





# MD Sohrab Ahmed

Senior Data Engineer

+60 - 173267957   
sohrabahmed@outlook.com   
linkedin.com/in/sohrabahmed   
Subang Jaya, Malaysia 

## Summary

Around 7 years of profound experience as **Senior Data Engineer** for all types of **Data** and Advanced Analytics to drive high priority business initiatives on **Azure** Cloud, Cloudera and **Databricks** Platform with passion to identify competitive advantages by transforming raw data into powerful information with cutting-edge technologies.

## Professional Experience

### PPG Industries | Subang Jaya, Malaysia

#### Data Engineer, Senior

Nov 2019 – Present

Skills: Microsoft Azure, Databricks, Data Factory, Key Vault, Scala, HDInsight, DeltaLake, Logic App, Synapse, AAS, DevOps, CLI, SQL-Server, Talend

- 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 ADF, Databricks, Logic App, AAS, ADLSGen2 and Azure Synapse
- Implemented logics to **integrate** multiple ERPs invoice data into centralized spend visibility report developed in Spark Scala using Azure resources
- Developed **CI CD for Databricks** code 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**

### Neustar | Hyderabad, India

#### Big Data Engineer

Dec 2018 – Oct 2019

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

- Developed process to create **data quality report** on fill rate generation for business analyst to review and approve as per tolerated threshold limit
- Build automated data **quality framework** to fix inconsistencies and unexpected values using Spark SQL 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 for different business use case

### Sphota | Bengaluru, India

#### Data Engineer

Apr 2018 – Dec 2018

Client: **Hotels.com**, Texas

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

- Developed an automated data ingestion framework including job scheduling in Azkaban, monitoring in Power-BI and alerting in 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

### Cognizant | Gurugram, India

#### Programmer Analyst

Jun 2015 – Mar 2018

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

Skills: Hadoop, Hive, Sqoop, MySQL, Unix, Scala, Spark, SQL

- Developed Sqoop job to import data from various relational database to HDFS and Hive 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 Certified** – Data Engineer, Developer Associate, Data Analyst, Data Fundamentals, Fundamentals
- **Cloudera** – CCA Spark and Hadoop Developer
- **Trainer** – Microsoft Certified Trainer
- **Databricks** – Managing Delta Lake

## Education

**MTech** – Master of Technology in Computing Systems & Infrastructure from BITS Pilani, India

2022 – 2024

**BTech** – Bachelor of Technology in Computer Science & Engineering from Uttar Pradesh Technical University, India

2010 – 2014