



# Oleg B.

Senior Big Data / Machine Learning Engineer



---

## Professional summary

Over 15 years' experience in leading design, developing, and delivery of complex IT projects and high-performance solutions, 12+ years in business intelligence and in the data analytics field. Advanced hands-on experience in reactive, microservices-based, distributed system design and development including stream application platforms for advanced analytics including machine learning and data science. Strong foundation in statistical inference and probability; ability to clearly write, interpret, and communicate complex statistical models to a variety of audiences. Deep practical experience with creating ETL processes to source and link data. Data science researcher focused on the immediate benefits for the business using the Big Data tools with advanced analytical and visualization APIs, OLAP engine for Big Data. Real-world practical use cases with building next generation Streaming Analytics Application. Skilled in cloud technologies, blockchain technology stack including distributed ledgers, applications (DApp) and smart contracts (Golang) based on Hyper edger foundation and Composer.

## Education

State Polytechnic University. Qualification awarded Bachelor's Degree in Applied Physics.

## Industries

Transportation & Logistics, Technology, FinTech, Sales & Marketing

## Technologies and tools

- Python, C#, Scala, Visual Basic .NET, R
- Apache Spark, TensorFlow, Keras
- Apache Cassandra, Apache Hive, HBase, MS SQL
- Apache Kafka
- AWS, Microsoft Azure
- Hadoop, Sqoop, Teradata Aster, Airflow

# PROJECTS

---

## Project for SPINS

Duration: 10.2019 – till now

SPINS is a wellness-focused data technology company and trusted advocate for the natural and specialty products industry.

### Team

20 specialists

### Position

Senior Big Data / Machine Learning Engineer

### Responsibilities

- Providing high quality, professional services to help organization establishing a data-driven company that treats data as a strategic asset (Genomes as shared, enterprise-wide, reusable information assets)
- Delivering innovation projects in variety of business areas including Enhance Capabilities, Quality Management, DevOps implementation , Innovation model and Integrated portfolio based on ground-breaking Big Data, Machine Learning and AI technologies and frameworks

### Technologies and tools

- Scala, Python, Serverless Framework, Lambda, S3, Redshift, AWS Glue, AWS Batch, AWS Step functions

# Emirates Group

Duration: 03.2018 – 10.2019

The Emirates Group is a Dubai-based international aviation holding company headquartered in Garhoud, Dubai, United Arab Emirates, near Dubai International Airport. The Emirates Group comprises Dnata, an aviation services company providing ground handling services at 78 airports, and Emirates Airline, the largest airline in the Middle East.

## Position

Senior Big Data / Machine Learning Engineer

## Responsibilities

- Provided high quality, professional services to help organization establishing a data-driven company that treats data as a strategic asset (Genomes as shared, enterprise-wide, reusable information assets)
- Delivered innovation projects in variety of business areas including Enhance Capabilities, Quality Management, DevOps implementation, Innovation model and Integrated portfolio based on ground-breaking Big Data, Machine Learning and AI technologies and frameworks

## Technologies and tools

- Apache Spark, Dataiku, Azure (Data Lake, Kubernetes, Blob, HDInsight)

The Scotiabank Digital Factory is a production facility designed to drive collaboration and creativity as well as improve customer experience in areas such as mobile and digital banking transactions, account openings, loan adjudication and mortgage lending.

## Position

Senior Data Engineer

## Responsibilities

- Integration Big Data technology stack and machine learning models via microservices architecture
- Fast Data: Components which process data in-flight (streams) to identify actionable events and then determine next-best-action based on decision context and event profile data and persist in a durable storage system
- Reservoir: Economical, scale-out storage and parallel processing for data which does not have stringent requirements for data formalization or modelling
- Factory: Management and orchestration of data into and between the Data Reservoir and Enterprise Information Store as well as the rapid provisioning of data into the Discovery Lab for agile discovery
- Warehouse: Large scale formalized and modelled business critical data store, typically manifested by a Data Warehouse or Data Marts
- Data Lab: A set of data stores, processing engines, and analysis tools separate from the data management activities to facilitate the discovery of new knowledge. Key requirements include rapid data provisioning and subsetting, data security / governance, and rapid statistical processing for large data sets
- Business Analytics: A range of end user and analytic tools for business Intelligence, faceted navigation, and data mining analytic tools including dashboards, reports, and mobile access for timely and accurate reporting



Information Management Architecture Strategy (IMAS) in Nationwide Building Society.

## Position

Big Data Lead

## Responsibilities

- Identified process friction points for mortgages, where a member gets stuck or cycles back, in journey and identified improvements
- For unsecured loans, identifying which members in all segments that are likely to close their account early across any channel to enable Nationwide to proactively manage the relationship
- Identified real-time opportunities to improve member omni-channel experience by aligning web-initiated contact with relevant human interaction at branch or call centre
- Classifying direct debits and standing orders from current account transactions to identify where a member has a product elsewhere
- Full-scale machine learning techniques across multiple environments – Path Analysis (nPath), Attribution Modelling, Naïve Bayes (to analyze behavioral differences), Cluster Analysis (to identify key investor types, segmentation), Text Analytics (n-gram) for key trigger phrases from text, Graph analytics for analysis of process steps actually taken in member web journeys, Time series analysis for periodicity detection

## Technologies and tools

- Scala, Python, R, Apache Spark, TensorFlow, Keras, Apache Solr, Apache Cassandra, Apache Hive, Apache Zeppelin, Apache Mesos, Apache Kafka, Google Cloud, AWS, Jupyter, Teradata Aster, HDP

Canadian Tire Corporation, Limited is a Canadian retail company which sells a wide range of automotive, hardware, sports and leisure, and home products. Some stores also sell toys and food products.

## Position

Lead Data Scientist

## Responsibilities

- Developed and implemented multi-layer threat / linked data analysis platform hosted in Big Data environment (HDP) to proactively uncover hidden threats through cyber hunting
- Designed and modelled Security Data Lake as a common data repository for wide range of security tasks (behavioral monitoring, network anomaly detection, user scoring, correlation engines and so forth)
- Identification and importance of behavioral analysis features for network / users Anomaly Detection (SIEM, CarbonBlack, FireEye, AVT, firewalls etc.)
- Data cleaning and enriched representations for Anomaly Detection in system calls (R, Scala)
- Fast outlier detection for distributed high-dimensional data sets with mixed attributes, empirical evaluation using real enterprise scale datasets (R, Python)
- Leveraged once class SVM for detecting anomalous Windows registry and file system accesses (R, Scala)
- Implementation of combined approach for anomaly detection using neural networks (SOM) and unsupervised clustering techniques (R, Scala, Python)
- Feasibility study of using graph-based clustering for anomaly detection in IP networks and real-time alert correlation with type graphs
- Discovered novel attack strategies from INFOSEC Alerts using probabilistic (HMM) / Statistical based alert correlation to support root cause analysis
- Developed a tool for content anomaly detector resistant to mimicry attack based on Markov n-Grams and POC for early detection of cyber security threat (APT) using structural behaviour modeling (R, Scala)

## Responsibilities

- Implementation of advanced visualization techniques for exploration analysis of massive data sets to find insights in cyber security data (parallel coordinates, TreeMap, TimeFlow, time-based visualizer)
- Developed the hybrid malicious code detection method based on Deep Learning and the application of Deep learning on traffic identification (R, SparkR)

## Technologies and tools

- Scala, Python, R, Apache Spark, Apache Cassandra, Apache Hive, Apache Zeppelin, Jupyter, Apache NiFi, Apache Kafka, ELK, Google Cloud, AWS, RabbitMQ, HDP

Paytm is India's largest mobile commerce platform that does over 30 million transactions a month, a majority of them occur on mobile apps. Paytm started by offering mobile recharge and utility bill payments and today it offers a marketplace to consumers on its mobile apps.

## Position

Senior Big Data Engineer / Data Scientist

## Responsibilities

- Played a lead role in determination of overall solution architectures and designs consistent with architecture to support strategic Big Data initiatives across domains
- Was responsible for assemble, manufacture and test variety of features (feature engineering), model selection and performance evaluation for real-time fraud detection and recommendation systems including decision trees, parametric models (logistic regression), nonparametric approaches (SOM, K-NN, SVM), ensemble methods (GBM, random forest)
- Led the creation of observed rules jointly with business SME to improve detection and reduce false positives (achieved accuracy 86%), author the idea of "utility" score which enables transaction prioritization for investigation according to their importance to the business
- Productize predictive and prescriptive analytics model with Data Engineers team
- Being directly involved into design of the reporting system that drives the action – the data output can be sliced and diced, and reported via graphic dashboards to allow managers to see priorities for investigating transactions and spot trends and anomalies

## Technologies and tools

- Scala, Python, R, Apache Spark, Apache Cassandra, Apache Hive, Apache Zeppelin, Apache Sqoop, Apache Kafka, AWS, RabbitMQ, Jupyter, HDP

Kinross Gold Corporation is a Canadian-based gold and silver mining company founded in 1993 and headquartered in Toronto, Ontario, Canada. Kinross currently operates eight active gold mines and was ranked fourth of the "10 Top Gold-mining Companies" of 2017 by InvestNews. The company's mines are located in Brazil, Ghana, Mauritania, Russia and the United States. It trades under the KGC ticker in the New York Stock Exchange and under K in the Toronto Stock Exchange.

## Position

Senior BI and Data Science Developer / Project Manager

## Responsibilities

- Led the project teams including establishing project plans and milestones, analyzing risks, developing budgets, and delegating work assignments; accountable for results
- World-wide implementation of MicroStrategy 9.3/4 and MicroStrategy Distribution services, OLAP Cubes and MicroStrategy mobile across company sites in North and South Americas and Russia
- Expertise in Installing, Configuring of all MicroStrategy activities including MicroStrategy Desktop, Administrator, Intelligence Servers, Web Servers and mapping to Client machines
- Strong Knowledge of Data Extraction, Data Integration, and Data Mining for Decision Support System using ETL and OLAP tools
- Intensive experience and exposure to all aspects of BI and data mining applications such as Administration, Architecting and Development
- Strong understanding of Data warehouse concepts, dimensional modeling using various Schemas and Multi-Dimensional Model with respect to query and analysis requirements

## Technologies and tools

- C#, Visual Basic .NET, Apache Hive, Apache Sqoop, Hadoop, HBase, MS SQL, SSIS, SSRS, SharePoint, MicroStrategy 8-9