

Hiring software engineer as easy as calling a taxi.

# Ihor K

## Big Data & Data Science Engineer with BI & DevOps skills

### SUMMARY

- Data Engineer with a Ph.D. degree in Measurement methods, Master of industrial automation
- 16+ years experience with data-driven projects
- Strong background in statistics, machine learning, AI, and predictive modeling of big data sets.
- AWS Certified Data Analytics. AWS Certified Cloud Practitioner. Microsoft Azure services.
- Experience in ETL operations and data curation
- PostgreSQL, SQL, Microsoft SQL, MySQL, Snowflake
- Big Data Fundamentals via PySpark, Google Cloud, AWS.
- Python, Scala, C#, C++
- Skills and knowledge to design and build analytics reports, from data preparation to visualization in BI systems.

### SKILLS

<b>Main Technical Skills</b>	AWS big data services (5 yr.), Python, Kafka, ETL, Microsoft Azure (3 yr.)
<b>AI &amp; Machine Learning</b>	AWS ML (Amazon Machine learning services), Keras, Machine Learning, OpenCV, TensorFlow, Theano
<b>Programming Languages</b>	C#, C++, Python, Scala
<b>Java Frameworks</b>	Apache Spark
<b>Scala Frameworks</b>	Apache Spark, Apache Spark 2
<b>Python Libraries and Tools</b>	Big Data Fundamentals via PySpark, Deep Learning in Python, Keras, Linear Classifiers in Python, Pandas, PySpark, TensorFlow, Theano
<b>.NET Platform</b>	.NET, .NET Core, .NET Framework
<b>Data Analysis and Visualization Technologies</b>	Apache Airflow, Apache Hive, Apache Oozie 4, Apache Spark, Apache Spark 2, Data Analysis, ETL

<b>Databases &amp; Management Systems / ORM</b>	Apache Hadoop, Apache Hive, Apache Spark, Apache Spark 2, AWS Database, dbt, HDP, Microsoft SQL Server, pgSQL, PostgreSQL, Snowflake, SQL
<b>Cloud Platforms, Services &amp; Computing</b>	AWS, GCP
<b>Amazon Web Services</b>	AWS big data services (5 yr.), AWS Database, AWS ML (Amazon Machine learning services), AWS Quicksight, AWS Storage
<b>Google Cloud Platform</b>	GCP AI, GCP Big Data services
<b>Message/Queue/Task Brokers</b>	Apache Kafka 2, Apache Oozie 4, Kafka
<b>Deployment, CI/CD &amp; Administration</b>	Kubernetes
<b>BlockChain and Decentralized Software</b>	OpenZeppelin
<b>Third Party Tools / IDEs / SDK / Services</b>	Qt Framework, YARN 3
<b>Other Technical Skills</b>	SPLL, Superset

## WORK EXPERIENCE

### Data Engineer

Apr-2011 To Till now

#### Project: AWS ELT data pipeline and AWS cloud deployment architecture

(2022-06 – current)

Project Description: Creation of ELT pipelines deployed on AWS to collect data from e-commerce platforms

#### Responsibilities:

- architecture design of ELT pipeline that gathers data from e-commerce clients into a single data warehouse.
- using DBT for processing customer data and identifying similar attributes.
- setting up Airbyte connections, developing custom Airbyte connectors, deploying AWS architecture, Terraform scripting
- building custom data management tools, creating data flow security solutions

**Tools & Technologies:** Python, Airbyte, Kubernetes, AWS EC2, CI/CD, OpenVPN server, AWS Lambda, AWS SQS, Fargate, BigQuery, DBT, Airflow, AWS Cloudwatch, REST API, AWS ECR



## **Project: Batch and Streaming Data Ingest into DataLake**

(2021-10 – 2022-05)

### **Responsibilities:**

- Design data processing pipelines for medical/ marketing/ e-commerce applications.
- Data Modelling,
- Database Design,
- Database development,
- using DBT for processing patient data,
- Big data processing using Spark Scala,
- Distributed platform development,
- ETL Data Transformation
- ETL Architecture and ETL Solutions Design

**Tools & Technologies:** Python, Scala, DB (SQL, PostgreSQL), DBT, Spark, Hadoop, Terraform, Kubernetes, Helm, GitLab CI/CD, AWS, Keycloak, Swagger, AirFlow.

## **Project: Audience Segmentation**

(2018 - 2021)

Building a custom customer data platform for a marketing company. Build an ETL pipeline that allows retrieving the data from multiple sources and storing them in the private data warehouse in Hadoop. Create CloudFormation "infrastructure as a code" description of the pipeline and CI/CD to deploy it into the desired environment. Work with streaming data in Amazon Kinesis. Design sources for BI reports in AWS.

### **Responsibilities:**

- Design and implement batch and event-driven workflows for big data processing
- Automated tests for distributed applications
- Data analysis and visualization
- Develop applications for data ingestion and selection
- Develop a recommendation system
- Built reporting dashboards in QuickSight from Athena sources.

**Tools & Technologies:** Python, Scala, SQL, Kubernetes, Spark, Hadoop framework, Docker, AWS (Storage, Database, DocumentDB, Athena, Lambda, Glue, API Gateway, Kinesis, QuickSight, CI/CD AWS CloudFormation and CodePipeline), Grafana, Git.

## **Data scientist and Data/software engineer**

(Jan-2011 To 2018)

- data analysis
- applying machine learning algorithms
- image analysis
- image recognition
- neural networks developing and tuning



- Database development
- ETL operations engineering
- Development of backend services for data curation
- Automated tests for CI/CD workflows

**Tools & Technologies:** C#, Python, Keras, TensorFlow, Theano, OpenCV, Pandas, Microsoft SQL Server, SQL, .NET Framework,

### **Associate Professor**

(09/1999–Present)

Department of industrial automation

#### **Taught courses:**

- Database development
- Database management systems
- Object-oriented programming
- Parallel programming
- System programming
- Development .NET applications

## **EDUCATION AND TRAINING**

- Measurement methods and devices Ph.D. Degree, EQF level 8
- Master of industrial automation, EQF level 7

## **COMMUNICATION SKILLS**

- Communication skills both oral and written gained as a university professor and R&D projects participant
- Presentation skills gained as a scientific conference speaker

## **COURSES & CERTIFICATES:**

- AWS Certified Data Analytics
- HDP Overview: Apache Hadoop Essentials (SPLL)
- Feature Engineering with PySpark
- Big Data Fundamentals via PySpark
- Deep Learning in Python
- Intermediate Python for Data Science
- Linear Classifiers in Python
- Machine Learning with the Experts
- Python Data Science Toolbox

