

# Aleksandr Andreev

Tech Lead / Senior Data Engineer

✉ alealandreev@gmail.com    LinkedIn    @bigdatageek    Habr Blog

## Summary

---

Senior Data Engineer with 10 years of experience building BI-analytics and high-performance data platforms. Currently leading data infrastructure at AlfaStrakhovanie (major insurance company in Russia), managing 150+ production pipelines and building new ones from scratch. I specialize in Big Data, async Python services and functional programming (Scala).

## Technical Skills

---

<b>Languages:</b>	Python, Scala, SQL, VBA
<b>Big Data:</b>	Apache Spark, Databricks, Apache Kafka, Apache Flink, Apache NiFi, Apache Iceberg, HDFS, S3, Hive, YARN
<b>Databases:</b>	PostgreSQL, ClickHouse, Snowflake, Greenplum, ScyllaDB (Cassandra), Neo4j, NebulaGraph, Redis, DuckDB, Oracle, MS SQL Server
<b>Data Tools:</b>	Apache Airflow, dbt, Great Expectations, Soda, OpenMetadata, Feast, MLflow, Trino, Polars, Pandas
<b>Infrastructure:</b>	Git, Gitlab, Linux, Kubernetes (+OpenShift), Docker, Grafana, Prometheus, Victoria-Metrics, ArgoCD
<b>Frameworks:</b>	FastAPI, Locust
<b>BI &amp; Analytics:</b>	Qlik Sense, QlikView, Power BI, Tableau, OLAP cubes

## Professional Experience

---

### Senior Data Engineer

**AlfaStrakhovanie** (*Largest insurance company in Russia*)  
*Machine Learning Department*

**Jul 2025 – Present**

Moscow, Russia

- Developed and maintained FastAPI microservices integrated with Feast feature store for ML models. Covered all services with Grafana monitoring, achieving response times under 100ms.
- Designed a real-time feature store on Feast + FastAPI handling 2K+ requests/min with P99 latency <100ms, enabling instant underwriting decisions.
- Rewrote critical microservices from Python/FastAPI to Rust, reducing P99 latency to 12ms and cutting infrastructure costs by 60%.
- Architected a Medallion (Bronze/Silver/Gold) lakehouse with Apache Iceberg processing 1TB+ of data daily; reduced storage and processing costs by 15% through compression and Spark/Databricks pipeline optimizations.
- Maintained and created new pipelines (150+ production ETL pipelines total) on Spark/Databricks and Airflow with data quality frameworks (Great Expectations, Soda), achieving 97.5% SLA and reducing incidents by 50%.
- Implemented data governance on OpenMetadata, enabling data lineage for 300+ data sources.
- Configured Python code linters (ruff, isort, black). Actively used AI tools (Claude Code, Cursor) + agents.

**Stack:** HDFS, S3, ClickHouse, Snowflake, Apache Kafka, Apache Iceberg, Apache Spark, Databricks, Python, Polars, FastAPI, Airflow, OpenMetadata, dbt, PostgreSQL, Feast, MLflow, Neo4j

**Senior SRE Data Engineer**  
**Kryptonite** ("*Russian Palantir*")  
*Data Engineering Chapter*

**Feb 2025 – Jul 2025**  
Moscow, Russia

- Designed and built real-time pipelines on Apache Spark (Scala) and Apache NiFi + Spark Streaming, processing 200M+ events daily for an AI-powered analytics platform.
- Maintained data pipelines on Spark Scala, PySpark, and Apache NiFi. Orchestration: Airflow + ArgoCD. Code stored in GitLab; all services ran on OpenShift (Kubernetes) or Docker. Data stored in S3 and ScyllaDB.
- Designed a graph system on NebulaGraph linking 140M+ objects for semantic search.
- Developed type-safe Spark ETL pipelines in Scala with Cats Effect.
- Managed Kubernetes (OpenShift) infrastructure for the production data platform, operating 200+ pods with 99.95% uptime SLA and auto-scaling.
- Implemented a comprehensive load testing framework (Locust) for key services, achieving a 5× throughput improvement through ScyllaDB load optimization.
- Covered key ETL processes with baseline tests. Performed log analysis via Grafana + Prometheus and OpenShift pod logs, identifying and fixing data and pipeline errors.

**Stack:** OpenShift (Kubernetes), ScyllaDB, S3, NebulaGraph, Apache NiFi, Apache Spark, Databricks, Scala, Apache Kafka, GitLab + Gitea, Grafana, ZIO, Cats Effect

**Middle Data Engineer**  
**VK-IT (PJSC VimpelCom/VEON)**  
*Compliance and Risk Department*

**Nov 2023 – Feb 2025**  
Moscow, Russia

- Built an enterprise-scale compliance and anti-fraud data reporting platform on Apache Spark + Scala, processing 1TB+ of telecom transactions daily with real-time anomaly and suspicious activity detection. Full development cycle: from business analysis to Spark Scala code, GitLab deployment, fat JAR builds, and Airflow scheduling with VictoriaMetrics metadata sensors and YARN/Grafana performance analysis.
- Designed and executed migration from Oracle/MS SQL to Hadoop ecosystem (HDFS, Hive, Spark/Databricks). Converted legacy Oracle PL/SQL procedures to Scala/Spark, reducing report generation time from 8 hours to 1 hour.
- Created a self-serve analytics framework for generating compliance reports via an Excel interface, saving 40+ hours of engineering work per week.
- Designed real-time streaming pipelines on Apache Kafka processing 1M+ events/sec for fraud and extremism/terrorism detection.

**Stack:** Scala, Spark, Databricks, Kafka, Python, HDFS, Hive, Airflow, Oracle, MS SQL Server, PostgreSQL, Flink, GitLab, Git, Grafana, VictoriaMetrics, YARN

**Senior BI/DWH Developer**  
**Beeline (VEON group)** (*Major Russian telecom operator*)  
*Finance and Logistics Department*

**Mar 2018 – Oct 2023**  
Moscow, Russia

- Built a data warehouse on MS SQL Server serving 100+ colleagues across marketing, finance, and logistics departments. Developed full-cycle ETL processes (extraction, loading, transformation), created OLAP cubes, views, and stored procedures.
- Created a comprehensive BI platform on Qlik Sense with 50+ interactive dashboards; was among the first to introduce the Qlik BI system at the company. Trained 100+ business users on self-service analytics, significantly reducing ad-hoc query volume.
- Planned budgets for equipment logistics to retail locations, automated budget calculations, performed factor analysis, and reduced logistics costs by 15. %.

**Stack:** SQL, Qlik Sense, QlikView, MS SQL Server, MDX, DAX, OLAP, Power BI, Power Query, Power Pivot

## Business Analyst

Feb 2017 – Mar 2018

### LG Electronics

Moscow, Russia

*Finance and Sales Department*

- Built an automated sales, revenue, and inventory forecasting system using VBA macros and Excel tools (Power Pivot, Power Query), improving forecast accuracy by 25% across consumer electronics product lines. Developed VBA automation tools, reducing manual reporting time by 80%.
- Consolidated heterogeneous data from 5+ major retail partners, optimizing COI and COGS metrics.
- Created KPI presentations for LG Electronics RUS senior leadership.

Stack: VBA, Microsoft Excel, Power Pivot, Power Query, PowerPoint

## Intern

Mar 2016 – Apr 2016

### Ministry of Economic Development of the Russian Federation

Moscow, Russia

*Department of State Regulation*

- Worked with legal documentation and the Consultant Plus system; responded to inquiries from regional governments within the framework of Russian Federation legislation and allocated budgets.
- Prepared presentations for Deputy Prime Minister of the Russian Federation Arkady Dvorkovich.

## Education

---

**Moscow State University (MSU)** (*Top 1 university in Russia, QS World Rankings 2026 top 105*) **2026**

*Postgraduate Diploma in Machine Learning and Big Data Analytics*

Faculty of Mechanics and Mathematics

**National Research University Higher School of Economics (HSE)** (*Top 3 university in Russia, top 100-150 globally QS/THE Rankings*) **2016**

*Bachelor's Degree in Economics*

Specialization: Stock Market and Investments

## Languages

---

English (C1 – Advanced) · Russian (Native) · Spanish (A2 – Elementary)

## Interests & Publications

---

Technical blogging on [Habr](#) · Biohacking · Scientific literature · Industrial rock & metal, grunge, post-hardcore