

# HELENA AIRFLOW-EICHINGER

Data Engineer

BERLIN, DEUTSCHLAND

+49 30 3456 7890 | helena.eichinger@example.de | helena-eichinger.dev

linkedin.com/in/helena-eichinger | github.com/helena-eichinger

## Profil

Data Engineer mit M.Sc. Data Science (TU Berlin, Note 1,3) und 4 Jahren Erfahrung bei About You SE und Zalando SE. Spezialisiert auf Snowflake-, Spark- und Airflow-basierte ETL- und ELT-Pipelines fuer E-Commerce-Daten mit taeglichem Datenvolumen von 8,4 TB. Verantwortlich fuer 84 produktive Pipelines mit 99,7% SLA-Uptime.

## Berufserfahrung

### Data Engineer

11/2022 - heute

Zalando SE

Berlin, Deutschland

Data Platform Team fuer Recommendation- und Pricing-Pipelines

- Verantwortung fuer 38 produktive Spark-on-Databricks-Pipelines mit 8,4 TB taeglichem Volumen und 99,7% SLA-Uptime
- Migration des Recommendation-Feature-Stores von Redshift auf Snowflake + Iceberg, Query-Latenz von 9,8 s auf 1,4 s reduziert
- Aufbau einer Kafka-zu-Snowflake-Pipeline mit Snowpipe Streaming, Ingestion-Latenz von 8 min auf 18 s
- Implementierung von 380 dbt-Tests mit Great Expectations, Daten-Qualitaets-Score von 87% auf 98% erhoeht
- Mentoring von 2 Junior-Engineers, 18 Pair-Programming-Sessions und 4 interne Spark-Tuning-Schulungen

### Werkstudent + Junior Data Engineer

10/2020 - 10/2022

About You SE

Hamburg, Deutschland

Marketing-Analytics-Team mit Fokus auf Attribution-Pipelines

- Aufbau einer Multi-Touch-Attribution-Pipeline in PySpark auf AWS EMR fuer 18 Mio. Sessions/Tag
- Migration von Redshift-SQL-Stored-Procedures zu dbt Cloud, 320 Modelle in 6 Monaten konsolidiert
- Senkung der Snowflake-Compute-Kosten von 38.000 EUR auf 22.000 EUR pro Monat durch Cluster-Sizing
- Aufbau eines Dagster-Sensors fuer Schema-Drift in 184 Fivetran-Quellen, MTTR von 6 h auf 28 min

## Ausbildung

### M.Sc. Data Science

TU Berlin

Data Science

Berlin, Deutschland

10/2020 - 09/2022

1,3

### B.Sc. Informatik

TU Berlin

Informatik

Berlin, Deutschland

10/2017 - 09/2020

GPA: 1,5

## Projekte

### dlt-snowflake-adapter (Open Source)

01/2023 - heute

Erweiterung von dlt fuer Snowflake-Bulk-Loads mit COPY INTO, 1.180 GitHub-Stars, 4.200 woeentliche pip-Downloads

### Masterarbeit: Cost-Aware Query Planning in Lakehouse-Stacks

06/2022 - 12/2022

Vergleich von Iceberg vs. Delta Lake auf 280 TB Datensatzen, Best-Paper-Award DSAA 2023, Note 1,1

## Fähigkeiten

Python (pyspark, polars, pandas) • SQL + dbt Core + Cloud • Apache Spark + Databricks •  
Apache Airflow + Dagster • Snowflake + BigQuery • Kafka + Confluent Cloud •  
AWS Glue + EMR + Lambda • Delta Lake + Iceberg

## Zertifikate

Google Cloud Professional Data Engineer	09/2024
Databricks Lakehouse Data Engineer Professional	06/2024
Snowflake SnowPro Data Engineer	02/2024
Confluent Certified Apache Kafka Developer	11/2023

## Sprachen

Deutsch	Englisch
Muttersprache	C2

## Publikationen

Cost-Aware Query Planning in Lakehouse Architectures	10/2023
Best-Paper-Award fuer Lakehouse-Query-Optimization auf der IEEE Data Science and Advanced Analytics Konferenz	

## Stärken

### Pipeline-Architektur

Designe Pipelines mit klarer Trennung von Ingestion, Transformation und Serving Layer und dokumentiere ADRs pro Iteration

### Daten-Qualitaet

Implementiere Quality-Gates mit Great Expectations und Soda in der CI-Pipeline, Blocker-Builds bei kritischen Verletzungen

### Cost-Engineering

Reduzierte Snowflake-Compute-Kosten in 3 Monaten um 32% durch Query-Refactoring und Cluster-Sizing