Mohammed Kushalgarhwala

+1 346-539-9344 | Mohammedkushalgarh@gmail.com | LinkedIn

Summary

Data Engineer experienced in building cloud-native data pipelines and Lakehouse models using PySpark, SQL, and Delta Lake, enabling reliable analytics and decision-making.

Education

Master of Science in Data Science Jan 2025 – May 2026
Sam Houston State University, *Huntsville, TX*
Bachelor of Technology in Computer Science Aug 2018 – May 2022
Sri Aurobindo Institute of Technology, *Indore, India*

Technical Skills

Programming & Querying: Python, SQL, Pandas, NumPy, PySpark
Data Engineering (Azure): Azure Data Factory, Azure Databricks, Azure Synapse Analytics, Microsoft Fabric
(AWS): Amazon S3, AWS Glue, Amazon Redshift, Amazon Athena, AWS Lambda
Data Storage: Azure Data Lake, Azure Blob Storage, Azure SQL Database, Delta Lake
Analytics & Visualization: Power BI
Security: Azure Key Vault, AWS IAM
DevOps & Monitoring: Azure DevOps, Azure Monitor, Git

Experience

Data Automation Intern Jan 2026 – Present
DM Clinical Research *Houston, TX*

- Automated **Google Workspace** workflows using **Apps Script** and APIs to integrate Gmail, Sheets, Chat, and Calendar
- Centralized scheduling and tracking system for 36–50 studies across 20+ sites, reducing request handling time by 5 minutes
- Automated access request and approval workflows with audit-ready access tracking, reducing IT effort by 70%
- Designed an IT inventory and asset tracking pipeline by assigning unique asset IDs, categorizing equipment, and automating role-based notifications, enabling **near-complete asset accountability**
- Implemented automated data ingestion from Google Forms with validation and eligibility filtering for research studies

Graduate Assistant – Data Engineering Jun 2025 – Dec 2025
Sam Houston State University *Huntsville, TX*

- Built a data-driven eCommerce platform using **Flutter, NestJS** and **MySQL** to capture behavioral and transactional data.
- Implemented a Real-time **clickstream ingestion pipeline** using **AWS Lambda** and **Amazon S3** for customer behavior.
- Enabled serverless analytics with **AWS Glue** and **Amazon Athena** to analyze user engagement, funnels, and interaction.
- Designed and optimized PostgreSQL schemas on **AWS RDS** to support ACID transactions and analytical feature.
- Provisioned secure cloud infrastructure on **EC2 and VPC**, leveraging **S3** for data lakes and **GitHub Actions**

Projects

End-to-End Azure Data Engineering Pipeline Using Medallion Architecture
Azure Data Factory, Azure Databricks, Azure Synapse Analytics, Delta Lake

- Ingested multi-source data from GitHub into **Azure Data Lake** (Bronze layer) using Azure Data Factory**(ADF)**
- Implemented **incremental loading** with watermarking to reduce reprocessing overhead and control compute costs
- Evolving source schemas using ADF dynamic mappings and **Delta Lake** schema evolution to maintain pipeline stability
- Developed scalable **PySpark** transformations in **Azure Databricks** to produce validated (Silver layer) datasets and aggregated, analytics-ready (Gold layer) datasets in **Azure Synapse** Analytics for reporting and KPI analysis

Customer Behavior Analytics Lakehouse (Bronze–Silver–Gold)
Databricks, Apache Spark, Delta Lake, PySpark, SQL

- Built an end-to-end **Databricks** Lakehouse pipeline to analyze customer Behavioral and transactional data.
- Ingested raw **CSV data** with **explicit schemas**, implementing **Delta Lake** Bronze tables with audit metadata
- Developed Silver-layer PySpark transformations for data cleansing, deduplication, schema enforcement, and CDC-style upserts
- Applied data quality validations and a quarantine pattern to isolate invalid records and ensure analytics-ready datasets
- Created Gold customer KPIs (AOV, total revenue, recency, preferred channel) using **Spark SQL** and window functions