

ALPHAN AKSOYOĞLU, Ph.D.

Thomas-Dehler-Strasse 7, 10787 Berlin | M: +49 176 419 163 94 | alphanaks@protonmail.com
LinkedIn: <https://linkedin.com/in/ma-aksoyoglu> GitHub: <http://www.github.com/Alphan-Aksoyoglu>

Summary

Senior Data Engineer with a PhD in Physics, experienced in building, evolving and scaling large scale data platforms for business and product analytics. Experienced in owning and optimizing petabyte (PB) scale data pipelines on cloud infrastructure, delivering high quality, high-freshness datasets consumed by thousands of users running millions of queries per month.

I bring strong analytical and systems thinking, with a rigorous approach to data correctness, governance, and reliability at a scale, covering ETL, ELT design, data modeling, cost and process optimization. Frequently operate as a technical lead within existing platforms, translating business goals and OKRs into scalable data solutions, mentoring engineers, and contributing to architectural and governance decisions across teams.

Technical Skills

Core Data Engineering <ul style="list-style-type: none">Python, SQL (advanced; data pipelines, analytics, automation)BigQuery, PostgreSQL, RedshiftApache Airflow, batch & incremental pipelinesKafka, Pub/Sub, CDCAnalytics & operational data modelingQuery optimization, cloud cost management	DevOps & Infrastructure <ul style="list-style-type: none">Git, GitHub, GitLab, CI/CD pipelinesDocker, KubernetesGoogle Cloud Platform (GCP)Amazon Web Services (AWS)Terraform
Other Programming and Tools <ul style="list-style-type: none">Java, JavaScript, C, C++, MATLAB, Mathematica, Bash	

Work experience (Industry):

Senior Data Engineer

08.2023 – Present

DeliveryHero SE – Berlin, Germany

- Designed, built, and owned end-to-end data curation pipelines for an internally developed Product Analytics platform, processing datasets in the PB (petabyte) scale. Pipelines built around BigQuery and Apache Airflow curated about 6 to 10 TB of event and session data per day, enabling reliable analytics at scale. Further optimized these largescale batch pipelines to support incremental processing up to six runs per day, significantly improving data freshness while maintaining performance and cost efficiency on high volume workloads.
- Architected and implemented a data archival and lifecycle management solution for historical Product Analytics data, resulting in ~€100,000/year in cloud cost savings without impacting analytical use cases.
- Owned and curated business critical core data models consumed by ~1,500 internal users running ~2 million queries per month. Ensuring high data quality, consistency, and trust across analytics, reporting, and operational use cases.
- Proposed, planned, and executed pipeline and data model optimizations (including partitioning, clustering, execution strategy changes, and query rewrites), delivering up to €20,000 per month in cost savings and reducing pipeline runtimes by ~70% (from ~4 hours to ~1 hour).
- Following the merger of batch curation and streaming ingestion teams, contributed to the design and implementation of a prototype CDC (Change Data Capture) based upsert pipeline, utilizing Pub/Sub, for a previously insert only streaming table, demonstrating realtime deduplication and improved data correctness, and establishing a proof of concept for reducing redundant data and downstream processing costs.
- Took full ownership of ingestion and curation pipelines of the Customer Experience team, integrating Qualtrics data and related sources. Refactored pipelines for robustness and efficiency, achieving > 98% uptime and reliable downstream consumption.

- Regularly translated OKRs into Epics and technical execution plans, created and prioritized tickets, delegated tasks, and led teammates through delivery, frequently compensating for gaps in product ownership to ensure successful completion of high impact initiatives.

Data Engineer II

01.2022 – 08.2023

DeliveryHero SE – Berlin, Germany

- Migrated business critical datasets from AWS/Snowflake style of Data Warehousing to GCP based data mesh architecture.
- Built and maintained business critical data pipelines used for C-level KPI reporting with a focus on reliability and data quality.

Data Engineer/Data Analyst – Freelance

05.2021 – 12.2023

Quandoo GmbH – Berlin, Germany

06-2021-09.2021

MegaWest Enterprise Inc. – Los Angeles, CA, USA

05.2021-12.2023

- (Quandoo) Built and deployed end to end ETL pipelines on GCP and AWS for ingestion and transformation of business-critical data.
- (MegaWest) Built end to end ingestion and analytics pipelines for global P&L assessment across Amazon, Shopify, Etsy, and eBay operations, enabling accurate financial insights and operational monitoring.

Business Intelligence Manager – Full Time

2016 – 2019

Nexify Inc. – Los Angeles, CA, USA

- Automated advertising and reporting workflows reducing manual effort from days/multiple employees to hours/single employee.
- Built data integration solutions for online retail metrics across Amazon, Shopify, Etsy, and eBay, enabling more reliable business insights.

Other Projects:

Vaultwarden - Raspberry Pi (<https://github.com/Alphan-Aksoyoglu/vaultwarden-rpi>)

A fully dockerized self-hosted, secure and private, VPN accessible, deployment of the Vaultwarden Password Manager.

Supermarket Simulation (https://github.com/Alphan-Aksoyoglu/supermarket_markov_simulation)

Modeled customer behavior for a fictional supermarket, using a Markov chain model, open-cv2 library, for visualization.

Tweeter ETL Pipeline (<https://github.com/Alphan-Aksoyoglu/tweeter-etl-pipeline>)

Fully dockerized ETL pipeline in Python, real-time Twitter streaming data and sentiment analysis stored in MongoDB/SQL.

Work experience (Academic):

Postdoctoral Research Assistant

2019 – 2020

Jacobs University Bremen – Bremen, Germany

- Led a research team completing four projects, performed mathematical modeling, data analysis, and project management.

Visiting Scientist

2016

National Institutes of Health – Bethesda, MD, USA

- Initiated and completed a follow-up project, performing modeling, data analysis, and experimental interpretation

Research Assistant

2011 – 2016

University of Massachusetts – Amherst, MA, USA

- Conducted experiments, computational modeling, and data analysis on focus topics (X-Ray scattering, electrophysiology).
- Taught introductory and advanced physics courses for scientists and engineers, physics labs and MATLAB.

Education

PhD in Physics University of Massachusetts – Amherst, MA, USA

2009-2016

Thesis: "Equilibrium Partitioning of Binary Polymer Mixtures into Biological Nanopores"

BSc in Physics Boğaziçi University – Istanbul, Turkey

2003-2008

Languages

English: Native Proficiency, **German:** Intermediate (B1 Certificate), **Turkish:** Native Language, **Romanian:** Beginner