



Chiheb Mhamdi

Data Engineer | Streaming and Production AI

✉ mhamdichiheb600@gmail.com ☎ +49 17684757673 📍 Schweinauer Hauptstraße 6b , Nürnberg

🔗 LinkedIn 🗣 Medium 🐙 Github

Data Engineer with 4+ years of experience in enterprise data and AI platforms. Built Kafka streaming and anomaly-detection pipelines on AWS at about 800,000 events per day, and redesigned BigQuery and Dataflow on GCP to cut analytics reports from 5 hours to 55 minutes. Reduced data lake ingestion from 3.5 hours to 28 minutes, cut LLM inference costs by about 35 percent on Kubernetes, and improved RAG chat latency from 45 seconds to 18 seconds on OpenShift.

Certificates

Databricks Certified Data Engineer - Associate

In progress (target: 2026)

Certified Kubernetes administrator

2025/02

AWS Certified Solutions Architect - Associate

2024/01

AWS Certified Cloud Practitioner

2023/12

Work Experience

10/2025 – Present
Nürnberg

Data Engineer, Bayerisches Landesamt für Steuern (LFST)

- Diagnosed PostgreSQL connection-pool saturation (393 of 500 connections in use, with about 120 idle connections per API pod), corrected pool limits, and documented a repeatable validation workflow for the team.
- Tuned asynchronous processing with Celery and Redis by analyzing queue depth, worker isolation, and concurrent user load, reducing p95 chat response time from about 45 seconds to about 18 seconds during peak office hours.
- Hardened Red Hat OpenShift deployments (route and nginx timeouts, Celery Beat access controls, readiness probes, and resource limits) and wrote runbooks for deploy, rollback, and on-call handoff.
- Created architecture diagrams and internal guides on incident response, RAG retrieval behavior, and LiteLLM/vLLM gateway routing for developers and operations.

Technologies used: Python, Red Hat OpenShift, PostgreSQL, Celery, Redis, nginx, LiteLLM, vLLM, retrieval-augmented generation (RAG), document indexing, REST APIs, Git, incident response / runbooks

12/2021 – 09/2025
Nürnberg, Germany

Data Engineer, Ancud IT

Selected clients: Continental AG, Valeo, Fraunhofer, Maisel, HKN

- Built Apache Kafka sensor ingestion and anomaly-detection pipelines on Amazon Web Services (AWS) for Continental AG using Virtual Private Cloud (VPC), AWS Fargate, and Elastic Container Registry (ECR), processing about 800,000 events per day across two pilot production lines; maintenance alerts surfaced anomalies within about three minutes and contributed to an about 14 percent drop in unplanned downtime over a six-month rollout on those lines.
- Reduced SharePoint-to-data-lake ingestion for Maisel from about 3.5 hours to about 28 minutes using nine Apache Airflow DAGs and Python pipelines; added 62 automated data-quality unit tests and Prometheus/Grafana monitoring for a MinIO data lake on a three-node Kubernetes cluster hosted on Hetzner.
- Set up a hybrid large language model (LLM) platform on Kubernetes for Maisel (Ollama, OpenWebUI, and an LLM proxy) for 45 internal users, routing on-premises and cloud models to meet European Union data-residency needs, and cut monthly public-API inference spend by about 35 percent (about 1,200 euros per month) versus sending all traffic to public APIs.
- Redesigned Google BigQuery schemas and Google Cloud Dataflow ingestion for Valeo, loading about 22 gigabytes per day and cutting a recurring plant analytics report from about 5 hours to about 55 minutes.
- Automated email infrastructure for HKN by customizing an open-source exchange stack, integrating SOAP/XML services, containerizing with Docker, and shipping continuous integration and continuous delivery (CI/CD) with GitHub Actions for 280 mailboxes with nightly archiving of about 40 gigabytes per month.
- Delivered a reinforcement-learning traffic-control pipeline for Fraunhofer (SUMO simulation, Kafka and Protocol Buffers aggregation, MLflow and Kubeflow experiment tracking), logging 68 reproducible experiments and achieving about an 11 percent reduction in mean vehicle wait time on the simulation benchmark network (not live city traffic).

Technologies used: Python, Apache Kafka, Apache Airflow, Amazon Web Services (AWS: VPC, EC2, ECR, Fargate), Google Cloud Platform (GCP: BigQuery, Dataflow, Cloud Storage), Docker, Kubernetes, MinIO, Hetzner, Nginx / Ingress, Prometheus, Grafana, SharePoint (ingestion source), Ollama, OpenWebUI, LLM proxy, MLflow, Kubeflow, SUMO, Protocol Buffers (Protobuf), Git, GitHub Actions, CI/CD, SOAP, XML, Streamlit, SQL, unit testing

Education

09/2020 – 07/2022 Ariana, Tunisia	Master in computer science , <i>Esprit , Higher School of Engineering and Technology</i>
09/2017 – 07/2020 Ariana, Tunisia	Bachelor in computer science , <i>Esprit , Higher School of Engineering and Technology</i>

Languages

Arabic — Mother tongue

French — Bilingual

English — Excellent

German — Intermediate (B1, currently studying for B2)

Publications

06/11/2023	Data Engineering : Istio Service Mesh and Envoy, <i>medium</i>
15/01/2024	LocalGPT for Data engineers, <i>Medium</i>
01/09/2023	Real Time Anomaly detection for data engineers, <i>medium</i>
16/03/2023	How to Understand Data Warehousing: A Comprehensive Guide, <i>medium</i>
30/11/2022	CDC with Debezium Kafka connector, <i>medium</i>
09/11/2022	How to scale your application server with celery, <i>medium</i>
07/11/2022	Real Time Notification using Celery , Redis and web sockets with Django, <i>medium</i>

Skills

CORE: — Python, SQL, Apache Kafka, Apache Airflow, Docker, Kubernetes • **CLOUD & DATA:** — Amazon Web Services (AWS: VPC, EC2, ECR, Fargate), Google Cloud Platform (GCP: BigQuery, Dataflow, Cloud Storage), MinIO, ETL/ELT, data lakes, data warehouses • **DEVOPS & PLATFORM:** — Git, GitHub Actions, GitLab CI, Ansible, Prometheus, Grafana, CI/CD, Infrastructure as Code (IaC), Red Hat OpenShift, nginx • **AI / ML OPS:** — retrieval-augmented generation (RAG), LiteLLM, vLLM, Ollama, OpenWebUI, MLflow, Kubeflow, anomaly detection • **DATA STORES & QUEUES:** — PostgreSQL, Celery, Redis, Protocol Buffers (Protobuf) • **INTEGRATION:** — REST, SOAP, XML, GraphQL • **CERTIFICATIONS:** — Certified Kubernetes Administrator (CKA), AWS Solutions Architect Associate, AWS Cloud Practitioner