

# Curriculum Vitae

Daniel David Kovacs

## Contact

[daniel.david.kovacs@gmail.com](mailto:daniel.david.kovacs@gmail.com)

+41 79 309 26 21

[github.com/kicsikrump](https://github.com/kicsikrump)

[linkedin.com/in/danieldavidkovacs](https://linkedin.com/in/danieldavidkovacs)

## Summary

With over 10 years of experience in software engineering and data engineering, I currently work as a Software Architect and a Lead Data Engineer at EPAM Systems, a leading global provider of digital platform engineering and software development services. I have a keen interest in Machine Learning and a strong background in Data, Data Architectures, and ETL. As a Solutions Architect, I design and implement scalable, secure, and cost-effective solutions for complex data challenges, leveraging AWS technologies and best practices. I also mentor and coach data engineers and software engineers, and collaborate with cross-functional teams and clients to deliver high-quality products and services. Previously, I was a Lead Data Engineer at EPAM Systems, where I managed and executed data engineering projects for travel, life sciences, finance and insurance domains. I have a passion for learning new technologies and skills, and I aim to create innovative and impactful solutions that can transform businesses and society.

## Work Experience

May 2014 – Apr 2022

Epam Systems, Budapest

Apr 2022 –

Epam Systems, Switzerland

### Mar 2024 –

*Lead Data Engineer for a global reinsurance leader*

Implemented ELT pipelines. Improved configurability and resiliency of existing pipeline components. Improved onboarding process.

Technologies and tools: Databricks, Spark, Python, Azure Data Factory, Azure SQL

### Feb 2023 – Mar 2024

*Solution Architect for a global reinsurance leader*

Designed and lead implementation of a data ingestion platform. Solution has achieved efficiency through automating business record creation from emails between insurance brokers and underwriters.

- Created cloud native serverless architecture in Azure with a data model, that scales seamlessly to multiple future insurance domains. This allows an ever growing portion of the business to benefit from the platform.
- Authored and illustrated Software Architecture Documentation, which guided sponsors to an educated decision to green-light the project; it was the vehicle for communicating with technical stakeholders, and guiding the project quickly and successfully through the strict in house review and governance process.
- Defined work breakdown structure, and provided accurate estimations, which has been a foundation for efficiently forming a cross functional project team, and delivering on time despite short deadlines.

Technologies and tools: Event Sourcing, Azure Function Apps, Event Grid, Entra ID, CosmosDB, Key Vault, Spring Cloud Function

## Sep 2022 – Feb 2023

*Lead Data Engineer for Industry-leading Swiss insurer*

Designed and implemented improvements to an existing Python-based in-house data transformation pipeline used by actuarial team to evaluate models for life insurance risk management.

- Designed a gradual migration of the framework backend from Pandas to Polars. This facilitated continued tool utilization, offering early benefits of improved model execution times.
- Designed new DAG visualisation backend based on GraphViz. This contributed to significantly accelerating the development cycle for insurance models.
- Implemented data quality tooling to enhance the reliability of the models.
- Migrated the framework to run jobs in Azure ML, cutting down the overall model evaluation time from a magnitude of hours to a magnitude of minutes.

Technologies and tools: Python, Numba, Pandas, Polars, Azure ML, Data Lake, Blob Storage, Docker, Azure Devops

## Nov 2021 – Apr 2022

*Solution Architect for a Business Information and Media company*

As a solution architect, took part in the productisation of an industrial computer vision application. It has been responsible for controlling and monitoring mining conveyor belts for contamination through cctv cameras. Additional capabilities include reporting based on the classification of the ore.

- Assumed role of Solution Architect in a red project status, tasked to produce a quality Solution Architecture Document, and to design for additional requirements. Timely delivery of the artefacts was key in a positive project assessment.
- Designed a sampling image processing pipeline on top of ZeroMQ, replacing the original monolithic solution. Decoupling the computer vision component enabled a tenfold performance gain.
- Defined software development process, fostered a culture of effective communication and cooperation, resulting in a cohesive project team and contributing to the project's overall success.

Technologies and tools: Python, PostgreSQL, SQLAlchemy, RabbitMQ, ZeroMQ, OpenCV, MLflow

## Jan 2021 – Dec 2021

*Lead Data Engineer for a leading global health technology company recognised for pioneering solutions in electronic health records and healthcare information technologies.*

As a data engineer I have migrated big data pipelines from Cloudera CDH platform to AWS EMR Hadoop cluster.

- Implemented changes in Oozie workflow to migrate from CDH to AWS. Despite a challenging documentation landscape, complex workflows, and slow development cycle, I have delivered over expectations, which has contributed to the team meeting it's planned commitments.
- Designed and implemented testing tools that have streamlined the development process, strengthened the confidence in the delivered solution, and contributed to the overall success of the project.
- Developed thorough documentation for an in-house framework that lacked proper documentation, leading to increased efficiency in the team's development processes.

Technologies and tools: Java, Python, Oozie, Vertica, Hue, Hadoop, AWS EMR, S3, EC2, Zookeeper, HDFS

## **Mar 2020 – Dec 2020**

*Lead Data Engineer for a US-based global pharmaceutical company renowned for its commitment to research, development, and production of innovative healthcare solutions.*

As a data engineer I have worked on a data platform for drug research and discovery process that enables integration to analytical tools and provides data integrity and provenance. The value is efficiency, improved decision making, decision capture, and data lineage.

- Designed and implemented a DSL in Python to provide a scalable developer experience for defining the ingestion schema, transformations, and lineage cataloging. This has significantly reduced the learning curve for the platform, and resulted in the scaling up ingestion pipeline development.
- Implemented systematic data quality checks within the DSL, enhancing overall data quality and paving the way for future automation opportunities.
- Optimised Airflow DAG execution performance by introducing async processing, multithreading, and execution on alternative runtimes. This optimization resulted in increased efficiency and a substantial reduction in runtime costs.
- Introduced Poetry as a dependency management tool, streamlining the development and deployment process. This initiative significantly reduced turnaround time by minimizing dependency-related errors.

Technologies and tools: Python, Pypy, Airflow, PostgreSQL, Neo4J, Poetry

## **May 2019 – Mar 2020**

*Lead Data Engineer for a leading international mobility provider in the global transportation industry.*

As a data engineer I have worked on a self service data discovery platform,

- Implemented robust data pipelines with Debezium, to migrate data from MySQL to data lake via Kafka. This has removed a scalability bottleneck, and improved overall data accessibility and empowered users to derive insights promptly, contributing to more informed decision-making processes.
- Shifted data pipelines from AWS ECS to EKS, which enhanced the overall efficiency, and aided in delivering timely data processing capabilities.

Technologies and tools: Scala, Play Framework, Python, Lenses, Slick, Kafka, Debezium, Confluent Platform, PostgreSQL, AWS S3, Kubernetes

## **May 2014 – May 2019**

*From Junior Software Engineer to Lead Software Engineer for a prominent online travel platform.*

As a software engineer I have worked on the mobile backend services for native applications. From 2016 I have assumed the responsibility of development team lead.

- Introduced improvements to the development process, fostered collaboration in the development team, and an innovative atmosphere, which not only elevated the teams profile, but also inspired changes within the greater project
- Lead innovative experiments, such as a chatbot interface, improved map search for public transportation and personal choice of POIs, monitoring and alerting with anomaly detection, which have lead to an increased surface for business opportunities

Technologies and tools: Java, Maven, Spring Boot / Cloud Stream / Data / Aop, Graphana, Apache ActiveMQ, Kafka, Cassandra, Docker, Kubernetes, Helm

# Education

2013 Budapest University of Technology: BSc in Software Engineering

2007 Budapest University of Technology: MSc in Architecture and Engineering

2004 Högskolan Dalarna, Sweden: Master in Solar Energy Engineering

