Mark Andreev, Senior Software Engineer

London, UK   Email: mark.andreev@gmail.com

Phone: +44 7458 900595   GitHub: https://github.com/mrk-andreev

---

Education

Sep 2016 - Jun 2018. Master of Computer Science

Lomonosov Moscow State University, Master of Applied Mathematics and Informatics.

Thesis: "NLP in Macroeconomic Monitoring".

Sep 2012 - Jun 2016. Bachelor of Computer Science

Moscow Power Engineering Institute (National Research University). Mathematical modelling. Thesis: "Face recognition".

Experience

Senior Software Engineer, May 2024 - present

Man Group plc, London, UK

Engineer and enhance the AHL OMS (Market Platform) underpinning AHL's systematic investment process, with contributions spanning asset-class onboarding, performance/design optimisation, and third-line support for 24/5 mission-critical systems.

- Onboarded new asset classes and flows into the trading pipeline
- Optimised Oracle SQL*Plus queries in the user-facing flow through manual analysis of Autotrace | Most critical query - decreased latency by 60%.
- Optimised the user-facing flow through manual analysis of AsyncProfiler | Decreased latency by 70%
- Adopted Vaadin for internal developer-focused UI | Moved teams off local snippets and improved cross-team knowledge sharing

Tech: Java, Oracle Database, Kafka, Python, Linux

Senior Software Engineer, **Sep 2022 - May 2024**

Conundrum.AI, London, UK

Development of Machine Learning Platform for Industrial AI, improved the platform to scale the Data Science team taking into account updated requirements for load, response time, reliability and security. Performance optimizations resulted in improved prediction quality by reducing data latency. Compliance with information security requirements enabled transition of the project to the stage of industrial operation.

- Implemented low-level optimization for Python SDK that helps integration with the platform | Decreased RAM usage by 30%
- Implemented low-level optimization for feature store on top of Kafka & ClickHouse (projections, application-level query planner) | Reduced execution time of the most popular queries by 63%
- Covered 80% of queries to eliminate major performance degradations (Gatling)
- Created performance optimizations for Kafka subscription proxy (Java 21 virtual threads, shared subscription) | Decreased CPU load by 80%
- Implemented security improvements for Platform (audit, L4 network policies, L7 network filter) | Apply security IS requirements at network level

- Instrumented platform's services with health performance metrics (Prometheus, Grafana, Alerts) | Decrease issue investigation time by 64%

Tech: Java, Kafka, ClickHouse, PostgreSQL, Kubernetes, Hazelcast, Python, Linux

Senior Software Engineer, **Sep 2019 - Aug 2022**
Conundrum.AI, Moscow, Russia
Developed a Machine Learning Platform for Industrial AI
- Developed microservices with Spring Boot, PostgreSQL and gRPC
- Migrated feature store to Kafka & ClickHouse (columnar OLAP database) | Decreased query execution time by 90%
- Created low-level connectors for Industrial Data Exchange formats (MQTT, OPC UA, Historian) | Decreased CPU load to exchange server by 68%
- Migrated the model-serving runtime to Kubernetes (Kubernetes API, Helm)
- Deployed platform to AKS (Azure Cloud) & K3s (on-premises, no internet)
- Migrated to Keycloak (SSO) for Authz

Tech: Java, Kafka, ClickHouse, PostgreSQL, Kubernetes, Python, AWS, Azure, Linux

Mid-level Software Engineer, **Nov 2017 - Sep 2019**
Conundrum.AI, Moscow, Russia
Developed Machine learning & Data pipelines for Industrial AI
- Created feature store for sensors' time series data (Java, Spring, PostgreSQL, TimescaleDB)
- Created model serving runtime server (Python, Processes)
- Created incident management service (Java, Spring, Spring State Machine)
- Created ETL pipelines using S3, SQS, S3-SFTP

Tech: Java, PostgreSQL, Python, AWS, Linux

Junior Software Engineer, **May 2017 - Oct 2017**
Conundrum.AI, Moscow, Russia
Developed Machine Learning Solutions as consulting projects
- Airline data clustering; created an approach for data splitting for offline A/B tests
- Telecom data churn; offline churn scoring based on telecom data activity
- Web data gender detection; offline gender detection based on web activity
- Mobile data geo-analysis; created reports about geolocation activity based on mobile location data
- Industrial time-series data; created data pipeline for failure prediction

Tech: Python, Java, Linux

Software Engineer, **Oct 2016 - May 2017**
Big Data Indicators · Internship, Moscow, Russia
Developed end-to-end data-mining pipeline for economic indicators
- Created data collection & processing pipeline
- Used topic models to discover trends
- Created sentiment analysis models for trend prediction

Tech: Python, Java, MongoDB, RabbitMQ, Redis, Linux

Development stack

- **Java**. Spring: MVC, Data, AMQP, Kafka, Security, State Machine, Apache Camel, Vert.x, GraalVM
- PostgreSQL, Oracle Database, ClickHouse, Kafka, MongoDB, RabbitMQ, Redis, Keycloak, Prometheus, Docker, Kubernetes, Helm, Airflow, Spark, Cassandra, Hadoop, Linux, Terraform, Ansible, MLflow, Git
- **Cloud**. **AWS**: EC2, S3, RDS, CloudFront, SQS, SNS, Lambda, Batch, IAM; **Azure**: VM, Blob.
- **Python**. Flask, FastAPI, Pandas, Scikit-learn, XGBoost, LightGBM, PyTorch, CatBoost, Tornado.

Contributions to Open Source.

- Apache Spark. ValidateExternalType should return a child in error. SPARK-49044
- Apache Airflow. Add ipc_mode for DockerOperator. AIRFLOW-27553
- Apache Airflow. Add logging support for init containers in KubernetesPodOperator AIRFLOW-43853
- Apache Ignite. Implemented target encoding preprocessor. IGNITE-13713
- Apache Ignite. Implemented Yandex Catboost inference integration. IGNITE-13714
- Apache Ignite. Implemented new distances (BrayCurtis, Canberra, JensenShannon etc). IGNITE-13386
- Apache Camel. Fix Azure Blob Storage and Azure Blob Queue interaction. CAMEL-16092
- Keycloak. Fix null username in LDAP. KEYCLOAK-19743
- Keycloak. Fix NPE if the user does not exist in PolicyEvaluationRequest. KEYLOCK-16239
- Vaadin Flow. fix: add nodeVersion in Gradle plugin settings. Vaadin-20059
- Tornado Swagger. Swagger API Documentation builder for tornado server.

Conferences / Public Speaking

Dec, 2022. Talk "Kafka 101**"** in MLSystemdesign_2022_ods Meetup.

Feb, 2019. Talk "ML in production" in FunTech Meetup

May 2018. Volunteer researcher at EnduringNet (founded by Ser-Huang Poon, Professor at The University of Manchester)

July 2017. Big Data approach to measure inflation expectations: the case of the Russian economy (IFABS 2017 Oxford Conference), Goloshchapova, I., & Andreev M.

May 2017. Measuring inflation expectations of the Russian population with the help of machine learning (Voprosy Ekonomiki), Goloshchapova, I., & Andreev M.