Lukas Hoffmann | Senior Data Engineer –Cloud & Big Data

Friedrichstraße 123, Berlin, Berlin, Deutschland, 10117
+49 30 12345678 • lukas.hoffmann@example.de • https://www.lukashoffmann.de
@lukashoffmann • @lukashoffmann

Basisinformationen

- Data Engineer mit über 6 Jahren Erfahrung in der Konzeption, Entwicklung und dem Betrieb skalierbarer Datenpipelines
- Spezialisiert auf Cloud-Datenplattformen (AWS, GCP), ETL-Prozesse und die Modellierung von Data-Warehouse-Strukturen
- Fundierte Kenntnisse in Python, SQL, Apache Spark, Kafka und Airflow sowie in der Automatisierung von Datenqualitätsprüfungen
- Teamorientiert, analytisch denkend und leidenschaftlich für saubere, wartbare und performante Datenarchitekturen

Ausbildung

Technische Universität München **https://www.tum.de**

Master, Informatik, Punkte: 1.3 *Okt. 2014–Sept. 2017*

- Schwerpunkt auf Datenbanken, verteilten Systemen und Big-Data-Technologien
- Masterarbeit: „Entwurf einer skalierbaren Echtzeit-Datenpipeline für IoT-Anwendungen“
- Praktische Erfahrung durch Projekte in Kooperation mit Industriepartnern

Kurse: Datenbanksysteme, Verteilte Systeme, Maschinelles Lernen, Big Data Analytics, Datenstromverarbeitung, Cloud Computing, Software Engineering

Universität Hamburg **https://www.uni-hamburg.de**

Bachelor, Wirtschaftsinformatik, Punkte: 1.7 *Okt. 2011–Sept. 2014*

- Grundlagen in Programmierung, Datenbanken und betriebswirtschaftlicher Informatik
- Bachelorarbeit: „Vergleich von ETL-Werkzeugen für heterogene Datenquellen“
- Mitglied der Fachschaft Informatik und Organisation von Workshops zu SQL

Kurse: Programmierung, Datenbanken, Betriebswirtschaftslehre, Statistik, Informationssysteme

Berufserfahrung

DataCraft GmbH **https://www.datacraft.de**

Senior Data Engineer *Jan. 2020–Heute*

- Leitung eines Teams von vier Data Engineers bei der Migration einer monolithischen Datenplattform in eine Cloud-native Architektur auf AWS
- Konzeption und Implementierung von ETL-Pipelines mit Apache Airflow, Spark und dbt zur Verarbeitung von über 5 TB täglich
- Einführung automatisierter Datenqualitätsprüfungen und Monitoring-Dashboards zur Sicherstellung der Datenintegrität
- Enge Zusammenarbeit mit Data Scientists, Analytics-Teams und Stakeholdern zur Ableitung datengetriebener Entscheidungen
- Mentoring von Junior-Engineers und Förderung von Best Practices in den Bereichen Code-Qualität und Dokumentation

Schlüsselwörter: Datenpipelines, ETL, AWS, Apache Spark, Airflow, Data Warehouse

CloudMetrics AG **https://www.cloudmetrics.de**

Data Engineer *Okt. 2017–Dez. 2019*

- Entwicklung und Wartung von Batch- und Streaming-Pipelines mit Python, SQL und Apache Kafka
- Aufbau eines Data Warehouse auf Google BigQuery inklusive Datenmodellierung und Performance-Optimierung
- Automatisierung von Deployments mit Docker, Kubernetes und Terraform
- Durchführung von Datenanalysen und Erstellung von Reports für das Management

Schlüsselwörter: Python, SQL, Kafka, BigQuery, GCP, Terraform

Sprachen

Deutsch: Muttersprache oder zweisprachig **Schlüsselwörter**: Muttersprache

Englisch: Verhandlungssichere Kenntnisse **Schlüsselwörter**: C1, Business fluent

Fähigkeiten

Programmierung & Datenbanken: Experte **Schlüsselwörter**: Python, SQL, Scala, PostgreSQL, MongoDB, Redis

Big Data & Streaming: Erfahren **Schlüsselwörter**: Apache Spark, Apache Kafka, Apache Flink, Hadoop, Pandas

Cloud & DevOps: Erfahren **Schlüsselwörter**: AWS, GCP, Docker, Kubernetes, Terraform, Airflow

Data Warehousing & BI: Erfahren **Schlüsselwörter**: Snowflake, BigQuery, Redshift, dbt, Tableau, Power BI

Auszeichnungen

Mitarbeiter des Jahres –Data Engineering

DataCraft GmbH *Dez. 2023*

Ausgezeichnet für die erfolgreiche Migration der Datenplattform in die AWS-Cloud und die Einführung automatisierter Datenqualitätsprüfungen.

Zertifikate

AWS Certified Data Analytics –Specialty **https://aws.amazon.com/certification/**

Amazon Web Services *Mai 2022*

Google Cloud Professional Data Engineer **https://cloud.google.com/certification/data-engineer**

Google Cloud *Aug. 2021*

Veröffentlichungen

Datenbank-Spektrum **https://link.springer.com/journal/13222**

Skalierbare Echtzeit-Datenpipelines mit Kafka und Spark *Nov. 2021*

Fachartikel über Architekturmuster und Betriebserfahrungen mit Streaming-Datenpipelines in der Cloud.

Projekte

Echtzeit-Analyseplattform für IoT-Daten **https://github.com/lukashoffmann/iot-analy**

Aufbau einer skalierbaren Streaming-Plattform zur Verarbeitung von Sensordaten in Echtzeit. *März 2022–Juni 2*

- Architektur und Implementierung einer Echtzeit-Pipeline mit Kafka, Spark Structured Streaming und AWS Kinesis
- Verarbeitung von ca. 10.000 Ereignissen pro Sekunde mit Sub-Sekunden-Latenz
- Aufbau eines Dashboards zur Visualisierung von Anomalien und Betriebskennzahlen

Schlüsselwörter: IoT, Kafka, Spark, AWS, Echtzeit

Open-Source Data Quality Framework **https://github.com/lukashoffmann/dq**

Entwicklung eines Python-Frameworks zur automatisierten Datenqualitätsprüfung in ETL-Pipelines. *Jan*

- Open-Source-Bibliothek zur Definition und Ausführung von Datenqualitätsregeln in Python
- Integration in Airflow und dbt zur automatischen Validierung von Pipeline-Ergebnissen
- Über 200 Sterne auf GitHub und Nutzung durch mehrere Unternehmen

Schlüsselwörter: Python, Data Quality, Open Source, Airflow

Interessen

Open Source: Apache Spark, Airflow, dbt

Sport: Laufen, Radfahren, Wandern

Ehrenamtliche Tätigkeiten

Data Literacy e.V. **https://www.data-literacy.de**

Dozent für Datenanalyse *Sept. 2020–Heute*

- Durchführung von Workshops zu SQL und Python für Mitarbeitende von Non-Profit-Organisationen
- Entwicklung von Lehrmaterialien und Übungsprojekten zur Datenanalyse
- Beratung von Vereinen bei der Auswahl und Einrichtung von Datenbanken