

# Muhammad Husnain

Senior Data Engineer

+92 333 4185494 • [muhhammad.husnain0996@gmail.com](mailto:muhhammad.husnain0996@gmail.com) • [linkedin.com/in/mohusnain](https://www.linkedin.com/in/mohusnain) • [mohusnain.dev](https://mohusnain.dev) • Open to Relocation

Senior Data Engineer with 6+ years designing and scaling multi-tenant analytics platforms on GCP and AWS across advertising, marketing, e-commerce, and SaaS. Strong hands-on experience with Python, BigQuery, Apache Airflow (Cloud Composer), Terraform, Pub/Sub, Cloud Functions, Cloud Run, and serverless architectures. Built orchestration systems powering customer-isolated analytics platforms and delivered 200+ production data pipelines. Expertise spanning ETL/ELT, data modelling, infrastructure-as-code, automation, and cloud-native data architecture.

**Core Stack:** Python • SQL • Apache Airflow • BigQuery • dbt • Terraform • Docker • GCP (BigQuery, Cloud Functions, Cloud Composer, Pub/Sub, Cloud Run) • AWS (EC2, S3, RDS, Redshift) • ETL/ELT • Data Modelling • Data Warehousing • CI/CD

## WORK EXPERIENCE

**Making Science** | Senior Data Engineer

08/2023 - 03/2026

- Designed and operated a multi-channel marketing data platform with a team of 6, integrating 15+ advertising sources (CM360, DV360, Google Ads, Facebook Ads, Criteo, Snapchat) into BigQuery, supporting analytics on \$50M+ ad spend across 2,000+ campaigns and 400+ clients with reporting cadences from daily to 15 minutes.
- Architected a hub-and-spoke BigQuery data model with centralized ingestion and per-client project isolation, building layered dbt models for ad performance metrics (spend, impressions, clicks, conversions, ROAS) consumed by analytics and BI teams.
- Built and orchestrated 20+ Airflow DAGs on Cloud Composer with dependency management, retry strategies, and SLA monitoring to support reliable ingestion and transformation across advertising data workflows.
- Developed and operated 200+ Cloud Scheduler and Cloud Functions workflows for parameterized and sub-daily processing, enabling scalable per-user/per-channel orchestration with reporting cadences ranging from hourly to 15 minutes.
- Migrated legacy Dataproc workloads to serverless Cloud Functions across 20+ customers, reducing infrastructure costs by ~58% (~\$12 → ~\$5 per customer/day).
- Led company-wide Terraform adoption by building reusable modules for 8 GCP services (Pub/Sub, Cloud Functions, Cloud Run, Cloud Scheduler, BigQuery, IAM, Secret Manager, GCS), reducing pipeline deployment time from ~1 hour to 10–15 minutes while eliminating manual provisioning errors.

**LiveEO** | Data Engineer

01/2023 – 08/2023

- Built 8 production-grade Python data workflows consolidating data from distributed AWS environments and third-party providers, supplying ML teams with training datasets for vegetation management, infrastructure monitoring, and deforestation models while processing workloads ranging from 20 GB to 100+ GB per run.
- Collaborated with the Anyscale team to design distributed compute infrastructure using a mix of on-demand and spot EC2 instances, supporting scalable batch ETL workloads while reducing monthly compute costs by ~80% (~\$500 → <\$100/month).
- Built CI/CD pipelines in GitLab CI for three ML teams, deploying containerized workflows to AWS (S3, ECR) via Prefect and replacing manual deployment processes with automated release workflows.
- Partnered with ML engineers and geospatial analysts to translate model training requirements into pipeline specifications, owning end-to-end data delivery SLAs for three downstream ML teams.

**Ryan-Miranda Partners** | Data Engineer

06/2022 – 01/2023

- Developed ETL pipelines processing 10–50+ GB of customer behavioral and transactional data from GA4 BigQuery exports, Postgres, and MySQL with ingestion cadences ranging from 30 minutes to daily, powering analytics workflows for customer behavior and campaign performance analysis.
- Designed and orchestrated AWS-based ingestion and transformation workflows using Python and Amazon MWAA (Apache Airflow), replacing manual spreadsheet-driven reporting processes across 3 client environments.
- Provisioned AWS infrastructure (EC2, S3, RDS, Redshift, Client VPN) using Terraform across development and production environments, reducing deployment time from ~2 days to under 2 hours while minimizing manual provisioning errors.
- Built and optimized Redshift transformation layers supporting downstream analytics, reporting, and customer performance dashboards across 3 client environments.

**Youpal Group** | Data Engineer

12/2020 – 05/2022

- Built ETL pipelines in Python and SQL powering data and ML workflows for an AI-driven job matching platform serving 50+ businesses and 10k+ candidates across Scandinavia.
- Built daily incremental ingestion pipelines integrating candidate profiles, CVs, job descriptions, and CRM data for rule-based and ML-assisted matching workflows, processing a 100+ GB historical dataset alongside ongoing user-driven syncs.
- Built GCP-based data pipelines integrating 10+ third-party video data sources and on-demand API feeds for computer vision workflows involving object tracking and logo detection across live video streams.
- Designed ETL workflows to migrate on-premises datasets to AWS and GCP, contributing to reduced infrastructure dependency and lower on-premises maintenance costs.

**Binary Tech** | Data Engineer

06/2019 – 11/2020

- Built Flask APIs powering the platform frontend, enabling 20+ client companies to configure data sources, build workflows, and trigger ML pipeline execution through a low-code web interface.
- Developed custom connectors for ingesting customer data from PostgreSQL, SQL Server, REST APIs, and third-party platforms while managing the underlying storage layer hosting client datasets within the platform.
- Built drag-and-drop ML model integrations for the platform's visual workflow builder, allowing non-technical users to compose ML pipelines without writing code.
- Contributed to the platform's visual ETL component built on Apache Nifi, supporting customer-built data pipelines as part of a 5-person scrum team.

## TECHNICAL SKILLS

**Languages:** Python • SQL

**Data & Orchestration:** Apache Airflow (Cloud Composer, MWAA) • BigQuery • dbt

**Data Architecture & Engineering:** ETL/ELT • Data Modelling • Data Warehousing • Data Integration • Star Schema & Snowflake Schema Design • Workflow Orchestration • Serverless Data Pipelines • Multi-Tenant Analytics Platforms • Cloud-Native Data Architecture

**Cloud Platforms:**

**GCP:** BigQuery • Cloud Functions • Cloud Run • Pub/Sub • Cloud Scheduler • Cloud Logging

**AWS:** S3 • Lambda • EC2 • MWAA (Managed Apache Airflow)

**Infrastructure & DevOps:** Terraform • Docker • GitHub Actions • GitLab CI • Bitbucket Pipelines

## CERTIFICATIONS

**GCP:** Professional Data Engineer • Cloud Digital Leader • Generative AI Leader

**AWS:** Certified Cloud Practitioner • Certified AI Practitioner

**Azure:** AI-900 (AI Fundamentals) • DP-900 (Data Fundamentals) • AZ-900 (Azure Fundamentals)

**HashiCorp:** Terraform Associate

**Snowflake:** SnowPro Core

## EDUCATION

**Bachelor of Science in Electrical Engineering** | University of Engineering and Technology, Lahore, Pakistan | 2015 – 2019