

Mateusz C.

Senior Data Engineer

TechHive
IT's simple with us

Professional summary

Data Engineer with almost 7 years of experience building and optimizing data solutions, with a strong focus on GCP and BigQuery. Hands-on experience designing data pipelines, data warehouse models and automated Airflow workflows, as well as developing data models with dbt, PySpark and SQL. Experienced in designing scalable and maintainable solutions, optimizing Spark workloads and reducing cloud resource consumption through architectural improvements. Strong Python and SQL engineering background with experience in OOP, ad-hoc API integrations and data transfer between GCP and on-premise environments. Comfortable working directly with business stakeholders, designing technical solutions, providing project updates and mentoring colleagues.

Education	Bachelor degree in Economics and Management Aston University
Certificates	Mastering Data Modeling Fundamentals
Skills	Python, SQL, JavaScript PySpark, Spark, dbt Apache Airflow, Google Cloud Composer, AWS MWAA GCP, AWS BigQuery, Cloud Storage (GCS), Cloud Composer S3, EMR, Athena Looker PyTest Google Analytics Universal & GA4, Facebook/Instagram Insights & Ads, LinkedIn Marketing APIs

Company	**GFT Technologies**
Project	Banking Project
Period	May 2023 – till now
Role	Senior Data Engineer
Technologies	Python, SQL, GCP, BigQuery, GCS, Cloud Composer, Airflow, dbt, PySpark, Apache Spark, YAML, Ibis, Dataproc

Responsibilities

- Redesigned and reimplemented existing solutions to improve adherence to engineering best practices, object-oriented programming principles, extensibility and scalability;
- Designed and implemented a BigQuery-based solution for exporting data to GCS, eliminating the need for Dataproc and reducing resource consumption;
- Designed, tested and implemented an on-demand data transfer solution from GCS to on-premise NAS servers;
- Designed and implemented data models using dbt and PySpark;
- Designed and implemented an automated approach for creating ad-hoc Airflow workflows based on YAML configuration files for the Ibis framework;
- Architected and implemented a factory pattern for scalable DAG creation based on YAML configuration;
- Optimized Spark jobs to improve data processing efficiency;
- Collaborated with business stakeholders and provided project status updates to upper management;
- Mentored junior colleagues and acted as a technical point of contact in the absence or unavailability of line managers.

Company	**Booksy**
	Internet Marketplace Platforms
Period	Apr 2022 – Apr 2023
Role	Data Engineer
Technologies	Python, SQL, GCP, BigQuery, Airflow, Kubernetes, GKE, Firebase, Looker, Google Ads, Meta Ads, OOP, APIs

Responsibilities

- Led a case study for automating Kubernetes pod deployments through Airflow for Booksy data extraction workflows, reducing GKE pod usage by over 90%;
- Designed and implemented an automated data pipeline for sending conversion events to Meta and Google Ads platforms for advertising optimization;
- Developed data models combining daily Firebase account registration events with business data and calculated LTV metrics for different account categories;
- Implemented utility Airflow DAGs for migrating complete BigQuery datasets between projects;
- Created and implemented new data integrations with the Booksy data warehouse;
- Developed ad-hoc third-party API integrations using object-oriented programming;
- Developed complex SQL queries to transform and transfer data from the data lake to the data warehouse;
- Created dashboards in Looker to support data analysis and business reporting.

Company	**Salts Healthcare**
	Medical Equipment Manufacturing
Period	Jan 2020 – Apr 2022
Role	Data Engineer
Technologies	Python, T-SQL, MSSQL, Pandas, Airflow, Selenium, BeautifulSoup, Django, Dash, Tableau, Advanced Excel, VBA, REST APIs

Responsibilities

- Created and maintained data pipelines for data acquisition, transformation and delivery;
- Identified new datasets and established ETL pipelines using various third-party data APIs;
- Automated recurring reporting processes and data workflows;
- Developed internal web applications to support business and sales teams;
- Developed data cleansing and transformation workflows using Python and Pandas;
- Automated dataset downloads and data collection through web scraping;
- Orchestrated ETL workflows using Airflow;
- Developed ad-hoc data visualizations and analytical applications for business users;
- Created and maintained recurring reports and visualizations for monthly and weekly business reporting;
- Supported legacy applications and reporting processes using VBA and Advanced Excel.