

MD Sohrab Ahmed

Senior Data Engineer

+60 1112625189 | sohrabahmed@outlook.com | [linkedin.com/in/sohrabahmed](https://www.linkedin.com/in/sohrabahmed) | Kuala Lumpur, Malaysia

SUMMARY

With 10 years of profound experience as **Senior Data Engineer** for all types of **Data** and Advanced Analytics to drive high priority business initiatives on **Azure, AWS, Cloudera, and Databricks** Platform with passion to identify competitive advantages by transforming raw data into powerful information with cutting-edge technologies.

EXPERIENCE

Senior Data Engineer | SEEK, Kuala Lumpur, Malaysia

Dec 2024 - Present

Skills: Microsoft Azure, Data Factory, Azure Data Lake Storage Gen2, Databricks, Apache Flink, Apache Kafka, MSK, Redshift, Delta Lake, Spark, Python, Scala, Real-time Data Processing, Data Lakehouse Architecture, Mentoring, Cost Optimization

- Spearheading the migration of SEEK's enterprise data warehouse from Amazon Redshift to Databricks Lakehouse to reduce infrastructure costs and enhance analytical performance and scalability.
- Designing and implementing a real-time data ingestion framework using Apache Flink, Kafka, and Databricks to enable faster business insights and operational decision-making.
- Building **multi-cloud compatible** pipelines for batch and streaming use cases, ensuring smooth transition from AWS to Azure while maintaining operational continuity.
- Enabling cross-functional data teams to leverage Databricks effectively by mentoring engineers and developing reusable frameworks, best practices, and onboarding documentation.
- Collaborating with analytics, product, and engineering teams to modernize data architecture and streamline data access across SEEK's data platform.
- Driving initiatives focused on **platform observability, cost optimization, security, and performance tuning** to ensure a reliable and secure data ecosystem.

Senior Data Engineer | FedEx, Hyderabad, India

Aug 2023 – Dec 2024

Skills: Microsoft Azure, Data Factory, Databricks, Synapse Analytics, SQL Databases, SQL, Spark Scala, PySpark, Python, Delta, Delta Lake, DevOps, Git, CI-CD, SAS, Datawarehouse

- Enhancing data ingestion pipeline with metadata driven **ABC (Audit, Balance, Control) framework** resulting enhance data quality, better operation, security, and reliability at lesser cost.
- Leading the re-architecture of the MARS Platform utilizing Azure cloud and Databricks Unity Catalog services which enabled enhance scalability, reliability, and security.
- Migrated SAS Enterprise APAC business reports and objects to Synapse serverless pool for better analytical experience for 500+ customers and cost saving by 30%.
- Implemented **data lineage dependencies** to track the flow of data across systems, ensuring transparency, integrity, and compliance throughout the data lifecycle.
- Established governance policy to standardized and accelerate code testing processes and ensure quality assurance compliance.
- Implementing **SAFe** methodology to synchronize **Agile** development efforts, aligning with business objectives and efficient cross-functional collaboration at scale.

Senior Data Engineer | PPG Industries, Subang Jaya, Malaysia

Nov 2019 – Dec 2022

Skills: Microsoft Azure, Data Factory, Logic App, Key vault, Databricks, Scala, Python, Delta, Delta Lake, DevOps, Snowflake, Power BI, Synapse, AAS, SQL-Server, Talend, ETL, SQL Datawarehouse

- Developed scalable **Delta Lake platform** with ACID, time travel, dynamic schema evolution, audit history features and hard delete, having capability of holding Type 1 & 2 SCDs data for different project requirement.
- Developed **metadata driven framework**, automated ingestion and integration pipeline for multiple structured and semi-structured sources using Azure SQL Databases, ADF, Databricks, Logic App, Analysis Services, ADLSGen2, Synapse Analytics, Key Vault and Snowflake.

- Improved quality and reliability, measured by 84% fewer customer defects by adopting a metadata-driven ABC framework, proactive quality rules, and health check.
- Developed Data Consumption layer with Composite model (Dual mode) using 200GB+ of data in **Snowflake** and frequently accessed data in **Power BI** Premium.
- Implemented logics to **integrate** multiple ERPs invoice data into centralized spend visibility report developed in Spark Scala using Azure resources.
- Eliminated 90% of manual effort and 75% of business impact by introducing **CI/CD process for Databricks** and Data Factory using Azure DevOps pipeline.
- Redesigned AAS models and Power BI Dataset refresh process to allow parallel refresh in secured with time saving and same cost using **Logic App**

Big Data Engineer | Neustar, Hyderabad, India

Dec 2018 – Oct 2019

Skills: Scala, Spark SQL, Hive, Airflow, Shell Script, Hadoop, Cloudera, Git, S3

- Developed process to create **data quality report** on fill rate generation for business analyst to review and approve as per tolerated threshold.
- Build automated data quality framework to detect and fix data inconsistencies, unexpected values using proactive rules setup with Spark Scala and Shell script.
- Developed centralized **data catalog** by applying consistent and business approved naming conventions.
- Designed data availability **health dashboard** to tracks, analyzes, and displays KPI such as refresh rate, fill rate generation, for different business use case.

Data Engineer | Sphota, Bengaluru, India

Apr 2018 – Dec 2018

Skills: Hadoop, Hive, Microsoft Flow, Azkaban, Spark, Teradata, Power BI, Git

- Developed data ingestion framework including job scheduling in Azkaban, monitoring job status in Power-BI and proactive alerting directly to business user on Slack.
- Implemented low latency data query in IBM Compose platform for ad-hoc access of clickstream data.
- Build UDF in Scala for GDPR Compliant in Hive and Teradata systems.

Programmer Analyst | Cognizant, Gurugram, India

Jun 2015 – Mar 2018

Client: **American Express**, Phoenix, Arizona

Skills: Big Data, Hadoop, MapR, Hive, Sqoop, HDFS, MySQL, Unix, Scala, Spark, SQL

- Developed Sqoop job to import data from various relational databases to HDFS and Hive Data Warehouse for Global Risk analysis.
- Collected and analyzed large-scale data sets using Hadoop ecosystem component.
- Automated ETL process using Hive script backed by external ORC tables with static and dynamic partitions.
- Developed custom UDFs and UDAFs for Hive in Scala to implement business logic.

CERTIFICATIONS

- Microsoft: Azure Fundamental, Data Fundamentals, Data Engineer, Developer, Power BI Data Analyst
- Databricks: Azure Databricks Platform Architect, Data Engineer Professional
- Cloudera: CCA Spark and Hadoop Developer
- Trainer: Microsoft Certified Trainer for Azure
- SAFe: SAFe 6 Practitioner

EDUCATION

- **MTech - Computing Systems and Infrastructure**
BITS Pilani, India | Completed: 2023
- **BTech - Computer Science and Engineering**
Uttar Pradesh Technical University, India | Completed: 2014