

Profile

Senior Data Platform leader with 10+ years of experience building scalable, governed data systems across AWS and GCP. Specialised in Lakehouse architecture, DataOps, and production-grade ML infrastructure. Proven track record delivering cost optimisation, platform scalability, and cross-functional leadership at scale.

Work Experience

July 2024 – **Tech Lead – ML Platform & Data Engineering**, *Synthesia*, London, UK.

- Present
 - Lead Data Platform and DataOps strategy enabling multimodal ML training (audio + video) across AWS and GCP.
 - Own quarterly roadmap and technical direction for a 40+ researcher R&D organisation.
 - Delivered self-service ML training framework with full data lineage, provenance tracking, and multimodal observability.
 - Architected Lakehouse platform (Delta + Trino) enabling federated query across distributed data domains.
 - Designed migration from WEKA to S3-based architecture, unifying workloads and reducing infrastructure costs.
 - Reduced storage spend by \$5,000+ per day via lifecycle governance and non-current deletion enforcement.
 - Built ML training cost visibility framework providing leadership real-time infrastructure spend insights.
 - Standardised inference evaluation pipelines (A/B, MUSHRA) integrating human annotation workflows.
 - Driving research into semantic layering and RAG/TAG systems for multimodal data discovery.
 - Introducing MCP patterns for dataset and evaluation workflows to speed up processes.

May 2022 – **Lead Data Engineer**, *7bridges*, London, UK, **full time**.

- July 2024
 - Championed Lakehouse architecture for batch and streaming, leveraging open-source technologies (delta, mlflow, spark on k8s, kafka, trino, dbt, airbyte). [[Datanova 2023](#)] [[Starburst Success Story](#)] [[Big Data London Conference Speaker](#)]
 - Achieved 50% reduction in compute and storage costs.
 - Slashed query latency by up to 98%.
 - Spearheaded innovative learning initiatives and hackathons to drive continuous innovation.
 - Documented team member improvement, delivering projects 1-2 quarters ahead of schedule.
 - Introduced efficient DataOps workflows.
 - Empowered teams with self-serve data access, fostering data-driven decision-making.

Dec 2021 – **Lead Data Engineer**, *Kainos*, London, UK, **Full Time Consultant**.

- May 2022
 - Working on architecture and guidance to current data platforms on consultant basis.

Aug 2020 – **Data Engineer**, *IMG Arena*, London, **full time**.

- Dec 2021
 - Designed Lakehouse architecture based on Apache Druid, data processing pipelines both streaming and batch.
 - Managed and maintained standards for Data Analysis tools used across the organisation.
 - Engaging with sports data providers about source data quality.
 - Contributing to future design/architecture solutions that directly impact the business.

Dec 2019 – **Data Engineer**, *GameAnalytics*, London, **full time**.

- June 2020
 - Improved current data pipeline built in PySpark and Hive to insert data into BigQuery.
 - Built up the company data science platform. Including being responsible for best practices for data engineering and implementing data science features.
 - Introduced Apache Airflow from scratch. Responsible for maintaining, teaching others (workshops) in how our schedule and workflow platform works in the data team, and how it can be used elsewhere.
 - Worked closely together with data scientist to put statistical AB testing models into production. This included both data engineering and data science aspects.

June 2019 – **Data Engineer**, *YoYo Wallet*, London, **full time**.

- Dec 2019
 - Engineered large scale data modeling. Data processing in real time and batch with analytics.
 - Managed the transitioning of the data warehouse from Postgres to utilise Redshift.

Aug 2018 – **Data Engineer**, *Yieldify*, London, **full time**.

- June 2019
 - Maintaining maintenance of Redshift with Alembic.
 - Working closely with Data Analysts to deliver business needs both ad-hoc and modifying existing architecture.

August 2018 – **Developer Connected Vehicle Cloud**, *Ericsson AB*, Gothenburg, Sweden, **full time**.

- Jan 2018
 - Worked closely towards OEMs and OEMs delivery partners. Developing connectivity for the modern automotive ecosystem.
 - Engineered micro-service infrastructure in Ericcsons cloud environment based on Apcera (similar to k8s).

- June 2011 - **Support & Field Test Engineer / Software Engineer**, *Drivec AB*, Helsingborg, Sweden, **full time**.
Dec 2017
- o **2017**: Transitioned into a Software Engineer role, focusing on fleet management system analytics, GPRS/GPS and CAN hardware integration, and data preparation for machine learning solutions.
 - o **2014 - 2017**: Served as a Support & Same role as previously including hands on Android coding and other duties along side studies.
 - o **2011 - 2014**: Served as a Support & Field Test Engineer, analyzing CAN-bus data, ensuring project rollouts and installations, and maintaining customer relationships.

Leadership skills

- o Leading small teams on large projects through mentorship and guidance to solve complex business problems.
- o Setting data direction together with CTO and senior management through board meetings.
- o Aligning with other squads/teams teaching new technologies.
- o Taking on product management duties to deliver end to end data products.
- o Pushing team members and companies to learn new areas of work such as infrastructure as code, end to end architecture sessions, understanding the importance of having skills across the board to deliver successful data projects.

Conferences And Tech Talks

Articles and conference appereances:

- o Starburst & 7Bridges Article Of Lakehouse Architecture: [\[Starburst Success Story\]](#)
- o Big Data London Conference Speaker: [\[Big Data London\]](#)
- o Conference Speech Of Lakehouse Architecture: [\[DataNova 2023\]](#)

Datapains - Social Media

I post content under #datapains on socials

- o Medium Data Posts Podcast: [\[Articles\]](#)
- o Datapains Youtube Channel Podcast: [\[Videos\]](#)
- o Data Team Success Podcast: [\[Interview With Ross Webb\]](#)

Publications

- o Publications ITMS - MLA²: [\[Thesis - Diva Portal\]](#)[\[IOP Science Publication\]](#)

Technical skills

Not limited to but a selection of:

- o Main Programming Language: ~10 years: **Python**
- o Infrastructure As Code: ~6 years: **Terraform**
- o Job Scheduler Platform: ~4 years: **Apache Airflow**, ~7 years: **Native K8S**, ~3 years: **Argo Workflow**, ~2 years: **AWS Serverless**
- o BI Tools: ~2 years: **Apache Superset**, ~2 years: **Metabase**
- o Distributed Computing: ~1 year: **Scala Spark**, ~6 years: **PySpark**, ~1 years: **Hive**
- o Stream Processing: (~3 years: **Apache Kafka**)
- o CI: ~9 years: **Jenkins**, **Travis**, **CircleCI**, **GitLab**
- o CD: (>8 years: **Docker**, **Docker-Compose**), >5 years: **ArgoCD**)
- o Main DB Technologies: (Columnar: **Trino**, **Druid**, **Redshift**, **BigQuery**, **Athena**), (NoSQL: **DynamoDB**), (RDBMS: **Postgres**, **MySQL**)
- o Main Cloud Technology: (>8 years: **AWS**), (1 year: **Google Cloud**)
- o Certifications: [\[Coursera - Machine Learning\]](#)

References

Portfolio: [\[GitHub\]](#)

Honors & Awards: [\[Scholarship - Best Thesis\]](#) [\[Scholarship - Sparbanken Kronan\]](#)

Education

- 2014 - 2017 **BA of Science in Engineering, Computer Science & Engineering**, *Halmstad University*, Sweden.
- **Machine Learning thesis project where we delivered a cloud module called ITMS - MLA**². Which was honored as the best thesis in the categories of Computer Science, Electronic, Mechatronic Engineering.
 - Delivered a prototype for an ingredient check up app for allergy at the hackathon Hack4Health in April 2016.
 - Attended education and preparations for 6 months to be able to be responsible for the team building process for new students at the introduction week at the IT section fall 2015.
 - Representative during the open day in March 2017 for the IT section at Halmstad University.
- 2016 - 2017 **Exchange Student BA Computer Science & Engineering**, *Korea University*, KOR, 87.2/100 GPA.
- Attended courses in Computer networks, Information security, Software Engineering, Internet Programming together with international and Korean students.
 - Developer for the anonymous location based messages application www.klotter.co together with international students, A+ received.
- 2012 - 2013 **Computer Science**, *Malmö University*, Sweden.

¹CAN - Controller Area Network

²ITMS MLA - Indirect Tire Monitoring System - Machine Learning Approach