

# Curriculum Vitae Daniel David Kovacs

Software Architect | Lead Data Engineer  
Winterthur, Switzerland (B permit)

Born: 07 April 1981, Budapest, Hungary  
Email: [daniel.david.kovacs@gmail.com](mailto:daniel.david.kovacs@gmail.com)  
Phone: +41 79 309 26 21  
GitHub: [github.com/kicsikrumppli](https://github.com/kicsikrumppli)  
LinkedIn: [linkedin.com/in/danieldavidkovacs](https://linkedin.com/in/danieldavidkovacs)



---

## Professional Summary

Hands-on software architect, and lead data engineer. I have over 11 years of experience in designing and implementing scalable software and data solutions. I have a keen interest in machine learning. I am skilled in software engineering, data modelling, and cloud platforms. Experienced in designing efficient data pipelines, messaging architectures, managing complex projects across industries including insurance, finance, life sciences, and transportation.

---

## Skills

Data Engineering: Python, PySpark, Databricks, Kafka, Airflow, Polars, Pandas, Dagster, DBT  
Cloud Technologies: Azure (Databricks, Function Apps, CosmosDB, Azure ML), AWS (EMR, S3, Lambda, EC2, EKS)  
Software Development: Python, Java, Scala  
Tools, Frameworks, Databases: FastAPI, Spring Boot, PostgreSQL, CosmosDB, Neo4J, Docker  
Languages: Hungarian (Native), English (C2), German (B2+)

---

## Professional Experience

EPAM, Switzerland	Apr 2022 – Present	Software Architect & Lead Data Engineer
EPAM, Hungary	May 2014 – Apr 2022	Software Architect & Lead Data Engineer

### Lead Data Engineer for Global Reinsurance Leader

Implemented ETL pipelines in a medallion architecture for financial data on Azure Databricks.

- Designed transformation-by-configuration library, enabling scaling to multiple use cases
- Improved pipeline resiliency and maintainability through modular design
- Established project onboarding process

Technologies: Azure Databricks, PySpark, Python, Azure Data Factory, Azure SQL

### Solutions Architect for Global Reinsurance Leader

Designed and lead the implementation of a cloud-native multi-tenant self-service data ingestion platform for insurance contract management with AI-driven intent and attribute extraction, and end-to-end lifecycle governance.

- Authored architecture documentation, securing project approval and successful delivery
- Designed, choreographed event-driven pipelines for on-demand, self-service ingestion across multiple business units
- Implemented append-only event sourced data model for replayable workflows and flexible downstream integrations
- Enforced governance with automated and manual validations, full audit trails and fine-grained RBAC

Technologies: Azure Event Grid, CosmosDB, Function Apps, Key Vault, Entra ID, Spring Boot

## Lead Data Engineer for a Major Swiss Insurer

Implemented cloud migration and general optimisations of existing in-house data transformation pipelines used by the actuarial team for life insurance risk management.

- Migrated local Python pipelines to Azure ML, reducing model evaluation times by 90%
- Improved pipeline scalability via Pandas-to-Polars migration
- Enhanced DAG visualisation backend, accelerating insurance model development cycle

Technologies: Python, Numba, Pandas, Polars, Azure ML, Azure Blob Storage, Docker

## Solutions Architect for a Business Information and Media Company

Scaled proof of concept into a production ready commercial product of on-prem industrial computer vision solution to monitor mining conveyor belts for foreign object alerting, and production reporting. Designed event driven architecture with modular image processing pipeline.

- Implemented modular image processing pipeline on a ZeroMQ backbone separating video sampling, preprocessing, monitoring, and image segmentation and recognition phases; achieved a 10x performance increase in processed frames per second
- Designed resilient messaging for object detection event handling on RabbitMQ
- Integrated MLflow for end-to-end model training, versioning, deployment and inference serving
- Authored solution architecture document. Turned around high risk of delivery into a positive assessment

Technologies: Python, RabbitMQ, ZeroMQ, OpenCV, SQLAlchemy, MLflow

## Lead Data Engineer for International Pharmaceutical R&D Company

Implemented components of a self-service data platform for drug research and discovery, providing cataloging, transformation and lineage of ingested biomedical database corpuses.

- Created Python DSL for schema ingestion, reducing onboarding time, and development cycle
- Enhanced Airflow execution, cutting down ingestion runtime costs with async processing
- Implemented systematic data quality checks

Technologies: Python, Airflow, PostgreSQL, Neo4J

---

## Education

- BSc in Software Engineering – Budapest University of Technology, 2013
- MSc in Architecture and Engineering – Budapest University of Technology, 2007
- Master in Solar Energy Engineering – Högskolan Dalarna, Sweden, 2004