

ANTONIO DAVIDE CALÌ

Staff Data Engineer

London, UK | +44 7572 957528 | antoniodavidecali@gmail.com | LinkedIn | GitHub

SUMMARY

Staff Data Engineer with 10+ years building large-scale streaming and batch data platforms across fintech, retail, insurance, and consumer tech. Specialised in Apache Flink, Kafka, and Iceberg lakehouse architecture, with a recurring focus on self-service data infrastructure and GDPR/PII governance built directly into the platform. Speaker at Flink Forward 2025. Proven track record leading cross-functional initiatives, defining technical strategy, and turning complex distributed systems into accessible developer tooling.

SPEAKING & OPEN SOURCE

- **Speaker, Flink Forward 2025** - Presented on production patterns for self-service streaming data products at Apache Flink's flagship conference.
- **Creator, oat-latte** - Open-source Go TUI framework (oat-latte.dev) with a declarative component model inspired by Flutter.
- **Contributor** - Apache Airflow, OLake (datazip-inc/olake-helm), and other open-source data infrastructure projects.

CORE SKILLS

- **Stream Processing:** Apache Flink (JVM & Python), Apache Kafka, Apache Beam, Flink SQL, Kafka Connect
- **Languages:** Python, Scala, Go, Java, SQL
- **Data Platforms:** Apache Spark, Databricks, BigQuery, Snowflake, dbt, Airflow, Apache Iceberg
- **Cloud & Infrastructure:** AWS (ECS, S3, Glue, MSK, MSF), GCP, Kubernetes, Crossplane, Terraform, Docker, GitHub Actions
- **Practices:** Distributed systems, schema evolution, GDPR/PII handling, microservices, functional programming, IaC

EXPERIENCE

Zego - Staff Data Engineer

Apr 2026 - Present

London, UK

Own the Data Engineering platform end-to-end - Iceberg lakehouse infrastructure and Snowflake/dbt analytics - leading the introduction of Apache Flink and Protobuf-based PII/GDPR governance across the platform.

- Fully scoped and led the introduction of Apache Flink to Zego's data platform, driving the migration of streaming ingestion from a legacy Kinesis/Scala/Protobuf pipeline to a Flink-based forwarder.
- Leading GDPR/PII governance by extending the company's Protobuf schema contract with field-level PII annotations, driving automatic PII derivation and masking across the data lake and Snowflake's B2B/B2C domains.
- Own Iceberg lakehouse infrastructure (Terraform + AWS S3/IAM, Polaris catalog) and its Snowflake integration across staging and production environments.
- Evaluated OLake (Debezium-based database-to-Iceberg replication) and a Kubernetes CRD for Polaris catalog management as ingestion PoCs, contributing upstream fixes to the open-source olake-helm Helm chart.
- Improved Airflow-based ingestion tooling, standardising schema generation and DAG scaffolding for new data sources via Claude Code skills, shortening the path from schema definition to production pipeline.

- **Tech:** Python, Scala, SQL, Protobuf, Snowflake, dbt, Apache Airflow, Apache Iceberg, Apache Flink, Terraform, Kubernetes, Helm, ArgoCD, AWS, Kinesis.

Just Eat Takeaway.com - Staff Data Engineer

Nov 2024 - Apr 2026

London, UK

Lead the Streaming Ingestion team, building self-service tooling that enables product teams to stream and manage real-time data across a multi-cloud estate. Drive the “left-shift” initiative: a declarative DSL on top of Apache Flink that lets domain teams design, deploy, and own their own data products.

- Architected a self-service streaming platform that abstracts Flink, Kafka, and multi-cloud connectivity behind a unified DSL, reducing onboarding time for new data products from weeks to days.
- Lead schema management and data connectivity strategy across GCP and AWS, standardising contracts between producers and consumers and improving pipeline reliability.
- Selected as speaker at Flink Forward 2025 to share the team’s approach to productionising self-service streaming.
- **Tech:** Scala, Python, Java, Go, Apache Flink, Apache Kafka, Apache Beam, GCP, AWS, Kubernetes, Terraform, BigQuery, GitHub Actions.

Dojo - Senior Data Engineer

Mar 2024 - Nov 2024

London, UK

Member of the Data Streaming Platform team, expanding Apache Kafka capabilities and introducing Apache Flink and Apache Iceberg into the company’s data stack.

- Designed and shipped Kubernetes CRDs that gave product teams a self-service path to ingest streaming data, abstracting Kafka and Flink configuration behind declarative manifests.
- Improved platform scalability and developer experience for streaming ingestion, reducing manual operational toil for the platform team.
- **Tech:** Go, Java, Kubernetes, GCP, Apache Kafka, Apache Flink, Apache Iceberg.

Raft AI - Senior Data Engineer

Jun 2023 - Feb 2024

London, UK

Led infrastructure refactoring and the rollout of a new data platform architecture across the company.

- Designed and implemented the company’s new data platform on dbt, Airbyte, dlthub, Airflow, and BigQuery, replacing a fragmented ad-hoc stack.
- Introduced Apache Spark and Apache Kafka to the organisation, and uplifted Airflow patterns and back-end SQL across teams.
- **Tech:** Python, Apache Spark, dbt, Airflow, GCP, Terraform.

The LEGO Group - Senior Data Engineer

Apr 2022 - May 2023

Copenhagen, Denmark

Designed and built a secure data platform for storing, processing, and analysing PII data under GDPR, then led its migration to Databricks with fine-grained access control.

- Built end-to-end PII detection, retention, and analytics on a cloud-agnostic microservice baseline, ensuring GDPR compliance and legal sign-off.
- Led the migration of the platform to Databricks and implemented fine-grained access control via Unity Catalog.
- **Tech:** Python, Scala, AWS, Databricks, Unity Catalog, Terraform.

The LEGO Group - Data Engineer

May 2021 - Mar 2022

Billund, Denmark

Built the next-generation LEGO Data Platform from scratch, replacing legacy architecture with a microservice-based stack.

- Led the introduction and full implementation of Apache Airflow pipelines for the platform's orchestration layer.
- Re-wrote the Data API in Scala and built a new REST API in Flask, modernising data access for downstream consumers.
- **Tech:** Python, Scala, AWS, Apache Airflow, Terraform.

IKEA - Data Engineer

Jul 2019 - Apr 2021

Malmö, Sweden

Part of the Inspirational Feed team, productionising ML algorithms that personalised the IKEA.com homepage for every visitor.

- Led the design, productionisation, and A/B testing of the ML-powered inspirational feed serving the IKEA website.
- Built tooling and platform features that gave data scientists a reliable environment to develop and ship multiple algorithm variants.
- Scaled the team and automated the model deployment lifecycle.
- **Tech:** Python, Scala, Apache Spark, Apache Beam, Apache Airflow, GCP.

Teradata - Data Engineer (Consultant)

Jul 2017 - Jul 2019

Prague, Czech Republic

- Delivered streaming and batch data engineering projects for enterprise clients across a broad technology surface.
- **Tech:** Apache Kafka, Apache NiFi, Apache Spark, Scala, Groovy, Python.

Exprivia Telco & Media - Data Engineer

Feb 2015 - Jun 2017

Milan, Italy

- Developed Java SE applications and optimised Scala/Spark jobs, improving data processing throughput and reliability.

EDUCATION

MSc, Computer Engineering - University of Bologna, Italy

Sep 2013 - Sep 2015

BSc, Computer Engineering - University of Padua, Italy

Sep 2008 - Aug 2013