OWAIS AHMAD

Data Engineer
Islamabad, Pakistan | +92 326 9796217 | owaisahmad.ds@gmail.com
linkedin.com/in/owais-ahmad-79a431292 | github.com/Owais4077

PROFESSIONAL SUMMARY

Data Engineer with production experience building ETL/ELT pipelines and Medallion Lakehouses on Databricks, PySpark, and Delta Lake. Built incremental pipelines with automated data quality checks, Airflow orchestration, and CI/CD, and integrated CRM and ERP systems (Odoo) through REST APIs. Focused on turning raw business data into reliable, analytics-ready datasets.

TECHNICAL SKILLS

Languages: Python, SQL

Big Data & Lakehouse: PySpark, Spark SQL, Databricks (Azure), Delta Lake, Auto Loader, Unity Catalog, Medallion Architecture

Data Engineering: ETL/ELT, Incremental Loading (Watermarks, Delta MERGE), Data Quality, Data Quarantine, Dimensional Modeling (Star Schema)

Orchestration & DevOps: Apache Airflow, Databricks Workflows, Docker, GitHub Actions (CI/CD), Pytest

Databases & APIs: PostgreSQL, MySQL, FastAPI, REST APIs, OAuth 2.0, JDBC

BI & Visualization: Power BI (DAX), Databricks SQL Dashboards

PROFESSIONAL EXPERIENCE

Data Engineer & Developer **Oct 2026 – Present**

UMS (Unified Marketing System) | Islamabad, Pakistan

- Build Python and SQL ETL pipelines that sync CRM data (leads, contacts, interactions) into PostgreSQL, processing 5,000+ records daily for reporting and analytics.
- Integrate CRM and business applications through REST APIs, replacing 5+ hours of manual work per week with automated syncs.
- Add validation and quality checks (nulls, duplicates, invalid formats) to CRM data, reducing duplicate and bad records by 30%.
- Work with sales and business teams to turn reporting requirements into clean, analytics-ready tables.

Data Engineer **May 2026 – Oct 2026**

YBIX | Lahore, Pakistan

- Built 5+ ETL/ELT pipelines on Databricks with PySpark and Spark SQL, processing 500,000+ records for analytics and reporting.
- Implemented Bronze/Silver/Gold layers on Delta Lake, using MERGE for incremental loads instead of full reloads, cutting pipeline runtime by 30%.
- Created data profiling, cleansing, and validation steps that made downstream datasets consistent for 3+ dashboards and reports.
- Delivered KPI datasets and dashboards used by 2+ teams for weekly business decisions.
- Documented pipeline architecture and data flows, and built reusable modules shared across 5+ pipelines.

EDUCATION

BS Computer Science (Specialization in Data Science) **2026**

University of Engineering & Technology (UET) Mardan | CGPA: 3.00/4.00

PROJECTS

Enterprise Sales Medallion Lakehouse

PySpark, *Delta Lake*, *PostgreSQL*, *Apache Airflow*, *Docker*, *GitHub Actions*, *Power BI* | *GitHub Repository*

- Built an end-to-end Bronze/Silver/Gold lakehouse ingesting PostgreSQL, REST API, and CSV/JSON sources into a Gold star schema of 145,000+ sales records, standardized to USD.
- Implemented incremental ingestion with timestamp watermarks, SHA-256 file tracking, and Delta MERGE upserts, so reruns never create duplicates.
- Designed a data quality engine with 29 rules (null, range, uniqueness, referential integrity) that quarantined 374 bad records without stopping the pipeline.
- Orchestrated daily and hourly Airflow DAGs and set up CI/CD running 162 automated tests; deployed a live dashboard with an AI assistant on Vercel.

Databricks End-to-End Lakehouse Pipeline

Databricks, *PySpark*, *Auto Loader*, *JDBC*, *Unity Catalog*, *Databricks Workflows*

- Automated ingestion from files (Auto Loader) and relational databases (JDBC) into Bronze/Silver/Gold Delta tables with cleansing, deduplication, and business rules.
- Built metadata-driven Databricks Workflows with monitoring, error handling, and alerts; secured access with Unity Catalog and secret scopes.

Online Retail Lakehouse & Analytics

Azure Databricks, *PySpark*, *Delta Lake*, *Power BI*

- Processed 541,909 retail transactions through Medallion layers, producing 527,390 validated records after deduplication, cancellation filtering, and null handling.
- Built Gold KPI tables for revenue, products, and customers with incremental MERGE loads, connected to Power BI dashboards.

Anviz Attendance to Odoo ERP Integration

Python, *FastAPI*, *PostgreSQL*, *Docker*, *Odoo API*

- Built an automated ETL pipeline pulling punch records from Anviz biometric devices into PostgreSQL and syncing them to Odoo ERP, eliminating manual attendance exports.
- Added employee ID mapping, duplicate prevention, retry handling, and sync-status tracking; containerized with Docker.

LinkedIn Connection Tracker

Python, *FastAPI*, *LinkedIn API*, *OAuth 2.0*, *PostgreSQL*

- Built a FastAPI application with OAuth 2.0 authentication to securely connect to the LinkedIn API and store connection data in PostgreSQL.
- Tracked the full connection lifecycle (Pending, Accepted, Rejected, Withdrawn) with history, a summary dashboard, and automated API tests.

CERTIFICATIONS & TRAINING

- Databricks Data Engineering Training – Databricks Academy
- Introduction to Data Engineering – IBM / Coursera, Feb 2025 (Credential ID: UKQRSDQUEUUA)
- BI Dashboards with IBM Cognos Analytics and Google Looker – IBM / Coursera, Feb 2025 (Credential ID: 5TQDTAHDYCP4)
- Data Analytics and Business Intelligence – IBM, Feb 2024 (Credential ID: PBQRG7QMK)
- Introduction to Data Mining – Dec 2024 (Certificate Code: 7647754)
- Get Started with Data Engineering – IBM
- Artificial Intelligence & Machine Learning – KPITB (Khyber Pakhtunkhwa Information Technology Board)