

Oleg



 Dubai, UAE

Sex M

Nationality Canadian

JOB APPLIED FOR
POSITION
PREFERRED JOB

AI ARCHITECT / PRINCIPAL BIG DATA ENGINEER/ GENERATIVE AI AND ML

If you torture the data long enough, it will confess ©

WORK EXPERIENCE

- Over 15 years' experience in leading design, developing, and delivery of complex IT projects and high-performance solutions, +10 year in business intelligence and in the data analytics field
- Generative AI effective adoption on enterprise level (multimodal) - implementation of machine customers ('custobots') that can autonomously negotiate and purchase goods and services including personalization and AI-assisted diagnostics (GenAI Stack - LLM, AI Mesh, Agents, RAG).
- Generative AI in NLP - information retrieval for 1) generate personalized recommendations for products or services based on a user's preferences and past behaviour 2) summarize legal documents and contracts, making it easier for lawyers and legal professionals to review and analyse large volumes of legal documents. 3) create content such as product descriptions, blog posts, and social media posts
- Generative AI - Q&A with multiple choices, pre-trained models (Hugging Faces ecosystem ,T5, BERT, GPT), ChatBot for on-line gambling platform (LangChain , Llamaindex, Pinecone, Cohere, Faiss, Hugging Face Hub)
- 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 the Big Data tools (Azure, AWS) with advanced analytical and visualization APIs (graph DB – Titan, Neo4J, Tinkerpop, software development – Scala, Python) with CI/CD pipelines – Jenkins, Circle CI, GitLab actions
- Recommendations platforms - mobile games platform (generate game recommendations based on player history, promo-offers, AWS Personalize), self-learning algorithms for a data-based risk management in agriculture (Monte-Carlo tree and Markov chains)
- Areas of expertise include outlier/anomaly detection (credit card/e-commerce fraud detection systems, interesting sensor events) , recommendation engines for customer segmentation using different type of basic and ensemble algorithms (proximity-based methods, high-dimension grids, multidimensional streaming, change-based outliers in temporal graphs (SNN), behavioural distance based), deep learning (PyTorch, PyTorch Geometric, PyG)
- Deployed to production cloud based Situational Awareness System - proved practical experience in IoT (AWS Greengrass) and Computer Vision (Object Detection , Image recognition and segmentation) integration
- Smart, intellectual platform for Property Management and Real Estate Marketplaces analysis for the EU based real estate agencies - well-defined combination of NLP (Transformers, BERT) and Computer Vision technologies to search and improve property value estimates, risk assessment (text-to-image , image tagging, photo compliance, watermark detection)
- Design and implement statistical / predictive models and cutting edge algorithms utilizing diverse sources of data to predict demand, risk and equipment reliability. Deep practical experience with creating ETL processes to source and link data.
- Hands-on experience on cloud technologies - AWS (EC2, Lambda, Kinesis, SageMaker) , Google Cloud - SDK, Deployment Manager, Source Repositoriess
- Hands-on experience with TOGAF as an enterprise architecture benchmark that ensures consistent standards, methodologies, and interaction among enterprise architecture specialists (Accenture UK, NBS bank), using Zachman Framework to create matrix where each cell is concentrating on one dimension or perspective of the organization.

WORK HISTORY

Full-time contracts, Dubai, UAE

2021 Nov – Currently

Positions: **Principal Big Data / Machine Learning / AI Architect**

- **Data Mesh / Fabric** - design and implementation, provides flexible, resilient integration of data sources across platforms and business users, making data available everywhere it's needed regardless of where the data lives (AWS, Azure)
- **Cloud-native platforms** – developing business-focused applications that leverage advanced technologies. It allows businesses to build new application architectures that are resilient, elastic, and agile — enabling companies to respond to rapid digital change
- **Composable applications** - next step in microservices approach, applications are built from business-centric modular components (Python, Golang, Javascript, Typescript)
- **Decision intelligence** - providing companies with methodology and roadmaps to improve organizational decision making. It models each decision as a set of processes, using intelligence and analytics to inform, learn from and refine decisions
- **AI Engineering** – leading projects to automate updates to data, models, and applications to streamline AI delivery.
- **Generative AI** - investigating a variety of opportunities ("known unknowns") to learn about artifacts from data, and generates innovative new creations that are similar to the original but don't repeat it. Generative AI has the great potential to create new

- forms of creative content, such as video, and accelerate R&D cycles in a variety of business fields.
- **NFT** - use cases (gaming, music industry), NFT Smart Contracts, NFT Minting, IPFS storage, and NFT Security (ERC-721 Smart Contracts)

Emirates Airlines, Dubai , UAE

2018 March – 2021 Nov

Position: **Principal Big Data / Machine Learning Engineer (contract)**

- 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:
 - Customer, Order, Competitors, Flights, Location, Finance & Planning Genome genomes (Big Data Lake on premises and in Azure, Dataiku Data Science Studio)
 - Enterprise Data Science Platform (Azure) for Advanced Analytics department – Dataiku Studio, Spark, HDInsight, Azure Kubernetes, Azure Blobs and Data Lake
 - Implementation of serverless framework in Kubernetes + Istio for deploying ML models to production
- Active participation in creation of Big Data CoE as the One-Stop Shop for all Big Data related challenges, design, development, review, implementation, support and training that works as a Build-Operate-Transfer mode
- Developing ML models for credit risk management (risk prediction associated with sales agents and agencies) - likelihood of default over time period, anomaly detection, predicting suitable cap
- Implementation data science platform to predict the contingency fuel required for a given flight considering the influencing factors such as weather, payload, recent performance, asset type and airport holding patterns etc
- Advanced Exploration Analysis applied to short-term and long-term planning of cash to identify surplus funds in station bank account for repatriation to HQ bank account and FX exposure on foreign currencies for hedging and converting to USD
- Automatic forecast analysis to help drive category saving strategies - to give the category teams visibility of their historical spend by mapping to the procurement categories, for example Fuel & Oil, Ground Handling, IT & Telecoms & Properties etc
- Practical experience in rule-based engine design (integrates data from different sources and in different formats for analysis), comprehensive monitoring of individual behaviour, detects fraud passing the results in real-time to authorization systems, prioritizes suspicious cases for investigation according to business value

Scotiabank Digital Factory , Toronto, Canada

2017 September – 2018 March

Position: **Senior Data Engineer**

- 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.

ACCENTURE UKI , London, UK

2016 June to 2017 August

Project: Information Management Architecture Strategy (IMAS) in Nationwide Building Society Position: **Big Data Lead:**

- **Implementation of Discovery Analytics / Data Science stream including:**
 - Identify process friction points for mortgages, where a member gets stuck or cycles back, in journey and identify improvements
 - For unsecured loans, identify which members in all segments that are likely to close their account early across any channel to enable Nationwide to proactively manage the relationship
 - Identify real-time opportunities to improve member omni-channel experience by aligning web initiated contact with relevant human interaction at branch or call centre
 - Classify 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 (analyse behavioural 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
- Environment and technologies: AWS, Google Cloud, Apache Spark, Apache Mesos, Cassandra, Kafka, Hive, HDP 2.3, Teradata Aster, Apache Solr,, Zeppelin, Jupyter, Scala, Python, TensorFlow, Keras, R*

CANADIAN TIRE CORPORATION , Toronto, Canada (largest retail company in Canada)
2015 August to 2016 June

Positions: **Lead Data Scientist (cyber security and intelligence)**

- Developing and implementation of multi-layer threat / linked data analysis platform hosted in Big Data environment (HDP) to proactively uncover hidden threats through cyber hunting
 - Design and modelling Security Data Lake (HDFS, Avro, Parquet, HBase, Cassandra) as a common data repository for wide range of security tasks (behavioural monitoring, network anomaly detection, user scoring, correlation engines and so forth)
 - Identification and importance analysis of behavioural 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)
 - Leveraging one 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
 - Discovering novel attack strategies from INFOSEC Alerts using probabilistic (HMM) / Statistical based alert correlation to support root cause analysis
 - Developing 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)
 - 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)
 - Developing the hybrid malicious code detection method based on Deep Learning and the application of Deep learning on traffic identification (R, SparkR)
- Environment and technologies: AWS, Google Cloud, Apache Spark, Apache NiFi, RabbitMQ, Cassandra, Kafka, Hive, HDP, ELK, Zeppelin, Jupyter, Scala, Python, R*

PAYTM LABS , Toronto, Canada (fastest growing Indian e-commerce)
2014 Dec to July 2015

Positions: **Senior Big Data Engineer / Data Scientist**

- Played a lead role in determination of overall solution architectures and designs consistent with architecture to support strategic Big Data initiatives across domains
 - 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)
 - Leading 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
- Environment and technologies: AWS, Apache Spark, Apache Sqoop, RabbitMQ, Cassandra, Kafka, Hive, HDP, Zeppelin, Jupyter, Scala, Python, R*

KINROSS , Toronto, Canada (one of the world's leading gold mining companies)
July 2008 to 2015 Feb

Positions: **Senior BI and Data Science Developer / Project Manager**

- Leads project teams including establishing project plans and milestones, analyzing risks, developing budgets, and delegating work assignments; accountable for results
- World-wide implementation 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.

Environment and technologies: Apache Hadoop, Sqoop, Hive, HBase, Visual.Net, MS SQL, SSIS, SSRS, SharePoint, C#, MicroStrategy 8-9

EDUCATION

Bachelor's degree - **Applied Physics** (Russia, State Polytechnic University, Nizhny Novgorod, 1984-1990),
Bachelor's degree - **Financial Management** (Russia, Moscow university, 2002-2005)

TRAININGS

Microsoft Certified Database Administrator (MS SQL), 2008

Microsoft Certified IT Professional (MS SQL Developer/Administrator), 2008

Microsoft Certified Technology Specialist (SharePoint) 2010

Languages

Russian, English, French
