Talha Ijlal (Data Engineeer)

 muhammad-talha-ijlal | talhaijlal | muhammad.talha.ijlal@gmail.com | +92.325.367.2778

SUMMARY

A dynamic and results-oriented Data Engineer, combining the experience of a seasoned professional with innovation and enthusiasm. With 2 years of experience, I bring a deep understanding of data science theory and practical skills, focusing on finding optimal, creative solutions tailored to each specific challenge rather than forcing problems to fit available tools.

WORK EXPERIENCE

Data Engineer at **DigiLawyer**   July 2024 - Present

– Implemented the data pipeline, fetching data from multiple sources, cleaning and restructuring them, and then feeding to the **PostgreSQL** Database. This also included **vectorizing** the data by contextual **embeddings** using **OpenAI** and **Gemini**'s text embeddings models.

– Employed **Qdrant** and then **pg_vector** to store vectors in **PG17**. Also used **MeiliSearch** and then **pgsearch** for Full Text Search **FTS**.

– Shifted the database from **Docker**ized **instance** to **Kubernetes** cluster. Worked with Elastic K8s Service **EKS** as the destination for the database migration. Also worked with engineering team to setup the cluster.

Data Engineer at **Adara - A Rategain Company**   Jun 2023 - July 2024

– Automated the alerts and **Salesforce** case creation for Support in case of delivery failure. This was done using **Apache Airflow DAGs** which was scheduled to run every hour to check for delivery failures. The environment was setup using **Google Composer**.

– Developed a pipeline to get the DV360 audience sizes for the active segments at Adara's end, and perform basic **analysis** based on sent and displayed sizes, trigger **automated actions** for segments refresh based on the analysis. Tools and Technologies used in this project included **Google API and Services**, Pandas and other Python libraries.

– Designed the workflow for Adara's FBR process which involved writing **Scheduled Google BigQuery queries**, sync services, hosting on **Compute Engine** and uploading of files using **SFTP**. Worked with TC and Data Science teams closely to get the automation part done correctly.

PROJECTS

Digitizing Pakistan's Statutes and Codes   DigiLawyer

Leading the digitization of Pakistan's legal statutes and codes, involving **modeling** a structured schema for legislative documents, **setting up** infrastructure for high-throughput digitization, **coordinating** the team for pipeline execution, and **ensuring** zero-downtime ingestion into DigiLawyer's production database.

ETL pipline for Binary data files   DigiLawyer

Developing an ETL pipeline for **Legal tech startup**, which involves **scraping** judgements from multiple judicial courts' site, **processing** them using **LLM** (OpenAI's GPT 3.5) to get summary, and uploading them to DigiLawyer's Database, hosted on Digital Ocean's **Virtual Machine** setup with Linux **Ubuntu**.

Feed pipeline for an OC store SecondBite

Designed and developed a pipeline to **extract** data from multiple sites, **transform** it for the OC feed format, perform the match check with the existing data and **load** the new entries OR update the previous one based on the latest data. This involved **Selenium** for driving **ChromeDriver** for scraping, **Pandas** for simple transformations and **Google Sheets API** to interact with Gsheets.

EDUCATION

2021 - 2025	Bachelor's of Computer Science at **UET Lahore**	(GPA: 3.48/4.0)
2019 - 2021	FSc. Pre Engineering at **BISE Lahore**	(Grades: 1015/1100)

SKILLS

Automation	Python, Selenium, Chromedriver, Apache Airflow DAGs
Cloud	GCP (BigQuery, Google APIs), AWS (EC2 instances, EKS), Azure (Virtual Machine)
Database	MySQL, PostgreSQL (pgsearch, pg_vector, pgvectorscale), MongoDB, Qdrant, MeiliSearch
Data analysis	Pandas, NumPy, Looker Studio
Web Scraping	Requests, bs4, IP rotating, captcha solving
DevOps	Docker, Kubernetes (K8s)
Also worked with	NodeJS, ExpressJS, Flask, RESTful APIs, GraphQL