

Roman Grebennikov

[Software, ML] Engineer

email: grv@dfdx.me, **skype:** rgrebennikov, **github:** shuttie, **stackoverflow:** shutty

location: Berlin, Germany

About:

I'm a ML big data full stack devops buzzword beep-beep professional, with a track record of designing and implementing large data processing ML-heavy applications.

Mainly interested in functional programming (Scala), machine learning problems (learn-to-rank, recommendations, natural language processing) and modern infrastructure (AWS/GCP, Kubernetes).

Experience:

- **ML Engineer**, Findify/Namaste Technologies March 2016 till now
Leading a small team building smart search and recommendations for 2K+ merchants:
 - Recommendation engine with Flink and Cassandra.
 - Real-time search personalization back-end with Scala, Akka, Kafka, Cassandra, ElasticSearch and a bit of machine learning: LambdaMART, ALS, SVD.
 - Search back-end with Scala, Akka-stream & Lucene.
 - A bit of devops with Kubernetes on AWS. CI/CD with Jenkins, ChatOps, monitoring with Grafana/Prometheus, logging with Kibana/Logstash/Loki.
 - Near-realtime analytics with Clickhouse and Kafka.
- **Lead developer**, DSR Sep 2014 – Feb 2016
Designing and implementing data-intensive applications:
 - 100Tb+ analytics warehouse solution with live data replication from MySQL fleet to Cassandra using AWS, Scala, Kafka, Spark-sql and Tableau.
 - Social network data aggregation service for financial scoring: AWS, Apache Storm, Cassandra, Akka and Spray.
- **CTO**, Sociohub May 2012 – Sep 2015
Building a major part of the analytics pipeline from the ground up, creating and leading a team for:
 - Distributed large-scale web scraping (C++, Scala+akka, ~1.5b requests/month/server),
 - Data processing: creating automated data extraction system based on Spark+Cassandra and Hadoop+HBase,
 - Solving partial duplicate detection tasks in large databases,
 - Creating advanced client credit scoring models for banking and finance (Weka, R, Rapidminer),
 - Building fault-tolerant high-performance client API (Scala+spray, Cassandra, Elasticsearch)
- **Software engineer**, Numbers Jan 2010 – May 2012
Worked in a small team of professionals to build custom data mining solutions for banking and finance:
 - Data mining, OLAP: creating client segmentation schemes, client scoring.
 - Business research and modeling: building recommender systems.

- Social networks: advanced statistical research, building automated web crawlers/spiders, deploying large NoSQL databases.
- **Quant**, Lord Finance May 2008 – Jan 2010
While working on a Ph.D. thesis at night, I've designed, evaluated and implemented a number of algorithmic trading models for a small local quantitative hedge fund:
 - Algorithmic trading research in option pricing. Creating and testing financial models in Mathematica & C++. Neural networks, parameter optimization.
 - High-frequency trading robot research with custom distributed market simulator.
 - Design and prototyping of trading robots.

Education:

- Ph.D. in CS, Voronezh State University, 2007-2011
Theoretical Informatics, Crowd behavior modeling using local scalar fields.
- Specialist in CS, Voronezh State University, 2002-2007

Skills:

- Scala (advanced): akka (cluster, streams, http), cats/cats-effect, sbt, scalatest, scalacheck, http4s
- Devops (advanced): AWS (EC2, R53, RDS, Redshift, EKS, ECS), Linux, Docker, Terraform, Kubernetes, Ansible.
- Databases (ok): Cassandra, Clickhouse, Redshift, Postgres, Aerospike, RocksDB.
- Data processing (ok): Spark, Flink, Hadoop, Hbase, Kafka, Kinesis, ElasticSearch.
- Java (ok): Lucene, performance, jmh.
- Machine learning (ok): xgboost, weka, a bit of Python with numpy/scipy.
- C++ (was ok, not used for years): boost, Qt, OpenMP.

Languages:

- English: advanced.
- German: basic, can order a beer.
- Russian: native.

Activities:

- Conference speaking (scala.io, BeeScala, FpConf, Highload, Jetconf, Scalar, JPoint, MICES, FlinkForward)
- [Snow, Kite, Wake]boarding.