Nikolai
Senior Data Engineer

SUMMARY

Data Engineer with 7 years of expertise in data analytics/science, ETL, and cloud technologies, blending deep healthcare and pharma industry knowledge. Proficient in Python, SQL, and a suite of data engineering tools including, Apache Spark, Airflow, and BI tools such as Power BI. Implemented real-time data streaming using Kafka, and has experience with multiple cloud services from AWS, Azure, and GCP. Key achievements include optimizing SQL database performance, automating data quality checks, and uncovering new drug candidates through computational data discovery—demonstrating a strong fusion of domain knowledge and technical acumen.

TECHNICAL SKILLS

Main Technical Skills	Python (7 yr.), SQL (7 yr.), JMeter (7 yr.), PySpark (7 yr.), Apache Airflow (4 yr.)
Programming Languages	Python (7 yr.)
.NET Platform	Azure (2 yr.)
AI & Machine Learning	Hugging Face (1 yr.), NumPy (7 yr.), PyTorch (2 yr.), Scikit-learn (3 yr.), TensorFlow (2 yr.)
Python Libraries and Tools	Matplotlib (6 yr.), NumPy (7 yr.), Pandas (7 yr.), Plotly (2 yr.), PySpark (7 yr.), PyTorch (2 yr.), Scikit-learn (3 yr.), SciPy (1 yr.), TensorFlow (2 yr.)
Data Analysis and Visualization Technologies	Apache Airflow (4 yr.), Looker Studio, Pandas (7 yr.)
Databases & Management Systems / ORM	MongoDB (3 yr.), Neo4j (2 yr.), Oracle Database (2 yr.), PostgreSQL (3 yr.), Redis (3 yr.), SQL (7 yr.)
Cloud Platforms, Services & Computing	Azure (2 yr.)
Google Cloud Platform	GCP Cloudbuild (3 yr.)
QA, Test Automation, Security	JMeter (7 yr.)

Virtualization, Containers and Orchestration	Kubernetes (2 yr.)
Other Technical Skills	DB MS SQL (3 yr.)

WORK EXPERIENCE

Data Engineer, COMPUTATIONAL DRUG AND DRUG TARGETS DISCOVERY

Duration: 01.2023 - Present
Summary: Led cross-functional teams to develop and implement data strategies for computational drug and drug target discovery using extensive NLP annotated text databases.
Responsibilities: Led development and implementation of data strategies, created complex SQL scripts, ensured compliance with healthcare regulations, supported data governance initiatives, cleaned and analyzed data with Python and Pandas, developed ML models, created dashboards with Power BI, and more.
Technologies: Python, SQL, Apache Spark, PySpark, Power BI, Pandas, NumPy, SciPy, Matplotlib, HuggingFace, Scikit-learn, AWS, MS SQL, Kafka, Docker, GitHub

Data Engineer, CORPORATE DATA WAREHOUSE

Duration: 03.2020 - 12.2022
Summary: Extracted and transformed data from various web providers to enable a full operational view of an Internet-run company, implementing GDPR compliance and data processing through GCP.
Responsibilities: Created data pipelines from CRM systems, transformed data for GCP BigQuery DWH, ensured GDPR consent requirements, developed Spark-based data pipelines, created reports with Looker Studio, and more.
Technologies: Python, SQL, Apache Spark, PySpark, Apache Airflow, Looker (Data) Studio, Plotly, Pandas, NumPy, Matplotlib, PostgreSQL, MongoDB, GCP, Oracle, Scikit-learn, PyTorch, TensorFlow, Kafka, Docker, Kubernetes, GitHub

Data Engineer, SMART GARDEN

Duration: 01.2018 - 03.2020
Summary: Participated in creating a smart garden solution with cloud-based and local network configurations to track garden sensor readings.
Responsibilities: Analyzed business requirements, performed exploratory data analyses, developed reports and dashboards with Power BI, implemented ETL using Azure Data Factory, and more.
Technologies: Python, SQL, Apache Spark, PySpark, Apache Airflow, Power BI, Pandas, NumPy, Seaborn, Matplotlib, MS SQL, Redis, Neo4j, Azure, Kafka, Docker, GitHub, Azure DevOps

Data Engineer, CORPORATE EDUCATION PLATFORM

Duration: 03.2017 - 12.2017
Summary: Provided data engineering expertise to enable a platform for corporate education and coaching across diverse industries.
Responsibilities: Wrote optimized SQL queries, administered SQL databases, created dashboards with Power BI, developed data processing methods, and handled API management.

Technologies: Python, SQL, Apache Spark, PySpark, Pandas, NumPy, PostgreSQL, MongoDB, Redis, AWS, Power BI, Kafka, Docker, GitHub

Head of Pharmacological Sector, Pharmacological Sector

Duration: 01.2015 - 02.2017
Summary: Overlooked bioequivalence studies and registration dossier formation within the pharmacological sector of a pharmaceutical company.
Responsibilities: Reviewed study protocols and reports, managed regulatory authority liaisons, and negotiated with clinical research organizations.

Manager of Pharmacovigilance Department, Pharmacovigilance Department

Duration: 04.2013 - 01.2015
Summary: Managed the pharmacovigilance department, designed risk management plans and standard operative procedures, and ensured medicine safety.
Responsibilities: Reviewed medicine safety profiles, updated safety reports, managed case safety reports, and liaised with regulatory authorities.

Doctor, Surgeon, Medical Services

Duration: 08.2007 - 12.2012
Summary: Provided medical services as a Doctor and Surgeon, diagnosed and treated patients, and educated medical students and junior doctors.
Responsibilities: Investigated and diagnosed health issues, arranged treatments, liaised with nurses, consulted other departments, performed surgical procedures, and more.

EDUCATION

- **Medical Doctor Master's Degree Primary Care**
  University of Glasgow
  Not Provided

CERTIFICATION

- **Exam DP-203: Data Engineering on Microsoft Azure**
  Certification in Data Engineering on Azure
  2022