

# Vivien Head-of-Data-Sigwart



## Big Data Engineer (Spark/Hadoop)

✉ vivien.sigwart@example.de ☎ +49 621 0123 4567 📍 Ludwigshafen, Deutschland  
🌐 vivien-sigwart.dev 👤 linkedin.com/in/vivien-sigwart 📄 github.com/vivien-sigwart

### Profil

Big Data Engineer mit M.Sc. Informatik (TUM, Note 1,3) und 9 Jahren Spezialisierung auf Hadoop- und Spark-Stacks bei BASF SE und Bayer AG. Verantwortlich fuer 18 PB Daten-Estate mit 480 produktiven Spark-Jobs in Cloudera + EMR. Spezialisiert auf Migration von Legacy-Hadoop-zu-Lakehouse-Architekturen mit messbarer Cost- und Latenz-Wirkung.

### Berufserfahrung

**Senior Big Data Engineer** 06/2020 - heute  
BASF SE Ludwigshafen, Deutschland

Industrial Data Lake Team fuer Manufacturing- und Supply-Chain-Daten

- Verantwortung fuer 18 PB Daten-Estate auf Cloudera + EMR + Iceberg mit 480 produktiven Spark-Jobs
- Migration von 6 Hadoop-Clustern auf Iceberg + Spark-on-Kubernetes, Compute-Kosten um 2,8 Mio. EUR p.a. reduziert
- Reduktion der Pipeline-Laufzeit von 18 h auf 38 min durch Spark Adaptive Query Execution und Z-Ordering
- Aufbau einer Cassandra-zu-Iceberg-Streaming-Pipeline fuer 240.000 Sensor-Events/s aus 28 Chemie-Werken
- Mentoring von 4 Engineers, 18 Spark-Tuning-Workshops und 6 ADRs pro Jahr publiziert

**Big Data Engineer** 08/2014 - 05/2020  
Bayer AG Leverkusen, Deutschland

Internes Pharma-R&D-Data-Lake-Team

- Aufbau eines Hortonworks-Hadoop-Clusters mit 14 Knoten fuer 4,4 PB Pharma-R&D-Daten
- Migration von Oracle-DWH auf Hive + Spark, Query-Laufzeit von 6,4 h auf 28 min reduziert
- Implementierung von 184 Apache-Oozie-Workflows fuer ETL-Pipelines, MTTR von 4 h auf 18 min
- Co-Autor der Bayer-Konzern-Hadoop-Standards mit anderen 3 Big-Data-Architekten

### Ausbildung

**M.Sc. Informatik** 10/2012 - 09/2014  
Technische Universitaet Muenchen (TUM) Muenchen, Deutschland

Informatik 1,3

### Fähigkeiten

Apache Spark + PySpark + Spark SQL

Apache Hadoop + Hive + HDFS

Cloudera + Hortonworks + EMR

Hudi + Iceberg + Delta Lake

Scala + Python + Java

Apache Airflow + Oozie

Apache Cassandra + HBase

Kafka + Flink

### Zertifikate

Hadoop + Spark Certified  
Cloudera  
10/2024

Databricks Lakehouse Data  
Engineer Professional  
06/2024

AWS Certified Data Engineer  
Specialty  
03/2024

**B.Sc. Informatik**  
Technische Universitaet Muenchen (TUM)  
Informatik

10/2009 - 09/2012  
Muenchen, Deutschland  
GPA: 1,5

**Confluent Certified Apache  
Kafka Developer**  
11/2023

## Projekte

**spark-hudi-bridge (Open Source)** 06/2020 - heute

Open-Source-Spark-Plugin fuer Hudi-zu-Iceberg-Migration mit Schema-Inference,  
1.480 GitHub-Stars

**Masterarbeit: Spark Adaptive Query Execution auf 1 PB  
Datensaetzen**

06/2014 - 12/2014

TUM, Note 1,2, mit Fraunhofer IAIS Sankt Augustin kofinanziert

## Publikationen

**Hadoop-to-Iceberg Migration at Industrial Scale** 06/2023

Hauptredner auf Berlin Buzzwords 2023 mit 1.840 Live-Teilnehmern

**Snowflake SnowPro Data  
Engineer**

08/2023

## Sprachen

Deutsch Muttersprache

Englisch C2

## Stärken

**Spark-Tuning**

Reduktion der Spark-Job-Laufzeit um  
Faktor 4-12 durch Partition-Pruning,  
Bucketing und Adaptive Query  
Execution

**Hadoop-zu-Lakehouse-  
Migration**

8 produktive Hadoop-zu-Iceberg-  
Migrationen in Konzern-Umfeld mit 18  
PB konsolidiertem Estate

**Cost-Engineering**

Senkung der Hadoop-/EMR-Compute-  
Kosten um 2,8 Mio. EUR p.a. ueber 3  
Jahre bei BASF SE