Marko Kolarek

📍 Berlin, Germany　✉ marko.kolarek@proton.me　📞 0162 5182782　🔗 markokolarek.com

Summary

Senior data engineer with extensive experience designing and implementing scalable data architectures on AWS and MS Azure. Expertise includes building robust data pipelines, managing large-scale migrations, and optimizing cloud infrastructure for e-commerce, fintech, and legaltech companies.

Experience

Oct 2025 – Jan 2026　**Senior Data Platform Engineer**, Contentful GmbH – Berlin, Germany (hybrid)

- Designed new data platform architecture.
- Maintained and improved data ingestion pipelines on AWS (Lambda, S3).
- Migrated and backfilled raw log data on AWS (DataSync, S3, Redshift).

Jan 2025 – July 2025　**Fractional CTO**, ATMO GmbH – Berlin, Germany (hybrid)

- Built production-ready carbon management platform end-to-end, using AWS, Python (Django) and Terraform.
- Implemented LLM-based workflow automations using Google Gemini.

Apr 2024 – Oct 2025　**Senior Data Engineering Consultant**, Self-employed – Berlin, Germany (hybrid)

- Designed and implemented data platform architecture.
- Prototyped and productionized machine learning systems.

Mar 2023 – Mar 2024　**Staff Data Engineer**, GROPYUS Technologies GmbH – Berlin, Germany (hybrid)

- Designed and implemented data architecture on and data modeling on MS Azure to ensure integration with IT systems.
- Led daily operations of data teams and collaborated with product management on technical roadmaps.
- Contributed to service mesh integration planning and implementation on MS Azure (HCP Consul, linkerd, Terraform).

Apr 2022 – Feb 2023　**Head of Data**, VAAAM AG – Berlin, Germany (hybrid)

- Managed data and backend engineering teams.
- Designed and implemented data architecture and models (AWS, Python) while collaborating on product strategy.
- Maintained existing services and migrated to new platform (Python, Java).

Oct 2020 – Mar 2022　**Senior Data Engineer**, GROPYUS Technologies GmbH – Berlin, Germany (hybrid)

- Developed event-based streaming pipelines using serverless solutions on MS Azure.
- Implemented digital twin infrastructure based on semantic web stack (RDF).

Mar 2019 – Oct 2020　**Senior Data Engineering Consultant**, DAIN Studios GmbH – Berlin, Germany (hybrid)

- Developed data ingestion and processing pipelines in Apache Spark (Python & Scala) and Apache Hive.
- Designed and implemented customer data platform infrastructure on MS Azure.

Previous Experience

Data Engineer, FlixBus SE (2018): Implemented and maintained data pipelines for feature engineering using PySpark.

Research Engineer, Zalando SE (2016-2018): Developed and automated A/B test analyses using Apache Spark, Pandas.

Backend Engineer, Zalando SE (2015-2016): Built data ingestion and processing pipelines using Python.

Research Associate, Visage Technologies AB (2014-2015): Researched and implemented eye gaze detection models using OpenCV (C++, Python).

Education

Oct 2012 – July 2014   **FER, University of Zagreb**, MSc in Information Processing – Zagreb, Croatia

- Majored in computer vision, developed thesis in C++ (OpenCV).

Oct 2008 – July 2012   **FER, University of Zagreb**, BSc in Information Processing and Multimedia Systems – Zagreb, Croatia

- Majored in computer vision, developed thesis in C++ (OpenCV).

Skills

Programming: Python (PySpark, Pandas, Flask, Django), Scala (Spark), SQL, Terraform

Cloud infrastructure: Amazon Web Services (S3, RDS, Redshift, EMR, Lambda, EC2, ...), Microsoft Azure (AKS, Functions, Blob Storage, Event Hub, ...), Kubernetes, Docker

Databases: Postgres (Amazon RDS, Azure Database), Amazon Redshift, Google BigQuery, EXASOL, MongoDB, Cassandra

Data: Apache Spark, Apache Hive, Apache Kafka

Languages: English (fluent), Croatian (native), German (basic)

Interests

Running, hiking and cats.