

Dieses Projekt implementiert eine skalierbare Dateninfrastruktur zur Batch-Verarbeitung von über 1.000.000 Temperaturdaten deutscher Wetterstationen. Containerisierte Microservices wie Kafka für die Datenaufnahme, Spark für Datenaggregationen und HDFS für die Datenpersistierung gewährleisten eine robuste Datenpipeline. Docker sorgt für Isolation und Wartbarkeit der Architektur, während ausgewählte Maßnahmen Datenschutz, Datensicherheit und Data Governance sicherstellen.