

Muhammad Husnain

Senior Data Engineer

muhammadhusnain08@gmail.com | linkedin.com/in/mohusnain | mohusnain.dev | Open to Relocation | +92-333-4185494

Senior Data Engineer with six years and eight months of experience across advertising, Earth observation, e-commerce, and SaaS, primarily on GCP. Most recently, I built and operated a multi-tenant advertising data platform. Before that, I built pipelines for geospatial ML training data, e-commerce analytics, and talent matching, and started out building APIs and data connectors for a low-code ML platform. Across 200+ production pipelines in Python and SQL, I have owned four layers: orchestration in Airflow, data in BigQuery and dbt, compute on Cloud Functions and Cloud Run, and infrastructure in Terraform.

Core Stack: GCP (BigQuery, Cloud Composer, Cloud Functions, Pub/Sub, Cloud Run, Cloud Scheduler) | Python | SQL | dbt | Apache Airflow | Terraform | Docker | AWS (S3, EC2, RDS, Redshift) | ETL/ELT Data Pipelines | Data Modeling | System Design | CI/CD | Infrastructure-as-Code

Work Experience

Making Science | Senior Data Engineer

08/2023 – 03/2026

A marketing technology consultancy with over 1,200 employees across 15 markets, working in AdTech, MarTech, and cloud.

- Led backend development of a multi-channel marketing data platform in a team of six, building config-driven ingestion for 15+ ad platforms (CM360, DV360, Google Ads, Meta, Criteo, Snapchat) with idempotent loads, deduplication, and backoff retries into per-client BigQuery projects, covering \$50M+ ad spend and 400+ clients.
- 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) used by BI teams.
- Engineered and owned the Airflow orchestration layer on Cloud Composer for ad-tech pipelines, running 20+ DAGs organized by advertising channel, each serving multiple clients. Implemented dependency management, with SLA monitoring, error tracking, and root-cause analysis via GCP Cloud Logging and Cloud Monitoring.
- 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 to \$5 per customer/day).
- Drove company-wide Terraform adoption with reusable modules for eight GCP services (Pub/Sub, Cloud Functions, Cloud Run, Cloud Scheduler, BigQuery, IAM, Secret Manager, GCS), cutting deploys from one hour to 10–15 minutes.

LiveEO | Data Engineer

01/2023 – 08/2023

A Berlin-based Earth observation company using AI and satellite imagery to monitor railways, power lines, and pipelines, and to verify deforestation-free supply chains for EU compliance.

- Delivered eight Python workflows ingesting Sentinel-2 and commercial satellite imagery into S3 training datasets by date, location, and land patch for vegetation, infrastructure, and deforestation models at 20–100+ GB per run.
- Led the design of distributed compute on Ray with the Anyscale team, mixing on-demand and spot EC2 instances for batch ETL workloads and cutting monthly compute costs by 80% (\$500 to <\$100/month).
- Owned the Prefect orchestration layer, deploying workflows to Prefect Cloud with logging and observability, and took charge of the DataOps team's GitLab CI/CD, fixing existing pipelines and building new ones.
- Managed data delivery SLAs for three downstream ML teams, working with ML engineers and geospatial analysts to turn model training requirements into pipeline specifications.

Ryan-Miranda Partners | Data Engineer

06/2022 – 01/2023

A US data engineering consultancy building cloud data infrastructure for growth-stage businesses, with clients including Ruggable, the direct-to-consumer washable rug brand.

- 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 three client environments.
- Owned the AWS infrastructure layer (EC2, S3, RDS, Redshift, Client VPN), provisioning it with Terraform across development and production environments and reducing deployment time from two days to under two hours.
- Engineered and optimized Redshift transformation layers supporting downstream analytics, reporting, and customer performance dashboards across three client environments.

Youpal Group | Data Engineer

12/2020 – 05/2022

A Stockholm-based technology group that builds and invests in digital ventures, including Yougig, an AI-driven platform matching tech talent with businesses across Scandinavia.

- Owned the ingestion and dbt transformation layers for Yougig, writing and deploying the code that turned S3 and custom CRM data into the candidate and job models behind matching for 50+ businesses and 10k+ candidates.
- Implemented daily incremental ingestion of candidate profiles, CVs, job descriptions, and CRM data for rule-based and ML-assisted matching, processing a 100+ GB historical dataset alongside ongoing user-driven syncs.
- Integrated 10+ third-party video sources and API feeds into Cloud Storage for computer vision models tracking objects and detecting logos, and helped build Docker-based training infrastructure on Compute Engine for 50+ GB of video.
- 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 (now NexusMax) | Data Engineer

06/2019 – 11/2020

A Washington-based startup building a self-service AI and machine learning platform where analysts create predictive models through visual, low-code tools.

- Implemented 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.
- Created 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 five-person scrum team.

Technical Skills

Languages & Data: Python | SQL | Apache Airflow (Cloud Composer, MWAA) | BigQuery | dbt

Data Architecture: ETL/ELT | Data Modeling | Data Warehousing | Star & Snowflake Schemas | Multi-Tenancy

GCP: BigQuery | Cloud Functions | Cloud Run | Pub/Sub | Cloud Scheduler | Cloud Logging | Cloud Monitoring

AWS: S3 | Lambda | EC2 | MWAA (Managed Apache Airflow)

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

Certifications

Google Cloud Platform (GCP): Professional Data Engineer | Cloud Digital Leader | Generative AI Leader

Amazon Web Services (AWS): Certified Cloud Practitioner | Certified AI Practitioner

Microsoft Azure: AI-900 (AI Fundamentals) | DP-900 (Data Fundamentals) | AZ-900 (Azure Fundamentals)

HashiCorp: Terraform Associate

Education

Master of Science in Information Technology

2026 – Present

University of the People, Pasadena, California, USA (Part-time, Online, Weekends)

Bachelor of Science in Electrical Engineering

2015 – 2019

University of Engineering and Technology (UET), Lahore, Pakistan