ASIM RAZA

Data Engineer

Gojra, Punjab, Pakistan | razaasim5700@gmail.com | linkedin.com/in/asimraza3768
github.com/asimraza3768 | asimbuilds.dev

PROFESSIONAL SUMMARY

Data Engineer who turns raw data into reliable, analytics-ready datasets. Builds automated ETL/ELT pipelines with Python, SQL, PySpark, Databricks, Delta Lake, and Apache Airflow, with data quality checks teams can trust. Brings 4 years of client-focused development, hands-on troubleshooting, and team leadership from a national hackathon. Currently building leadership and communication skills through the Aspire Leaders and McKinsey Forward programs.

TECHNICAL SKILLS

Languages & Libraries: Python (pandas, Requests), SQL (joins, relational databases)
Data Engineering: ETL/ELT, Data Pipelines, Data Ingestion, Data Integration, Data Quality, Data Cleaning, Apache Airflow (DAGs), REST API Ingestion
Big Data: Apache Spark, PySpark, Delta Lake, Hadoop (coursework), Kafka (coursework)
Cloud & Platforms: AWS S3, Databricks, Unity Catalog, Snowflake SQL, Cloud Computing Concepts
Data Modeling & Warehousing: Dimensional Modeling, Star Schema, Data Warehousing, Relational Database Design, PostgreSQL, MySQL, MongoDB
Tools: Git, GitHub, Docker, CI/CD (GitHub Actions), Python Testing (pytest), FastAPI, Scikit-Learn, XGBoost
Professional Skills: Team Leadership, Communication, Presentation, Problem Solving, Time Management, Research and Fact-Checking

PROJECTS

Job Market Intelligence Platform | *Python, REST APIs, AWS S3, Databricks, PySpark, Delta Lake, SQL* **Sep 2026**

- Built automated REST API ingestion with cursor-based pagination, landing 400+ job listings in AWS S3.
- Designed a 3-layer ETL pipeline with 4 automated workflow tasks, producing 6 curated analytics datasets.
- Implemented data quality checks for required fields, duplicates, salaries, timestamps, remote values, and country codes.
- Used MERGE-based upserts and deduplication to keep job records consistent across all 3 layers.
- Found the root cause of API failures, restoring ingestion from HTTP 400 errors to successful 200 responses.

Weather Data Engineering Pipeline | *Python, PySpark, Databricks, Delta Lake, Databricks SQL* **Aug 2026**

- Built automated REST API ingestion pipelines for 5 cities using Python, PySpark, and Databricks.
- Developed ETL workflows that turn raw weather data into analytics-ready Delta Lake datasets.
- Added data quality validation to catch integration and transformation issues before analytics.
- Designed a star schema and queried Gold Delta tables with Databricks SQL.

EXPERIENCE

Intern | *Press Network of Pakistan (PNP), Islamabad | On-site* **Oct 2026 – Present**

- Working through a 6-week newsroom internship program covering news writing, fact-checking, media monitoring, and press clipping.
- Building research, clear-writing, and presentation skills through weekly editorial assignments.

Freelance WordPress Developer **2022 – 2026**

- Delivered 3–4 WordPress websites, turning client requirements into working solutions.
- Worked with MySQL databases, custom plugins, and third-party REST API integrations.
- Troubleshot performance, security, and data issues to keep live sites stable.

IT Lab Administrator (Intern) | *Virtual University of Pakistan, Gojra Campus* **2023 (6 months)**

- Set up and maintained lab computers, software, and network access so classes and exams ran without interruption.
- Diagnosed and resolved hardware, software, and connectivity issues, reducing downtime for users.
- Guided students and staff on lab systems and tracked equipment, usage, and reported issues.

Front Counter Associate | *Bonsaii Restaurant, Gojra* **May 2023 – Aug 2023**

- Processed POS transactions accurately and prepared daily expense sheets and reconciliations.

HACKATHON & LEADERSHIP

Team Lead & Data Engineer | *Sentinel AI, Qabil Bano Pakistan (National Hackathon)* **2026**

- Led a team through 2 selection stages to build Sentinel AI, a flood-risk prediction and emergency response platform.
- Ingested live weather data (Open-Meteo API, 17 climate predictors) to feed an XGBoost flood-risk model.
- Built FastAPI REST endpoints with role-based access control and created presentation decks.

AWARDS & LEADERSHIP PROGRAMS

Certificate of Appreciation | *All Pakistan Minorities Heritage Photo Contest, Islamabad* **Apr 2023**

Aspire Leaders Program | *In progress, joined Oct 2026* **Oct 2026**

McKinsey Forward Program | *In progress, joined Oct 2026* **Oct 2026**

EDUCATION

Bachelor of Computer Science (BSCS), Virtual University of Pakistan **2022 – Expected 2027**

CERTIFICATIONS & WHAT I LEARNED

DataCamp

- **Data Engineer in Python:** building end-to-end data pipelines with Python and SQL, from ingestion and cleaning to loading.
- **Associate Data Engineer / Associate Data Engineer in SQL:** SQL for data engineering, including joins, relational databases, and data warehousing basics.
- **Introduction to Apache Airflow in Python:** writing DAGs and scheduling, ordering, and monitoring pipeline tasks.
- **ETL and ELT in Python:** extracting, transforming, and loading data with Python, and when to use ETL or ELT.
- **Data Warehousing Concepts / Database Design:** how warehouses and schemas are structured, and how to design relational databases.
- **Introduction to Snowflake SQL:** querying and managing data in the Snowflake cloud data platform.
- **Software Engineering Principles / Writing Efficient Python Code:** clean, modular, testable Python and ways to make code faster.
- **Cleaning Data in Python / Data Ingestion with pandas:** importing data from files and APIs and fixing messy or missing data.
- **Understanding Cloud Computing:** core cloud concepts and service models.

Coursera and Virtual Programs

- **Introduction to Data Engineering (IBM):** the data engineering lifecycle, data repositories, and pipelines.
- **Python for Data Science, AI & Development (IBM):** Python fundamentals, data structures, pandas, and working with APIs.
- **Successful Presentation (CU Boulder):** structuring and delivering clear, persuasive presentations.
- **Deloitte Data Analytics and Technology; BCG Data Science (Forage):** virtual job simulations in analyzing data and presenting insights.