

PROFESSIONAL SUMMARY

Experienced, results-oriented and problem-solving Data Engineer with 5+ years of diverse experience in Information Technology, including Development and Implementation of various Big Data applications as well as DevOps culture (CI/CD), mainly in cloud environments. Strong team player with excellent communication skills, as well as effective working independently.

EXPERIENCE

BI4ALL - Data Engineer

Warsaw, Poland

02/2023 - Current

- Successfully designed and implemented a complex **ETL** pipeline using **Databricks (Python, PySpark, Spark SQL)** and **Azure Data Factory** to efficiently ingest and transform data from multiple sources into a **Data Lake** and **Azure SQL Database**.
- Automated data processing using **Databricks Jobs** and **Data Pipelines (DLT)**, improving overall efficiency and accuracy of the data processing with **Spark**.
- Created **CI/CD** pipelines in **Azure DevOps** to automate development implementation with **Databricks** and **Azure Data Factory**, enabling faster and more reliable deployment of data pipelines.

Procter & Gamble - DevOps / Data Engineer - IT Manager

Warsaw, Poland

03/2022 - 02/2023

- Developed and maintained a highly scalable **Big Data** application using **Python, Spark SQL, PySpark** and deployed it on **Azure Kubernetes Service** with **Spark** pods implementation (100M records per processing)
- Created a **Docker**-based app development environment (**Python, Spark, YAML**) and integrated it with **Redis** Cache for efficient app development.
- Created custom log queries using **Azure Log Analytics** Query Language (KQL) to search, filter, and analyze app performance.
- Ingested data with **Python, PySpark, Spark SQL** and optimised query, processing using **Spark**.
- Optimized **PostgreSQL** database to improve query performance.
- Built and managed **Data Pipelines** and **Jobs** with **Databricks (Delta Live Tables, PySpark, Python, SQL)** and **Azure Data Factory** to streamline data processing.
- Built and maintained **CI/CD** pipelines using **Azure DevOps** and managed infrastructure environments with **Terraform**.
- **Led a team** of 2 Data Engineers in the refactoring of the application architecture and implementation of the **Big Data** App solution, utilizing **Agile** methodology.

SKILLS

Programming Languages - Python, Spark, PySpark, SQL, Java and Shell Scripting.

Databases - MSSQL, PostgreSQL, MySQL, MongoDB, Redis, Neo4j.

ETL development, **Big Data** processing, **Performance** optimization.

Databricks - Delta Lake, Delta Live Tables, Jobs, Data Pipelines, .

Azure - Azure Data Factory, Azure Synapse Analytics, Azure Log Analytics Azure Data Lake, Azure DevOps, Functions, WebApps, AKS, Cosmos DB, SQL Databases, VMs.

AWS - Redshift, Athena, Glue, Lambda, S3, EC2.

Tools - Docker, Kubernetes, CI/CD pipelines, Terraform, Jira, GitHub.

Data Visualisation - PowerBi, Tableau, Databricks SQL Dashboards.

Methodologies - Agile, Scrum, Waterfall, DevOps.

Machine Learning algorithms - Linear and Logistic Regression, Decision tree, SVM, Naive Bayes, KNN, K-means, Random Forest.

Operating systems- Linux, MacOS, Windows.

CERTIFICATIONS

Databricks Certified: Data Engineer Associate

Databricks Certified: Associate Developer for Apache Spark

Databricks Certified: Lakehouse Fundamentals

Microsoft Certified: Azure Data Engineering Associate

Microsoft Certified: Azure Developer Associate

Microsoft Certified: Azure Data Scientist Associate

Microsoft Certified: Azure Data Fundamentals

Microsoft Certified: Power BI Data Analyst Associate

Atos - Junior Data Scientist

Warsaw, Poland

08/2021 - 02/2022

- Designed and implemented an **ETL** pipeline with **Databricks (Python, Pyspark, Spark SQL)** and **Azure Data Factory** to load data into **MySQL** database.
- Created reports and dashboards with **Tableau** and **PowerBI** for data-driven decision-making.
- Assisted in creating **Machine Learning** models for predictive analytics.
- Utilized **Azure Synapse Analytics** and **Azure Log Analytics** to troubleshoot and resolve data processing issues.

Nissan Sales CEE - Data Engineer

Warsaw, Poland

09/2019 - 08/2021

- Built scripts using **Python, PySpark and Spark SQL** to schedule **Databricks Jobs** for financial calculations and task automation.
- Developed and maintained data models in **MSSQL** database, migrating to **Azure SQL Database**.
- Built data pipelines with **Azure Data Factory** for data processing.
- Implemented **Azure Functions** with **Python and Spark** for data extractions.
- Built **Azure DevOps** project structure and **CI/CD** pipelines for consistent deployment of code and infrastructure. Integrated environments.

EDUCATION

06/2023

Warsaw University of Technology

Warsaw, Poland

Master of Engineering: Computer Science - Specialization in Data Engineering.

- Master's thesis: **"NoSQL databases in Big Data solutions"**.

01/2022

Warsaw University of Technology

Warsaw, Poland

Bachelor of Engineering: Electrical Engineering

- Engineering's thesis: **"Project of software to automatically select electric cars"**.
- Head of Software department in Scientific Association Proton Dynamic and Aria.

AWS Certified: AWS Cloud

Practitioner

DevOps Institute Certified: DevOps Foundation

ADDITIONAL TRAININGS

Databricks Data Engineer Academy - Accenture

Developing Data Engineering Solutions with **Azure** and **Databricks** using **Python, PySpark, Spark SQL, Delta Live Tables** and tools: **Delta Lake, Azure Data Factory, Databricks Jobs, CosmosDB, Azure Data Lake, Azure Log Analytics**.

Cloud Data Engineer Academy - Onwelo

Developing Data Engineering Solutions with **AWS** using **Python, PySpark, Spark SQL** and tools: **AWS Glue, Redshift Athena, S3, EC2**.

LANGUAGES

Polish: First Language

English: C2
Proficient

German: B1
Intermediate

Spanish: A1
Beginner

HOBBIES

Automotive, sports mainly boxing, football and tennis. I also love to be up to date with new technologies like blockchain or cryptocurrencies.