Nikolay Eremeev

Senior Data Infrastructure Engineer

GitHub · LinkedIn · nikolayeremeev.com · German/EU Citizen

hello@nikolayeremeev.com

7+ years building data infrastructure and distributed platforms. Currently owning department-wide initiatives at CrowdStrike spanning real-time pipelines, cloud cost, CI/CD, IaC, and LLM-powered security automation across AWS/GCP. Core stack: Go, Java, C++, Flink, Kafka, Spark, Kubernetes, Terraform.

Experience

CrowdStrike — Engineer III, Data Infrastructure · *Jul 2025 – Present*

- Flink-based data replay service with pluggable Kafka/S3 connectors, SPI storage/transform modules, checkpoint resumability, token-bucket rate limiting — driven by critical incidents.
- Owned EMR cost optimization across production Spark clusters — Grafana dashboards, profiling, right-sizing with consuming teams — **37.5% cost reduction** on the flagship workload.
- Built department-wide CI/CD platform from scratch — Gradle monorepo; CD via Airflow (Spark big-data clusters), GitOps (Kubernetes), and Chef/Cinc (bare metal). Multi-level caching: 10 min PR / 30 min main builds. Drove migration from gitflow (hundreds of long-lived branches) to trunk-based via org-wide tech talks. Onboarding 6+ teams and 80+ services.
- Migrated cloud environments to Terragrunt-based IaC, reducing Terraform codebase **~25×** across AWS/GCP. Custom Terraform providers for Chef/Cinc and AWS.
- Built LLM-powered vulnerability patching integrated into CI/CD — automated scans generate auto-fix PRs via a Claude-based SDK on Jenkins; migrated container images to distroless.

IBM zAIOps — Advisory Software Developer · *Aug 2023 – Jun 2025*

- Led Data Processing team (3 engineers) building real-time event correlation in Java + Flink for IBM Z observability — event grouping, time-series aggregation, topology graph construction.
- Owned CI/CD workstream and led ChatOps team (3 engineers). Rescued ChatOps after IBM China office closure — restored build/release within weeks as tech lead across both workstreams.
- Hired 3 developers; mentored juniors → one promotion to mid-level.

IBM zAIOps — Software Developer · *Apr 2022 – Jul 2023*

- Led containerization of Tivoli Enterprise Portal Server (team of 2): multi-day install → ~5 min container deployment. Multi-arch builds (amd64, s390x) via Jenkins.

TeamViewer — Backend Engineer · *Jul 2020 – Mar 2022*

- Backend features in C++ and Go for enterprise platform serving 500M+ users: audit logs, client settings, RBAC.
- Led 32-bit → 64-bit user ID migration — zero-downtime rollout across interdependent services.

Nuance Communications — Software Engineer · *Apr 2019 – Jun 2020*

- C++/Java features for Dragon speech recognition in Android Auto and Linux in-car systems. Reduced build times ~50% via Gradle migration.

Open Source

- **Diffctx** — Smart git diff context for LLM code review via graph-based scoring (Rust + Python). CLI, Python API, MCP server. PyPI; preprint on Zenodo.
- **life-as-code.com** — Self-hosted health analytics (Garmin/Hevy/HealthKit). Python, Dash, PostgreSQL, k3s.

Skills

Data: Flink · Kafka · Spark (EMR, Dataproc, K8s) · PostgreSQL · Cassandra · MS SQL
Languages: Go · Java · C++ · Python
Infra: Kubernetes · Docker · Terraform · Helm · Jenkins · Gradle · Chef/Cinc · AWS · GCP · Grafana · OpenTelemetry
AI/LLM: Claude API/SDK integrations · MCP servers · LLM-powered CI/CD workflows
Human: English (Fluent) · German (Fluent) · Russian (Native)

KCNA — Kubernetes and Cloud Native Associate · B.Sc. Computer Engineering, RWTH Aachen (2014–2019)