

# KELASH KUMAR

Lahore, Punjab, Pakistan · +92 330 3141082 · kelash.iba22@gmail.com  
linkedin.com/in/kelashkumar-iba · github.com/kelashkumar-iba · kelash.me

## SUMMARY

---

Computer Science graduate from Sukkur IBA University, focused on Data Engineering, with three end-to-end portfolio pipelines spanning real-time streaming, Change Data Capture, and Medallion-architecture lakehouse. Includes a production-deployed streaming pipeline live on Oracle Cloud at [streaming-pipeline.kelash.me](https://streaming-pipeline.kelash.me) with automated CI/CD and \$0/month infrastructure cost. Hands-on with Python, SQL, Airflow, dbt, Kafka/Redpanda, Debezium, and Docker. Every project is containerized, orchestrated, and publicly available on GitHub with full architecture docs.

## TECHNICAL SKILLS

---

**Languages & Query:** Python, SQL, Bash/Shell, Java

**Data Engineering:** Apache Airflow, dbt, Kafka/Redpanda, Debezium, Great Expectations, Pandas, ETL/ELT, Data Modeling, Data Warehousing, Medallion Architecture, CDC, SCD Type 2

**Databases & Storage:** PostgreSQL, DuckDB, MinIO (S3-compatible), Parquet, MongoDB

**DevOps & Cloud:** Docker, Docker Compose, Linux, Git, GitHub Actions, CI/CD, Oracle Cloud (OCI), pytest, pre-commit, Makefile, REST APIs, Caddy, iptables, SSH

**Data Visualization:** Metabase, Grafana, Plotly, Microsoft Power BI, Tableau

## PROJECTS

---

### Real-Time Crypto Streaming Pipeline

Live | GitHub

*Python, Redpanda, PostgreSQL, dbt, Airflow, Metabase, Docker, Oracle Cloud, GitHub Actions*

- Engineered a 9-service Docker Compose streaming pipeline ingesting live prices for 10 crypto assets from CoinGecko every ~60s into Redpanda, with a Python consumer implementing poison-pill protection and retry logic for malformed JSON and invalid UTF-8 bytes; sustained ~14,400 updates/day.
- Delivered sub-5-minute end-to-end freshness (API → Metabase) via a 3-layer dbt transform (staging/intermediate/mart) orchestrated by Airflow on a 5-min schedule with a freshness gate that auto-halts downstream runs on stale data.
- Deployed to Oracle Cloud Always Free ARM VM (\$0/month) with Caddy + Let's Encrypt HTTPS and 3-layer security hardening (cloud firewall + iptables + Docker isolation); shipped a 4-job GitHub Actions CI/CD pipeline with safety-gated SSH auto-deploy achieving ~90s push-to-live latency and 8 pytest unit tests passing in 0.14s.

### Change Data Capture Pipeline

GitHub

*PostgreSQL, Debezium, Redpanda, Python, dbt, Airflow, Docker*

- Cut replication lag from hourly-batch (~60 min) to <10s by implementing log-based CDC with Debezium capturing Postgres WAL events into Redpanda, then an idempotent Python consumer writing to the warehouse via primary-key upserts — eliminating duplicates on consumer restart under at-least-once delivery.
- Built SCD Type 2 historical dimensions in dbt for point-in-time analytics on mutable source tables; isolated malformed events via a Python consumer dead-letter queue, preventing bad records from blocking the pipeline.

### Medallion Data Lakehouse with Quality Gates

GitHub

*MinIO, DuckDB, Pandas, Great Expectations, Airflow, Parquet*

- Architected a Bronze → Silver → Gold lakehouse on MinIO with Parquet storage and DuckDB as the in-memory SQL engine, achieving cloud-native lakehouse semantics (object storage + columnar format + SQL compute) at zero cost.
- Enforced data quality at every layer transition with 31 Great Expectations checks (11 Bronze + 20 Silver, including compound (city, date) uniqueness) — blocking bad partitions from promotion to Gold; paired with partition-aware idempotent Airflow DAGs enabling safe backfills and reruns.

## EDUCATION

---

### Sukkur IBA University

Sukkur, Pakistan

Bachelor of Science in Computer Science

Aug 2022 – May 2026

**Relevant Coursework:** Database Systems, Data Structures & Algorithms, Operating Systems, Computer Networks, Software Engineering, System Design

## CERTIFICATIONS

---

- IBM Data Engineering Professional Certificate — Coursera (2026)
- Snowflake Data Engineering Professional Certificate — Coursera (2026)
- Google Data Analytics Professional Certificate — Coursera (2025)

## ACHIEVEMENTS & HONORS

---

**National Skill Competency Test (NSCT), 2026** — Ranked in the 92.2 percentile nationally; top marks in Programming, Database, and AI/ML & Data Analytics.

**Sindh Talent Hunt Program (STHP) Scholarship, 2022** — Fully-funded merit scholarship covering the BS Computer Science program at Sukkur IBA University.