed PySpark transformations in Databricks following medallion architecture collaborating with business analysts to define cleansing rules and validation logic
- Worked with Delta Lake to maintain data history, supporting compliance and audit across 4 business domains
- Contributed to Kafka and Databricks Structured Streaming workflows, improving claims data availability from hourly batch refreshes to near real-time delivery for downstream consumers
- Automated CI/CD deployment of ADF pipelines via Azure DevOps with ARM templates, coordinating releases across dev, UAT, and prod environments with the DevOps team
- Delivered analytics-ready datasets for BI, improving reporting turnaround time by up to 50% for business stakeholders
- Tech:** Azure Data Factory, Databricks, PySpark, Delta Lake, ADLS Gen2, Synapse, Kafka, SQL, Power BI, CI/CD

PROJECTS

AWS Lakehouse & Analytics Platform | S3, Glue, Athena, Redshift, Airflow, PySpark, Iceberg - [GitHub](#)

- Built a serverless AWS data lakehouse to ingest and analyze public crime datasets. Used Lambda, Kinesis Data Firehose, S3, AWS Glue, and Apache Iceberg tables for ETL pipelines. Enabled scalable, ad-hoc SQL analytics with Athena and enforced column-level security via Lake Formation.

Big Data Processing with Spark | Apache Spark, HDFS, GCP Dataproc, SQL - [GitHub](#)

- Processed multi-terabyte datasets on Google Cloud Dataproc using Spark and Hadoop clusters. Optimized ETL performance threefold through efficient join strategies, dynamic partition pruning, and scalable distributed processing

EDUCATION

Master of Science in Computer Science

Aug. 2023 – May 2025

Illinois Institute of Technology

Chicago, IL

Bachelor of Engineering in Computer Science

Aug. 2018 – June 2022

Osmania University

Hyderabad, India

TECHNICAL SKILLS

Data Engineering: ETL/ELT, Data Modeling, Star Schema, SCD Type 2, Batch & Streaming Pipelines, Unity Catalog
Cloud Platforms: AWS (S3, Glue, Redshift, Lambda, EC2, EventBridge), Azure (ADF, Databricks, Synapse, ADLS Gen2)
Big Data: Apache Spark, PySpark, Kafka, HDFS, Hadoop, Hive, Delta Lake, Iceberg, Parquet, Medallion Architecture
Orchestration & DevOps: Apache Airflow, CI/CD, Azure DevOps, Git, GitHub, Docker, CloudWatch
Programming: Python, SQL, Java, Linux, Bash
Certifications: AWS Certified Solutions Architect, AWS Certified Data Engineer, Microsoft Azure Fundamentals