

Trainings

## Data Engineering Challenges: Umgang mit personenbezogenen Daten



Das Training geht auf die Herausforderungen ein, die sich für Data Engineers aus der DSGVO für ihre Datenprodukte ergeben. Gemeinsam erarbeiten wir den gezielten Einsatz von Technologien, um diese Herausforderungen erfolgreich [auf einer Cloud Plattform] zu bewältigen.

Duration: 1 Tag

Target group: Data Engineers, Data Scientists, Product Owner

Die Datenschutz-Grundverordnung (DSGVO) ist 2018 in Kraft getreten und hat zum Ziel, die Datenschutzrechte von EU-Bürgern sicherzustellen und zu vereinheitlichen. Unternehmen und Organisationen haben dadurch einen gesetzlichen Rahmen, wie sie personenbezogene Daten erheben, verarbeiten und speichern können. Um aufwändige Korrekturen zu vermeiden, ist es wichtig, die rechtlichen Anforderungen bereits bei der Konzeption der eigenen Datenplattform zu berücksichtigen. Obwohl die verschiedenen Aspekte der DSGVO dabei als Hindernisse auf dem Weg zu innovativen Lösungen wahrgenommen werden können, bieten moderne Data-Engineering-Ansätze und Konzepte sehr gute Möglichkeiten, um auf rechtlich sichere Weise Mehrwerte aus Daten für Unternehmen liefern zu können.

Ausgehend von allgemeinen Anforderungen der DSGVO im Kontext von personenbezogenen Daten lernen die Teilnehmenden unterschiedliche methodische und technische Konzepte zum Umgang bspw. mit Anonymisierung, Löschanfragen, Aufbewahrungsfristen und Metadaten-Management kennen. Dieser Kurs nimmt dabei eine praktische Perspektive ein, bei der die Teilnehmer:innen prototypische Hands-on-Implementierungen durchführen, die sich in produktiven Praxisprojekten bewährt haben. Im Fokus stehen dabei jeweils ein konkreter Use Case und eine darauf abgestimmte Architektur.

Nach diesem Training können Teilnehmende die sich aus der DSGVO ergebenden Problemstellungen für ihre Datenprodukte beurteilen und durch den gezielten Einsatz von Technologien die Herausforderungen erfolgreich [auf einer Cloud Plattform] bewältigen.

## Agenda

- DSGVO & Anforderungen an Data Engineers
- Löschkonzepte & Architekturen
  - Umgang mit personenbezogenen Daten (PII, Data Privacy)
  - Verschlüsselung/Anonymisierung (Hashing, SHA-256, k-Anonymity)
  - Privacy-aware Table Design (Metadata, Tagging, Schema)
  - Retention Policies, Archival Processes
- Hands-on-Implementierung
  - Updates & Deletes auf großen Datenmengen mit Spark & Delta Lake
  - Metadaten-Management mit Data Catalogs (Data Discovery)
- Data Governance
  - Notwendige Prozesse in Ihrer Organisation
  - IAM & Rechtekonzept

*Die Trainingsumgebung wird auf GCP aufgesetzt. Die verwendeten Technologien lassen sich aber auch auf den anderen Public Clouds verwenden (oder es gibt ähnliche Services).*

*Das Training ist keine Rechtsberatung.*