



# Konstantin Lakehouse-Holzweissig

## Senior Data Engineer

konstantin.holzweissig@example.de +49 89 4567 8901

Muenchen, Deutschland

konstantin-holzweissig.dev

linkedin.com/in/konstantin-holzweissig

github.com/konstantin-holzweissig

## PROFIL

Senior Data Engineer mit M.Sc. Informatik (TUM) und 8 Jahren Erfahrung in skalierbaren Datenplattformen bei Delivery Hero SE und Allianz SE. Tech-Lead eines 6-koeufigen Data-Engineering-Squads mit Fokus auf Spark, Snowflake, Kafka und Databricks. Verantwortlich fuer PB-Skala-Datenpipelines mit BaFin-/VAIT-Anforderungen und 1,8 Mio. EUR jaehrlichen Cost-Einsparungen.

## BERUFSERFAHRUNG

Senior Data Engineer / Tech Lead 06/2022 - heute  
Allianz SE Muenchen, Deutschland

Tech-Lead in der Group Data Platform mit VAIT- und ISO-27001-Anforderungen

- Architektur einer PB-Skala-Lakehouse-Plattform auf Databricks + Iceberg fuer 28 Tochtergesellschaften
- Reduktion der Snowflake- und Databricks-Compute-Kosten um 1,8 Mio. EUR p.a. ueber 18 Monate
- Migration von 184 Legacy-Spark-Jobs auf Databricks Workflows mit Delta Live Tables, MTTR von 4,2 h auf 38 min
- Aufbau eines BaFin-konformen Audit-Layers in dbt + OpenMetadata mit 320 Datenklassifikationen
- Mentoring von 6 Engineers, 4 zur Senior-Stufe befoerdert, 18 ADRs pro Jahr publiziert

Data Engineer 08/2018 - 05/2022  
Delivery Hero SE Berlin, Deutschland

Global Data Platform Team fuer Marketing- und Logistik-Pipelines

- Entwicklung der Echtzeit-Delivery-Tracking-Pipeline auf Kafka + Flink mit Sub-Sekunden-Latenz fuer 18 Maerkte
- Senkung der ETL-Laufzeit von 6,8 h auf 28 min durch Spark-Partition-Pruning und Adaptive Query Execution
- Migration von Airflow 1.x auf 2.x mit 280 DAGs in 4 Monaten ohne Downtime
- Aufbau einer Data-Mesh-Domain-Struktur mit 12 Datendomains und dezentralen dbt-Projekten

## AUSBILDUNG

M.Sc. Informatik 10/2018 - 09/2020  
Technische Universitaet Muenchen (TUM) Muenchen, Deutschland  
Informatik 1,2

B.Sc. Informatik 10/2014 - 09/2018  
Technische Universitaet Muenchen (TUM) Muenchen, Deutschland  
Informatik GPA: 1,4

## FÄHIGKEITEN

Python (pyspark, ray, polars)  
Apache Spark + Databricks  
Snowflake + BigQuery + Redshift  
Apache Kafka + Flink  
Apache Airflow + Dagster + Prefect  
Delta Lake + Iceberg + Hudi  
Terraform + Kubernetes  
dbt + Soda + Great Expectations

## ZERTIFIKATE

Databricks Lakehouse Data Engineer Professional  
11/2024

Google Cloud Professional Data Engineer  
08/2024

AWS Certified Data Engineer Specialty  
05/2024

Confluent Certified Apache Kafka Developer  
11/2023

## SPRACHEN

Deutsch Muttersprache  
Englisch C2  
Franzoesisch B2

## PROJEKTE

iceberg-snowflake-bridge (Open Source) 01/2022 - heute

Open-Source-Bibliothek fuer Iceberg-zu-Snowflake-External-Tables, 2.180 GitHub-Stars, 38 Mitwirkende

Masterarbeit: Streaming-Lakehouse-Architekturen

06/2020 - 12/2020

Evaluierung Flink + Iceberg + dbt fuer Sub-Sekunden-Latenz, Note 1,1, Best Thesis Award

## PUBLIKATIONEN

Streaming Lakehouse at PB Scale 06/2024

Hauptredner auf dem Data + AI Summit in San Francisco mit 4.800 Konferenzteilnehmern

## STÄRKEN

Architektur-Verantwortung

Verantwortet 14 produktive ADRs fuer Datenplattformen mit klarer Trennung von Compute, Storage und Governance

Cost-Optimierung

Senkung der Snowflake-Kosten um 1,8 Mio. EUR p.a. durch Query-Refactoring, Cluster-Sizing und Materialisierungs-Reviews

Mentoring

Begleitet 4 Junior-Engineers zur Beforderung in 2 Jahren, schreibt woechentlich technische Reviews