Artem Boiko

Ottawa, Canada m@artemboiko.com (343) 202-7838 artemboiko.com artem1001

Summary

Staff Data Engineer / Staff Software Engineer with 12+ years across fintech, enterprise banking, retail, and music streaming. Hands-on builder of low-latency streaming, lakehouse/query-engine, and high-load backend systems on GCP/AWS — including sub-second Apache Flink SLOs and a 100TB+ data-quality platform for tier-1 banks. Combines deep coding and performance engineering (concurrency, JVM/GC tuning, profiling) with staff-level technical leadership across teams.

Skills

Languages: Python, Go, Java, Scala, SQL, C

Streaming & Processing: Apache Flink, Apache Kafka, Apache Spark, Apache Beam / Scio / Dataflow, Google Cloud Pub/Sub, CDC, Apache Hadoop, Apache Hive

Lakehouse & Analytics: Apache Iceberg, Trino, BigQuery, dbt, Airflow / Composer, Cube.js semantic layer, data quality, lineage / OpenLineage

Databases: PostgreSQL, SQL Server, MongoDB, Apache Cassandra

Cloud & Infrastructure: GCP, AWS, Kubernetes / GKE / EKS, KEDA, Terraform, Ansible, Linux, Prometheus, Grafana

Reliability & Performance: SLOs / SLIs, incident response, HA / DR, concurrency, JVM / GC profiling, high-load tuning, cost optimization / FinOps

ML & AI: BigQuery ML, anomaly detection (ARIMA), MLflow, ZenML, JAX

Experience

Alpaca.markets, Staff Data Engineer — Remote
API-first stock-brokerage infrastructure fintech — Mar 2026 – present — 5 months

- Own reliability, operability, and cost for the company-wide Trino/Iceberg/Airflow/dbt lakehouse; established SLIs, HA/recovery posture, and Prometheus/Grafana observability and alerting.
- Took ownership of a fragmented data stack; introduced maintenance windows and upgrade strategies, retired legacy applications, and defined the modernization path.
- Eliminated recurring Trino cluster crashes — from ~10 production incidents/month to zero — through JVM/GC tuning, query optimization, and workload-aware cluster sizing; reduced dbt job failures by 50%.
- Apply application-aware KEDA autoscaling and Kubernetes right-sizing for heterogeneous analytical workloads, reducing query latency and failed-query rate while holding spend flat.
- Drive the future-state lakehouse architecture and centralized governance model, defining how data access, ownership, and platform evolution scale across the company.
- Provision and scale Cube.js semantic-layer infrastructure for reporting consumers, connecting governed data access with business-facing metrics.

Badal.io, Staff Engineer — Toronto, Canada
GCP data & ML consultancy; tier-1 Canadian banks and national retail — June 2022 – Mar 2026 — 3 years 10 months

- Led enterprise GCP data-platform engagements for tier-1 Canadian banks and retail clients, setting architectures, reusable standards, delivery plans, and mentoring engineers.
- Led a team of senior and DevOps engineers delivering low-latency Apache Flink streaming analytics for a tier-1 investment bank; designed fault-tolerant market-event processing with checkpointing, DR, and JVM profiling to hit strict sub-second SLOs.
- Optimized Flink streaming jobs per workload class — some latency-optimized, others throughput-optimized — using Flink SQL and custom UDFs, redesigning stream topologies for reductions, with SQL Server tuning and indexing strategy across both paths.
- Designed and built AutoDQ for a tier-1 Canadian bank: a dbt/BigQuery/GCP data-quality system over 100TB+ datasets with anomaly detection, DQIM integration, scorecards, and alert deduplication that reduced noise by ~80%.
- Standardized dbt and Airflow/Composer delivery patterns, including SLIM CI, PR impact reports, reusable DAG framework, and optimizations that cut validation/build time by ~80%.
- Architected on-prem-to-GCP migration patterns, including Hadoop/Hive CDC, SQL Server replication, cross-region DR/fallback, low-downtime cutovers, and lineage across Spark, BigQuery, Qlik, and OpenLineage.

Nordstrom (contract via Luxoft / DXC Technology), Lead Software Engineer — Kyiv, Ukraine
Jan 2021 – May 2022 — 1 year 5 months

- Led delivery and operations for the Hadoop/Spark core of Nordstrom Analytical Platform, coordinating releases and production support for internal analytics teams.
- Led a mixed senior/junior engineering team; interviewed, onboarded, and mentored engineers while raising operational and documentation standards.
- Owned on-call and incident response for Hadoop/Spark platform issues, debugging cluster failures and restoring critical production workflows under pressure.
- Delivered platform features across Python, Go, Scala, Terraform, Kubernetes, AWS EMR, Hadoop, Spark, and Hive, coordinating rollouts with Nordstrom stakeholders.

Spotify (contract via Hyper Core AB), Senior Software Engineer / Data Engineer — Stockholm, Sweden
Jan 2019 – Dec 2020 — 2 years

- Built and operated extreme-scale GCP data systems for Spotify royalties reporting and customer messaging, owning production incidents for high-value analytical and product workflows.
- Designed complex reporting strategies with Scala/Scio/Apache Beam and BigQuery pre-aggregation/partitioning, processing terabyte-scale datasets for royalty and messaging analytics.
- Resolved Dataflow/JVM hot-node incidents caused by skewed data, using profiling, key redistribution, and pipeline-level tuning to stabilize critical jobs.
- Developed Python, Java, and functional Scala backend services and ingestion workflows with data-quality checks, stronger testing, code review, and on-call practices.

Ancestry.com (contract via Svitla Systems), Senior Software Engineer — Kyiv, Ukraine
Feb 2016 – June 2018 — 2 years 5 months

- Built Ancestry.com Shopping Cart backend and web features across Python/Flask, AngularJS, RabbitMQ, PostgreSQL, AWS, and Kubernetes, owning requirements through implementation.
- Migrated on-prem services to AWS EKS/Kubernetes, optimized high-load Python/asyncio latency, and implemented GDPR compliance changes.

Education

BS **International Scientific and Technical University named after Academician Yuriy Bugai**, Computer Science — Kyiv, Ukraine