

Carolyn dbt-Pries

Staff Data Engineer

✉ carolin.pries@example.de

🌐 carolin-pries.dev

☎ +49 89 5678 9012

🌐 linkedin.com/in/carolin-pries

📍 Muenchen, Deutschland

📄 github.com/carolin-pries

Profil

Staff Data Engineer mit M.Sc. Informatik (RWTH Aachen, Note 1,1) und 11 Jahren Erfahrung in Datenplattform-Architektur bei BMW Group und Siemens AG. Cross-Squad-Architekt fuer 4 Data-Engineering-Teams (28 Engineers) mit Fokus auf Lakehouse, Streaming und Data Governance. Verantwortlich fuer 38 PB Daten-Estate mit 99,9% Plattform-SLA.

Berufserfahrung

Staff Data Engineer

BMW Group AI Lab • Muenchen, Deutschland • 01/2021 - heute

Cross-Squad-Architekt fuer Connected-Mobility-Daten mit 38 PB Estate

- Architektur der Group Data Platform 2.0 auf Databricks Unity Catalog + Iceberg fuer 4 Squads (28 Engineers)
- Reduktion der Cloud-Datenkosten um 8,4 Mio. EUR p.a. ueber 3 Jahre durch Cluster-Sizing und Auto-Scaling-Tuning
- Aufbau eines Echtzeit-Streaming-Stacks auf Kafka + Flink fuer 2,4 Mio. verbundene Fahrzeuge mit Sub-Sekunden-Latenz
- Co-Autor der internen Daten-Governance-Policy mit BMW Compliance, ISO 27001 Re-Zertifizierung
- Mentoring von 12 Senior-Engineers, 4 ADRs pro Quartal und 4 Vorstands-Reviews pro Jahr

Senior Data Engineer / Tech Lead

Siemens AG • Erlangen, Deutschland • 08/2015 - 12/2020

Industrial Analytics Platform fuer Manufacturing- und Energy-Daten

- Aufbau des Hadoop-zu-Snowflake-Migrationsprogramms fuer 18 Werke mit 1,4 PB Industrie-Daten
- Reduktion der Pipeline-Laufzeit von 14 h auf 38 min durch Spark Adaptive Query Execution und Z-Ordering
- Tech-Lead von 8 Engineers in 2 Squads ueber 4 Jahre, 3 Beforderungen zum Senior-Level
- Aufbau eines Apache-Atlas-Data-Catalogs mit 12.800 katalogisierten Datensatzen und 4.200 Lineage-Edges

Ausbildung

M.Sc. Informatik

RWTH Aachen • Aachen, Deutschland • 10/2013 - 09/2015

Informatik • 1,1

B.Sc. Informatik

RWTH Aachen • Aachen, Deutschland • 10/2010 - 09/2013

Informatik • GPA: 1,3

Fähigkeiten

Apache Spark + Databricks

Snowflake + BigQuery + Redshift

Apache Kafka + Flink + Pulsar

Delta Lake + Iceberg + Hudi

Python + Scala + SQL

Terraform + Kubernetes + Helm

OpenMetadata + DataHub

dbt + Soda + Monte Carlo

Zertifikate

Databricks Lakehouse Data Engineer Professional

• 10/2024

Google Cloud Professional Data Engineer

• 06/2024

AWS Certified Data Engineer Specialty

• 03/2024

Snowflake SnowPro Data Engineer

• 11/2023

Confluent Certified Apache Kafka Developer

• 08/2023

Projekte

spark-cost-explainer (Open Source)

- 06/2021 - heute

Spark-UI-Plugin fuer Compute-Cost-Attribution pro Job/Stage, 3.840 GitHub-Stars, im Cloudera-Marketplace verfuegbar

Masterarbeit: Adaptive Query Optimization in Federated Lakehouses

- 01/2014 - 06/2015

Best Thesis Award RWTH Aachen, h-Index 6 mit 8 Publikationen

Publikationen

Adaptive Query Optimization in Federated Lakehouses

08/2024 •

Erste VLDB-Industry-Track-Publikation der BMW Group AI Lab, h-Index 6 mit 8 Publikationen

Stärken

Plattform-Strategie

Verantwortet die 5-Jahres-Roadmap fuer die Group Data Platform mit Quartals-Reviews zum Vorstand

Tiefes Cost-Engineering

Reduktion der Cloud-Datenkosten um 8,4 Mio. EUR p.a. ueber 3 Jahre bei BMW Group

Cross-Squad-Mentoring

Mentor fuer 12 Senior-Engineers, fuehrt monatliche internal Data Eng Tech Talks fuer 280 Mitarbeitende

Sprachen

Deutsch

Muttersprache

Mandarin

C1

Englisch

C2