

Muhammad Hassaan Sohail

Chicago, IL | sohailmuhammadhassaan@gmail.com | +1 (773) 457-8468 | heyhassaan.com

SUMMARY

Dynamic Data Engineer with expertise in building scalable data pipelines, real-time streaming systems, and advanced analytics solutions. Proficient in Python, SQL, and cloud platforms (AWS, GCP). Demonstrated success in optimizing data workflows and delivering high-volume, fault-tolerant data engineering systems that enable data-driven decision-making.

EDUCATION

The University of Chicago | Chicago, IL | MS Applied Data Science **Expected December 2025**
Coursework: Data Engineering Platforms, Machine Learning I & II, Advanced AI Systems, Time Series Analysis, Natural Language Processing, Data Visualization, Bayesian Methods

Lahore School of Economics | Lahore, PK | BBA Finance & Marketing (Honors) **May 2022**
Coursework: Mathematics, Statistics, Financial Management, Computing for Business, Risk Management

WORK EXPERIENCE

Findd Inc, Provo, UT

Data Engineering Lead

March 2024 – September 2024

- Designed and implemented scalable data pipelines and real-time streaming solutions using Kafka, boosting retrieval speed by 45%, achieving 99% uptime, and ensuring 20% higher compliance accuracy.
- Automated workflows for 70,000+ employee records with Apache Spark and Airflow, improving HR efficiency by 50% and enabling precise workforce forecasting with 90% accuracy.
- Collaborated with cross-functional teams to deliver predictive analytics models, enhancing decision-making and streamlining data-driven strategies.

Well-Being & Equity (WE) in the World, Boston, MA

Data Scientist, (Part-Time)

February 2022 - September 2024

- Architected and optimized 45+ ETL pipelines, reducing data inconsistencies by 95% and improving reliability by 40%, while implementing secure data governance frameworks for 300,000+ individuals and 200+ partners.
- Applied Spark for batch processing and Flink for real-time analytics, enabling 50+ policy recommendations and significantly enhancing public health outcomes.
- Partnered with federal organizations to deploy innovative data solutions, cutting insight processing time by 30% and supporting data-driven decision-making.

Enrichers Private Limited, Lahore, PK

Data Analyst Intern

June 2019 - August 2019

- Developed ETL pipelines to analyze 10 years of trade data using Python and Tableau, uncovering profit drivers and improving trade decision-making.
- Implemented machine learning models (regression, random forest) to enhance trade forecasts, increasing accuracy by 75%.
- Automated reporting systems, reducing manual effort by 20%, while delivering actionable insights to improve operational efficiency.

PROFICIENCIES & SKILLS

Languages: Python, Scala, SQL, R

Libraries: Pandas, NumPy, Scikit-Learn, TensorFlow, PyTorch, Matplotlib, Seaborn

Frameworks: Apache Spark, Hadoop, Flink, Kafka, Airflow

Development Tools: Git, Docker, Kubernetes, AWS (S3, Glue, EMR), Google Cloud Platform (BigQuery, Dataflow), Azure (Data Factory, Synapse Analytics)

Databases: PostgreSQL, MySQL, MongoDB, Neo4j, Snowflake, Redshift

Data Visualization Tools: Tableau, Power BI