

IGOR MOSYAGIN

igor@mosyag.in • +46793325392 • [linkedin.com/in/shrimpsizemoose/](https://www.linkedin.com/in/shrimpsizemoose/) • Linköping, Sweden

SUMMARY

Senior Data Engineer with 10+ years of experience across data engineering, backend development, and machine learning systems. Built production data pipelines, backend services, streaming systems, and data platforms using Python, Go, Spark, Kafka, Kubernetes, and cloud infrastructure. Experienced in technical ownership, cross-team support, mentoring, and working across data, software, and infrastructure.

SKILLS

Programming: Python, SQL, Go; coursework experience with Scala and Kotlin

Data & Streaming: Airflow, Kafka, Spark, PySpark, Spark Structured Streaming, Apache Beam / Dataflow, dbt, Flink

Backend: gRPC, FastAPI, aiohttp, event-driven services, Django

Infrastructure & Cloud: Kubernetes, Docker, Terraform, AWS, GCP

Data Systems: BigQuery, Redshift, PostgreSQL, Redis, Solr, Aerospike, HBase

EXPERIENCE

Data Consultant — shrimpsizemoose AB

Jul 2023 — Present

Data engineering, analytics, and machine learning consulting

- Designed, reviewed, and implemented data pipelines and data platforms for enterprise clients
- Worked across data engineering, analytics, cloud infrastructure, and machine learning projects
- Delivered engineering workshops on tools including Airflow, dbt, Terraform, BigQuery, and modern DataOps practices
- Coached engineers on technical communication, public speaking, and knowledge sharing

Data Engineer — INGKA Group (IKEA) [contract]

Oct 2024 — Aug 2026

Marketing Activation Platform — customer data, segmentation, and audience activation

- Supported Apache Beam / Dataflow pipelines processing customer profile, consent, enrichment, and event data for audience segmentation and activation
- Integrated data flows with Adobe Experience Platform, including scheduled exports and source-system integrations
- Built and maintained Terraform-managed GCP infrastructure across development and production, including BigQuery, Pub/Sub, GCS, Cloud Scheduler, IAM, and observability
- Designed reusable, config-driven infrastructure patterns for onboarding multiple markets with consistent access control and data governance

Data Metrics — data products and engineering platform insights

- Built internal data products combining engineering data from GitHub, Jira, and internal developer platforms
- Developed ingestion and transformation workflows with Python, dbt, BigQuery, and Cloud Run
- Managed infrastructure and deployments with Terraform and CI/CD practices
- Prototyped and benchmarked new data technologies and worked with stakeholders on adoption and technical trade-offs

Developer Advocate — DoubleCloud

Jun 2024 — Oct 2024

Data engineering products and technical prototyping

- Acted as "Customer Zero", testing managed data services and providing product feedback from a data engineer's point of view
- Built technical prototypes and tutorials using company data products
- Supported engineering, product, and marketing teams with technical content, webinars, and conference talks

Developer Advocate — Bytewax

Sep 2023 — Jan 2024

Stream processing, technical prototyping, and developer tooling

- Built and evaluated streaming PoCs using Kafka, Redpanda, Pulsar, and Flink

- Refactored developer documentation for a major framework release and worked with the engineering team on developer experience
- Produced technical examples, partner materials, and conference content around stream processing systems

Senior Data Engineer — Klarna

Nov 2020 — Aug 2023

Data Insight and Analytics Platforms — core data infrastructure and internal data tooling

- Co-led the move towards a multi-cluster Redshift data platform used by internal analytics and engineering teams
- Built and improved internal data tooling and production workflows using Python, Airflow, Dagster, Terraform, AWS, and Amazon Redshift
- Supported and debugged Spark-based AWS Glue workloads as part of the internal data platform
- Improved DataOps practices through CI/CD, automated testing, monitoring, and production observability
- Prototyped and evaluated technologies for internal platform improvements
- Led the internal Python community, developed training, ran workshops, and supported engineers across teams

Machine Learning Engineer — Lamoda

Sep 2017 — Oct 2020

Personalization, search, and advertising data systems

- Built and maintained batch and streaming pipelines with PySpark and Spark Structured Streaming, including event processing at approximately 15K events/sec
- Developed production backend services in Go using gRPC and event-driven patterns for experimentation, state tracking, and service integration
- Built services supporting search suggestions and synonym management on top of Solr, handling brand misspellings, fashion terminology, and user search language
- Worked extensively with production Kafka flows, including producers, consumers, schemas, partitioning, operations, and migration of integrations from other systems
- Migrated ML and backend services from Nomad to an on-prem Kubernetes platform, working with Helm, manifests, custom operators, and CI/CD
- Developed and maintained personalization pipelines powering product recommendations and advertising analytics

Backend Developer — CHTD

Mar 2011 — Jan 2012

- Developed Python/Django backend functionality around an internal network database
- Integrated internal data with customer BI and reporting systems using PostgreSQL and SQL

EDUCATION

Linköping University, PhD in Theoretical Physics

2012 — 2017

- Teaching assistant for Quantum Mechanics course
- Lab assistant for Theory & Modeling Group

NUST MISIS, PhD in Condensed Matter Physics

2010 — 2015

- Administered a research HPC environment with 100+ nodes, including Linux operations, user support, monitoring, and resource allocation automation

Moscow Institute of Physics and Technology, MSc in Applied Math and Physics

2008 — 2010

Moscow Institute of Physics and Technology, BSc in Applied Math and Physics

2004 — 2008