

# Oleg B.

## Architect/Team-lead ML Engineer/Big Data Architect

### SUMMARY

- Over 15 years experience in leading the design, developing, and delivery of complex IT projects and high-performance solutions, +10 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
- Proficient Data Engineer-researcher focused on the immediate benefits for the business using Big Data tools (AWS Glue, AWS Greengrass, AWS EMR, AWS Data Lake) with advanced analytical and visualization APIs (graph DB - Titan, Neo4J, Tinkerpop, software development - Scala, Python) with CI/CD pipelines - Jenkins, Circle CI, GitLab actions
- Generative AI - Q&A with multiple choices, pre-trained models (Hugging Faces ecosystem, T5, BERT, GPT), ChatBot for online gambling platform (LangChain, Pinecone, Cohere, Faiss, Hugging Face Hub)
- Generative AI in NLP - information retrieval for 1) generate personalized recommendations for products or services based on a user's preferences and past behavior 2) summarize legal documents and contracts, making it easier for lawyers and legal professionals to review and analyze large volumes of legal documents. 3) create content such as product descriptions, blog posts, and social media posts
- Recommendations platforms - mobile games platform (generate game recommendations based on player history, promo-offers, AWS Personalize ), self-learning algorithms for data-based risk management in agriculture (Monte-Carlo tree and Markov chains)
- Upper-intermediate English.
- Availability starting from ASAP

### TECHNICAL SKILLS

<b>Main Technical Skills</b>	ML
<b>Programming Languages</b>	Python, R, Scala
<b>Scala Frameworks</b>	Akka, Apache Spark
<b>Scala Libraries and Tools</b>	Akka
<b>Java Frameworks</b>	Apache Spark
<b>AI &amp; Machine Learning</b>	AWS SageMaker (Amazon SageMaker), Keras, Kubeflow, Mlflow, PyTorch, TensorFlow



<b>.NET Platform</b>	Azure
<b>Python Libraries and Tools</b>	BentoML, Dask, Keras, Matplotlib, Metaflow, Pandas, PyTorch, Seaborn, TensorFlow
<b>Python Frameworks</b>	Django
<b>Data Analysis and Visualization Technologies</b>	Apache Airflow, Apache Hive, Apache Spark, HBase, Jupyter Notebook, ML, Pandas, Power BI, Sqoop
<b>Databases &amp; Management Systems / ORM</b>	Apache Hadoop, Apache Hive, Apache Kylin, Apache Spark, AWS ElasticSearch, AWS Redshift, Cassandra, ELK stack (Elasticsearch, Logstash, Kibana), Microsoft SQL Server, MongoDB, MySQL, Neo4j, Oracle Database, PostgreSQL, Redis, Snowflake, SQL
<b>Cloud Platforms, Services &amp; Computing</b>	AWS, Azure, Azure ML, GCP
<b>Amazon Web Services</b>	AWS EC2, AWS ElasticSearch, AWS Glue, AWS Kinesis, AWS Lambda, AWS RDS (Amazon Relational Database Service), AWS Redshift, AWS S3, AWS SageMaker (Amazon SageMaker), AWS SAM, AWS VPC
<b>Deployment, CI/CD &amp; Administration</b>	Ansible, CI/CD, Helm
<b>Web/App Servers, Middleware</b>	Apache HTTP Server
<b>Platforms</b>	Apache Mesos
<b>SDK / API and Integrations</b>	API
<b>Mail / Network Protocols / Data transfer</b>	Consul
<b>Operating Systems</b>	Debian, Linux, Ubuntu, Windows
<b>Virtualization, Containers and Orchestration</b>	Docker, Kubernetes, Terraform
<b>Version Control</b>	Git
<b>Collaboration, Task &amp; Issue Tracking</b>	Jira, Redmine
<b>Message/Queue/ Task Brokers</b>	Kafka
<b>Other Technical Skills</b>	Hashicorp, Pachyderm, Raspberry



## EXPERIENCE

### Data Engineer

September 2021 - now

### ML Engineer

April 2020 - now

### ML Engineer

March 2018 - April 2020

### Data Engineer, Scotiabank Digital Factory

September 2017 - March 2018

### Big Data Architect, ACCENTURE UKI

June 2016 - August 2017

### Data Scientist, Canadian Tire Corporation

August 2015 - June 2016

### Big Data Engineer, RAYTM LABS

December 2014 - July 2015

### Data Science Developer, KINROSS

July 2008 - February 2015

## PROJECTS

### Big Data Developer / Data Engineer

Nov 2022 - now

**Description:** Data Engineer for Palantir Foundry

**Responsibilities:** Design and development ETL pipelines (process flows) and models based on Palantir Ontology and Apache Spark to handle structure batch data. Using Palantir API to connect different data sources to the corporate data lake

**Technologies:** Palantir Foundry (PySpark)



## Data Engineer

Jun 2022 – Nov 2022

**Description:** A platform for online tests, Q&A

**Responsibilities:** Design and development AWS-based ETL pipelines to handle structured and unstructured data for online assessment and educational platforms. Setup and configuration of AWS data like using Databricks platform and Apache Spark Platform performance analysis and troubleshooting. POC for Online assistant leveraging NLP algorithms (PyTorch, Transformers). Ingestion pipelines (based on AWS Glue + Data Catalog), developing Redshift Db data models (dist keys and sort keys)

**Project link:** <https://www.inspera.com/>

**Technologies:** AWS S3, Lambda, Glue, Redshift, Step functions, Databricks Live tables (Delta lake), Elasticsearch, powerBI

## Data Engineer

Jan 2022 – Jun 2022

**Description:** One of the biggest airline companies in the world

**Responsibilities:** Providing high-quality, professional services to help organizations establish a data-driven company that treats data as a strategic asset. Delivering innovation projects in a 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. Active participation in the creation of Big Data CoE as the One-Stop Shop. Developing ML models for credit risk management. Implementation data science platform to predict the contingency fuel required for a given flight considering the influencing factors. Advanced Exploration Analysis applied to short-term and long-term planning. Automatic forecast analysis

**Technologies:** AWS, Google Cloud, Apache Spark, Apache Mesos, Cassandra, Kafka, Hive, Zeppelin, Jupyter, Scala, Python, TensorFlow, PowerBI

## Big Data Engineer

Sep 2021 – Jan 2022

**Description:** Bank which offers personal and commercial banking, wealth management and private banking, corporate and investment banking, and capital markets, through its global team

**Responsibilities:**

- Integration Big Data technology stack and machine learning models via microservices architecture

**Technologies:** Google Cloud, Apache Spark, Apache Mesos, Cassandra, Kafka, HDP 2.3, Teradata Aster, Zeppelin, Jupyter, Scala, Python, Keras, R, PowerBI

## Data Engineer

Sep 2017 – Mar 2018

**Description:** Information Management Architecture Strategy (IMAS) in Nationwide Building Society

**Responsibilities:**

- Implementation of Discovery Analytics / Data Science stream including Full-scale machine learning techniques across multiple environments - Path Analysis (nPath), Attribution Modelling, Naïve Bayes (analyze behavioral differences), Cluster Analysis (to identify key investor types, segmentation),



Text Analytics (n-gram) for key trigger phrases from the text, Graph analytics for analysis of process steps actually taken in member web journeys, Time series analysis for periodicity detection. Ingestion pipelines (based on AWS GLue + Data Catalog), developing Redshift Db data models( dist keys and sort keys)

**Technologies:** AWS (Glue, Redshift, etc), Google Cloud, Apache Spark, Apache Mesos, Cassandra, Kafka, Hive, HDP 2.3, Teradata Aster, Apache Solr, Zeppelin, Jupyter, Scala, Python, TensorFlow, Keras, R

### **Data Engineer**

Jun 2016 - Aug 2017

**Description:** A corporation, one of the leaders in the retail industry in Canada, owning a network of stores in all provinces and territories of the country

#### **Responsibilities:**

- Developing and implementing of multi-layer threat / linked data analysis platform hosted in a Big Data environment. Design and modeling Security Data Lake (HDFS, Avro, Parquet, HBase, Cassandra) Identification and importance analysis of behavioral 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). Implementation of a combined approach for anomaly detection using neural networks (SOM) and unsupervised clustering techniques (R, Scala, Python). Developing the hybrid malicious code detection method based on Deep Learning and the application of Deep learning on traffic identification (R, SparkR). Ingestion pipelines (based on AWS GLue + Data Catalog), developing Redshift Db data models( dist keys and sort keys)

**Technologies:** AWS (Glue, Redshift, etc), Google Cloud, Apache Spark, Apache NiFi, RabbitMQ, Cassandra, Kafka, Hive, HDP, ELK, Zeppelin, Jupyter, Scala, Python, R

### **Senior Big Data Engineer / Data Scientist**

Aug 2015 - Jun 2016

**Description:** Fastest growing Indian e-commerce

#### **Responsibilities:**

- Integration of Big Data technology stack

**Technologies:** AWS, Apache Spark, Apache Sqoop, RabbitMQ, Cassandra, Kafka, Hive, HDP, Zeppelin, Jupyter, Scala, Python, R

### **Data Engineer**

Dec 2014 - Jul 2015

**Description:** One of the world's leading gold mining companies

#### **Responsibilities:**

- Leads project teams. Worldwide 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 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 Systems 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 Models with respect to query and analysis requirements

**Technologies:** Apache Hadoop, Sqoop, Hive, HBase, Visual.Net, MS SQL, SSIS, SSRS, SharePoint, C++, Java, C#, MicroStrategy 8-9

