

Best Technology Talent For Companies Around The World



Matheus

Senior Data Engineer

Summary

Certified Data Engineer with 7+ years of experience designing and implementing scalable, cloud-based data solutions for e-commerce, retail, and technology consulting. Currently serving as Data Platform Tech Lead, spearheading a multi-cloud Big Data platform built on Kubernetes, Kafka, Apache Spark, and Airflow handling massive-scale data ingestion and processing.

Implemented automated data lineage tracking for Spark and Airflow workflows using LangChain and LLM, ensuring traceability and compliance. Built high-performance ETL pipelines in Python, managed SQL transformations with dbt and Cosmos, optimised BigQuery data warehouse architectures, and developed custom Kubernetes operators in Python and Golang. Earlier delivered a Machine Learning Platform on Azure Kubernetes, and large-scale GCP data pipelines.

Holds four industry certifications: GCP Professional Data Engineer, Certified Kubernetes Administrator, Certified Kubernetes Application Developer, and HashiCorp Terraform Associate.

TECHNICAL SKILLS:

- GCP (BigQuery, Cloud Storage, Cloud Build), Azure, multi-cloud architectures
- Apache Spark (PySpark), Apache Airflow (Cosmos/dbt), Apache Kafka, Kafka Connect (CDC), Apache Druid, Apache Superset, Apache Beam
- ETL/ELT pipelines, dbt (Data Build Tool), data lineage, data lake design, SQL transformations, Elasticsearch integration
- Kubernetes: CKA/CKAD certified, custom operators (Python, Golang), Helm Charts, kubectl, cluster debugging, GitLab CI/CD deployment
- Terraform, Docker, Prometheus, Grafana, Alertmanager, Slack integration, GitLab CI
- Python, SQL, Golang, YAML
- AI/ML: LangChain, LLMs (data lineage automation), Kubeflow, ArgoCD, MLOps pipelines, Flask/Falcon APIs, recommendation systems, time series
- MySQL, PostgreSQL, BigQuery

Professional Career

Data Engineer / Consulting Tech Lead

July 2023 – Present | Brasília, Brazil

Project Info:

Technical Leader of the Data Platform for one of Brazil's largest e-commerce and retail groups. Responsible for leading the development of Big Data solutions and scaling the company's multi-cloud data infrastructure.

Responsibilities:

- Led and guided the development of the multi-cloud Big Data platform
- Migrated scheduled SQL from BigQuery to dbt running on Airflow using Cosmos
- Implemented effective monitoring with Prometheus, Grafana, Alertmanager, and Slack integration to ensure maximum availability and performance
- Created innovative data ingestion products simplifying the data lake ingestion process using Python and YAML
- Developed native Kubernetes operators in Golang, optimising orchestration and automation processes
- Implemented automated data lineage tracking for Spark and Airflow workflows using LangChain and LLMs, ensuring traceability and compliance
- Established a robust Apache Spark on Kubernetes data processing environment for scalable batch workloads
- Technically led the data platform team, providing mentorship, guidance, and strategic decisions on multi-cloud data infrastructure
- Collaborated with cross-functional teams and stakeholders to align goals, improve processes, and implement best practices

Software and technologies:

Python, SQL, Golang, Kubernetes, Docker, Apache Spark, Apache Airflow, Kafka, Kafka Connect, Druid, Superset, dbt, Cosmos, Prometheus, Grafana, Alertmanager, GCP, BigQuery, MySQL, PostgreSQL, LangChain, LLMs

Kubernetes Developer

July 2022 – July 2023 | Brasília, Brazil

Project Info:

Belivered and maintained the MFlow Machine Learning Platform: a UI/API running on Kubernetes (Azure) that enables users to train and deploy ML models, using Apache Airflow as the orchestration engine.

Responsibilities:

- Managed all interaction between the MFlow API and Apache Airflow using the Python API, creating Airflow DAGs for ML training and deployment orchestration
- Maintained all platform applications (API, UI, Apache Airflow) on Kubernetes using Helm Charts and CI/CD pipelines
- Debugged issues across the Kubernetes cluster using kubectl command-line tooling

Software and technologies:

Python, Azure, Docker, Kubernetes, Helm Charts, Apache Airflow, CI/CD

Senior Data Engineer

November 2020 – July 2022 | São Paulo, Brazil

Project Info:

Data engineering on GCP, building ETL/ELT pipelines to migrate data from databases to a Data Lake on Google Cloud Storage, and creating semantic tables in BigQuery.

Responsibilities:

- Built ETL/ELT pipelines using Python, SQL, and Apache Spark (PySpark) to migrate data from databases to the Data Lake on Google Cloud Storage
- Created semantic tables on Google BigQuery using Airflow as the scheduler; stored all pipelines and code in Git with GitLab CI deployment
- Provisioned and managed all GCP Data Platform resources using Terraform
- Integrated Data Lake data with external systems including Elasticsearch, Kafka, and REST APIs using Python, PySpark, and Apache Beam; deployed on Kubernetes via GitLab CI

Software and technologies:

Python, Apache Spark (PySpark), Apache Beam, SQL, GCP, BigQuery, Cloud Storage, Airflow, Terraform, Kubernetes, Elasticsearch, Kafka, GitLab CI, Git

Data Scientist

August 2019 – November 2020 | São Paulo, Brazil

Project Info:

Built MLOps pipelines and real-time prediction APIs for recommendation systems and time series models on Google Cloud Platform.

Responsibilities:

- Built MLOps pipelines using Docker, Kubernetes, Kubeflow, ArgoCD, Git, and Cloud Build on GCP
- Developed recommendation system and time series solutions
- Built backend applications with Python (Flask and Falcon) to expose Machine Learning models and serve real-time predictions

Software and technologies:

Python, GCP, Kubernetes, Docker, Kubeflow, ArgoCD, Cloud Build, Airflow, Terraform, Apache Spark, SQL, Flask, Falcon

Education

2013 – 2018

University of Brasilia

Master's Degree in Electrical and Electronics Engineering

2019 – 2020

Brasília Center for Higher Education

Artificial Intelligence

Certifications

GCP — Google Cloud Professional Data Engineer (July 2022)

The Linux Foundation — CKA: Certified Kubernetes Administrator (May 2021)

The Linux Foundation — CKAD: Certified Kubernetes Application Developer (November 2020)

HashiCorp Certified — Terraform Associate (May 2021)

Foreign Languages Skills

Portuguese: Native

English: Advanced