

Indir Lal

Data Engineer • Python, SQL & PySpark • Azure & GCP • ETL/ELT & Cloud Data Warehousing

+92 305 7091899 | indir.lal.de@gmail.com | linkedin.com/in/indir-lal | github.com/Indir07

SUMMARY

Data engineer and Computer Science graduate with hands-on experience building scalable ETL/ELT pipelines and cloud data warehouses on Azure and GCP. Built a 60M+ record BigQuery warehouse from 500M+ raw records, backed by 3+ years building data-engineering and ML projects. Strong in Python, advanced SQL, PySpark, data modeling, and pipeline reliability turning messy, multi-source data into trusted, query-ready assets.

TECHNICAL SKILLS

Languages & APIs: Python (Pandas, NumPy), SQL, Java, JavaScript / TypeScript, FastAPI, REST APIs

Cloud & Big Data: Azure (ADF, Synapse, ADLS, Databricks), GCP (BigQuery, Cloud Storage), PySpark, Spark SQL, Delta Lake, Medallion / Lakehouse

Data Engineering & Databases: ETL/ELT Pipelines, Data Modeling (Star / Snowflake), Warehouse & Data Lake Design, Data Quality & Validation, Query Optimization, PostgreSQL, MySQL, MongoDB, SQLite

Tools & Practices: Terraform (IaC), Git, GitHub Actions (CI/CD), Docker, Linux, Power BI, Agile

EXPERIENCE

Data Engineer Intern

Jun 2026 – Present

Noble Ventures Data Inc

St. Augustine, Florida, USA (Remote)

- Designed and built a 60M+ record Master B2B/EIN data warehouse in **Google BigQuery**, consolidating and enriching 500M+ raw firmographic records from disparate sources.
- Engineered fault-tolerant **Python ELT pipelines** on Cloud Storage and BigQuery APIs, ingesting and transforming massive datasets (24GB+ CSVs, chunked JSONL, compressed archives).
- Developed multi-phase **SQL matching engines** (exact, fuzzy, geographic) bridging 9.6M+ records, using advanced SQL (window functions, regex) for cleansing and deduplication.
- Built and deployed a high-performance **FastAPI** microservice over BigQuery with API-key authentication, request validation, and audit logging.

Freelance Data Engineer (Fiverr)

Dec 2023 – Present

Independent International Clients

Remote

- Built and maintained **ETL/ELT pipelines** in Python and SQL, transforming messy multi-source MySQL/PostgreSQL data into clean, query-ready warehouse models.
- Standardized a 400M+ row dataset, resolving pipeline failures and data discrepancies, and maintained Level 1 Seller status across international projects.

PROJECTS

CryptoVolt AI Algorithmic Trading Platform

FastAPI, React, TypeScript, PyTorch, XGBoost

- Built a full-stack platform (FastAPI backend, React/TypeScript frontend) with PyTorch LSTM + XGBoost models and sentiment analysis, improving the Sharpe ratio from 0.41 to 0.89 and cutting max drawdown from 34.2% to 19.4%.

Serverless AQI Forecasting Pipeline

Python, scikit-learn, Hopsworks, GitHub Actions

- Built a serverless ML pipeline forecasting air quality 72h ahead feature store, ensemble models, SHAP explainability, and scheduled GitHub Actions (hourly ingest, nightly retrain) with a live dashboard.

Azure Lakehouse Data Pipeline

ADF, Databricks (PySpark), ADLS, Delta Lake, Terraform

- Orchestrated an end-to-end pipeline via Azure Data Factory through ADLS and Databricks into a Medallion (Bronze/Silver/Gold) Lakehouse, with infrastructure as code (Terraform), CI/CD, and Key Vault secrets.

VeriDocs RAG Document Q&A Tool

Python, Hugging Face, TF-IDF Retrieval, Streamlit

- Built an LLM / retrieval (RAG) application answering questions over documents with citations, confidence scoring, and refusal behavior.

EDUCATION

Sukkur IBA University

June 2026

Bachelor of Computer Science

Sukkur, PK

- Sindh Talent Hunt Program Scholarship selected among the top 100 of 5,000 applicants for a four-year undergraduate scholarship. Relevant coursework: DBMS, Data Structures & Algorithms, Probability & Statistics, AI, OOP.

COURSES & CERTIFICATIONS

Intermediate SQL & Intermediate Python (DataCamp) • Data Engineering Essentials (IBM) • Big Data with Spark & Hadoop (IBM / Coursera) • Introduction to MongoDB (MongoDB, Inc.)