

# Vadym S.

## Middle Data Engineer

### SUMMARY

- 4+ years of experience as a Data Engineer, focused on ETL automation, data pipeline development, and optimization; - Strong skills in SQL, DBT, Airflow (Python), and experience with SAS, PostgreSQL, and BigQuery for building and optimizing ETL processes; - Experience working with Google Cloud (GCP) and AWS: utilizing GCP Storage, Pub/Sub, BigQuery, AWS S3, Glue, and Lambda for data processing and storage; - Built and automated ETL processes using DBT Cloud, integrated external APIs, and managed microservice deployments; - Optimized SDKs for data collection and transmission through Google Cloud Pub/Sub, used MongoDB for storing unstructured data; - Designed data pipelines for e-commerce: orchestrated complex processes with Druid, MinIO, Superset, and AWS for data analytics and processing; - Worked with big data and stream processing: using Apache Spark, Kafka, and Databricks for efficient transformation and analysis; - Amazon sales forecasting using ClickHouse, Vertex AI, integrated analytical models into business processes; - Experience in Data Lake migration and optimization of data storage, deploying cloud infrastructure and serverless solutions on AWS Lambda, Glue, and S3.

### TECHNICAL SKILLS

<b>Main Technical Skills</b>	Python, PySpark, Docker, Apache Airflow, Kubernetes
<b>Programming Languages</b>	Python, Scala
<b>C++ Libraries and Tools</b>	C/C++/C#
<b>Mobile Frameworks and Libraries</b>	Crashlytics
<b>AI &amp; Machine Learning</b>	NumPy, Scikit-learn, TensorFlow
<b>Python Libraries and Tools</b>	NumPy, Pandas, PySpark, Scikit-learn, TensorFlow
<b>Data Analysis and Visualization Technologies</b>	Airbyte, Apache Airflow, Apache Hive, AWS Athena, Databricks, Pandas
<b>Databases &amp; Management Systems / ORM</b>	Apache Druid, Apache Hive
<b>Amazon Web Services</b>	AWS Athena, AWS EMR, AWS Glue
<b>Azure Cloud Services</b>	Databricks
<b>SDK / API and Integrations</b>	API, Stripe
<b>Virtualization, Containers and Orchestration</b>	Docker, Kubernetes

## Other Technical Skills

Delta lake, DMS, Xano

## WORK EXPERIENCE

### Data Engineer, NDA

(July 2021 - Present)

#### Data Engineer, ETL Automation

**Summary:** Building and automating ETL data pipelines with a focus on optimizing PostgreSQL models in the DBT cloud, integrating with third-party APIs using Python, and refactoring Zeppelin notebooks.

**Responsibilities:** ETL Automation, designing and implementing storage systems, managing API integrations, developing PostgreSQL models in DBT Cloud, establishing microservice deployment jobs, and refactoring notebooks.

**Technologies:** Python, PySpark, Zeppelin, Docker, AirFlow, Kubernetes, MiniKube, S3, Athena, ECR, PubSub, DBT Cloud, Airbyte, API, BigQuery, PostgreSQL, HiveDB, GitHub, GitLab, Miro, Jira, Teams

#### Data Engineer, Analytic Platform

**Summary:** Advanced SDK development to optimize data reception from API endpoints, transformation into special format events, and efficient transmission to Google Cloud Pub/Sub, incorporating MongoDB and Google Cloud storage.

**Responsibilities:** Optimizing SDK data reception, transforming data into event formats, data transmission with Pub/Sub, integrating MongoDB, and using Google Cloud storage.

**Technologies:** Python, GCP storage, PubSub, API, MongoDB, GitHub, BitWarden, Jira, Confluence

#### Data Engineer, E-commerce platform

**Summary:** Orchestration of complex data pipeline for an e-commerce platform, focusing on data transfer, ingest, processing, and optimization using Airflow, Druid, Minio, and Superset for visualization, and building architectures with AWS.

**Responsibilities:** Data pipeline orchestration, architectural planning and visualization, workflow optimization, data processing with Spark and Kafka, and implementation with AWS services.

**Technologies:** Python, Airflow, Druid, Minio, MongoDB, Spark, Kafka, AppFlow, Glue, Athena, Quicksight, PostgreSQL, GitLab, Superset, InSight, Draw.io, Jira, and Confluence

#### Data Engineer, Retail Platform

**Summary:** Developed a technical pipeline for a retail platform, emphasizing economic efficiency, integrating key technologies like AWS, Firebase, and Stripe, and utilizing no-code solutions with Xano.

**Responsibilities:** Technical pipeline development, data transfer optimization, authenticating users with Firebase, payment integration with Stripe, enhancing data processing with AWS IoT, and utilizing Xano's no-code solution.

**Technologies:** Python, AWS, Xano, Firebase, API, Stripe



### **Data Scientist, E-commerce Analytic Platform**

**Summary:** Big data analysis and sales forecasting for Amazon product sales, utilizing advanced statistical and programming skills.

**Responsibilities:** Collecting historical data, preparing sales forecasts, big data analysis, and predictive modeling.

**Technologies:** Sales Prediction, ClickHouse, Vertex AI, AirFlow, Jenkins, Kibana logs, Keepa

### **Data Engineer, Simulation, and Automation Worker traffic**

**Summary:** Created a simulation and automation framework for worker traffic, automated EC2 instance management, and deployed solutions using containerization and AWS cloud services.

**Responsibilities:** Developing simulation and automation framework, managing EC2 instances, and deploying containerized solutions.

**Technologies:** Python, EC2, ECR, Docker, Windows

### **Data Engineer, Hotel & Restaurant**

**Summary:** Led the optimization of cloud infrastructure for the hotel industry using AWS services, improving performance, scalability, and cost-effectiveness.

**Responsibilities:** Cloud infrastructure review and optimization, code refactoring, enhancement of AWS services, and pipeline setup for room price prediction.

**Technologies:** AWS Lambda, Glue, ECR, DMS, EventBridge, SNS, API Gateway, S3, Python

### **Data Engineer / Big Data Engineer, Scalable ETL Pipeline with Databricks and AWS**

**Summary:** Designed and implemented an ETL pipeline with Databricks and AWS, processing large-scale data with Apache Spark and integrating with AWS services for schema management, data governance, and real-time processing.

**Responsibilities:** Designing and implementing end-to-end ETL pipeline, data transformation, and cleaning, metadata and schema management, querying and dashboard integration, and maintaining cost efficiency.

**Technologies:** Databricks, AWS (S3, Glue, Lambda, EMR), Apache Spark, Delta Lake, Python, SQL

## **EDUCATION**

### **Bachelor in Software Engineer**

West Ukrainian National University (WUNU) is a classical university of Ternopil, a leading modern education institution.

2020 - Present

## **CERTIFICATION**

- **Programming for Everybody (Getting Started with Python)**



Coursera certificate

- **What is Data Science?**

Coursera certificate

- **Introduction to Data Science in Python**

Coursera certificate

- **Applied Machine Learning in Python**

Coursera certificate

- **Amazium Data Science Camp**

Amazium certificate

- **Machine learning with Python**

Coursera certificate

- **Google Cloud Big Data and Machine Learning Fundamentals**

Coursera certificate

