Fariha Shah

(253) 392-9379 | Fariha_shah@outlook.com | Seattle, WA | LinkedIn | GitHub | Open to relocate

SUMMARY

•**Data Scientist and Data Engineer** experienced in building and maintaining **scalable data pipelines across HRTech and Oil, Gas sectors. MS** in Data Science from **Seattle University, Washington**

• Proficient in **Python, SQL, and Spark;** Specialized in building scalable **ETL pipelines, distributed data processing, and cloud-based data infrastructure**

SKILLS

• Languages, Tools: **Python, SQL (Postgres, MySQL, DynamoDB), Pyspark, C#, Bash, HTML, CSS, Angular,Java,Excel,R**

• Statistics: **Hypothesis Testing, A/B Testing, Predictive Modeling, Forecasting, Machine Learning, NLP**

• Big Data, Processing: **Apache Spark, Hadoop, Kafka, PySpark, Airflow**

• Cloud: A**WS (Lambda, S3), Kubernetes, Docker, Git, GCP Bigquery,**

• Data Tools: **Tableau, ETL Pipelines, REST APIs, DataBricks**

EDUCATION

Seattle University, Washington June 2026
Master of Science in Data Science, GPA: 3.58/4.0

Dawood University of Engineering and Technology, Karachi, Pakistan Dec 2020
Bachelor of Engineering in Petroleum and Natural Gas, GPA: 3.62/4.0

PROFESSIONAL EXPERIENCE

Data Engineer II | FLOWHCM | Karachi, Pakistan July 2020 – Aug 2024

• **Engineered scalable ETL pipelines using Spark, Redshift, S3, and Python, achieving 99.8%** pipeline uptime while integrating payroll, performance, and employee data from multiple HRM modules.

• **Automated the processing of billions of HR records**, cutting manual **efforts by 29% monthly** and enhancing data accuracy for HR reporting.

• **Ingested data from various HR systems**, including time-tracking, recruitment, and employee management platforms, using SQL, REST APIs, and custom connectors to create unified data views for HR analytics.

• **Developed and maintained data pipelines** that support HR reporting tools, enabling HR managers to track key metrics.

• **Collaborated with HR analysts and project managers** to optimize data workflows, resulting in a **26%** improvement in efficiency for HR-related KPIs.

Data Engineer Intern | Pakistan Petroleum Limited | Karachi, Pakistan Jan 2020 – June 2020

• **Developed and implemented ETL** processes to ingest transactional and sensor data from oil field operations, saving over **$85,000** annually by reducing reliance on external data vendors.

• **Collaborated with field engineers** and analysts to understand operational needs and translated these into actionable reports using Tableau, **saving 17 hours of manual work** each week.

• **Utilized Spark in Python to distribute data processing on large datasets** from drilling and production operations, improving data ingestion and processing speed by **67%.**

PROJECTS

Flood Risk Prediction using ML & SARIMA | Asia-wide Climate Data | *Python, Scikit-learn, Climate Datasets, Machine Learning* Apr 2025- June 2025

• Built a predictive framework for flood vulnerability across Asian countries using **SARIMA and regression models (Ridge, Lasso, RF).**

• Merged multi-source data (**historical flood records, GDP, CO_2, satellite zones**) and created a geospatial risk map using **GeoPandas**.

Educational Data Pipeline: ETL and Reporting for Academic Metrics | *Python, MYSQL, Tableau* Jan 2025 - Mar 2025

•Built a university academic database by integrating UOW APIs through **Python-based ETL pipelines,** ensuring data quality and consistency in **MySQL.** Designed **stored procedures, advanced SQL queries, and Tableau dashboards** to visualize student performance and institutional metrics.