Andrii
Middle Data Engineer, ETL, Data QA

SUMMARY

- 8+ years of experience as a Data Engineer, Data QA, and Data Science challenges. - AWS Certified Data Analytics.
- Experience creating and supporting of ELT data pipelines built on Snowflake and DBT
- Warehousing models, and support existing processes/ETL's (extract/transform/load), functions (in Python/SQL/DBT) in a cloud data warehouse environment using Snowflake, AWS services
- Software developer with good communication and organizational skills.

TECHNICAL SKILLS

Main Technical Skills	Python, AWS, Kafka, Apache Airflow, Snowflake
Programming Languages	C, C++, Java, Python
Java Frameworks	Apache Spark
Scala Frameworks	Apache Spark
Python Libraries and Tools	Dask, Keras, Matplotlib, NumPy, Pandas, PyTorch, TensorFlow
AI & Machine Learning	Keras, LSTM, NumPy, OpenCV, PCA, PyTorch, TensorFlow
Data Analysis and Visualization Technologies	Apache Airflow, Apache Spark, Decision Tree, k-NN, Logistic regression, Pandas, Random Forest
Databases & Management Systems / ORM	Apache Hadoop, Apache Spark, GC SQL, Snowflake, SQL
Cloud Platforms, Services & Computing	AWS, GCP
Amazon Web Services	AWS big data services, AWS S3
Google Cloud Platform	Google AI platform, Google BigQuery, Google Cloud AI, Google Data Studio, StackDriver
Deployment, CI/CD & Administration	CI/CD, GitLab CI
SDK / API and Integrations	GC API, Livescore API, messengers API
Version Control	GitLab
Message/Queue/Task Brokers	Kafka

Virtualization, Containers and Orchestration	Kubernetes
Mail / Network Protocols / Data transfer	LDAP
QA, Test Automation, Security	Selenium Webdriver
Other Technical Skills	BigData services, clusterization, Data QA, grab, Naive Bayes

PROJECTS

Data Engineer, cloud DevOps

April 2022 - now

Responsibilities

- setting up Airbyte connections, developing custom Airbyte connectors, deploying AWS architecture, building custom data management tools, creating data flow security solutions
- AWS ELT data pipelines
- AWS cloud architecture deployment
- Creation of ELT pipelines deployed on AWS to collect data from e-commerce platforms

Technologies: Python, Airbyte, Kubernetes, AWS EC2, CI/CD, OpenVPN server, AWS Lambda, AWS SQS, Fargate, Bigquery, Airflow, AWS Cloudwatch, REST API, AWS ECR

Data Engineer, Data management platform (Amazon e-Commerce Aggregator)

2021-present

Overview:

A next-generation consumer goods company reimagining how the world's most-loved products become accessible to everyone. We use a deep understanding of rankings, ratings, and reviews to identify and acquire quality brands and use world-class expertise and data science to make their products better or create new ones to meet changing customer demand.

Responsibilities:

- Create and support ELT data pipelines built on Snowflake and DBT while ensuring high-quality data
- Develop and deploy data warehousing models, and support existing processes/ETLs (extract/transform/load), functions (in Python/SQL/DBT) in a cloud data warehouse environment using Snowflake, AWS services
- SQL statements and developing in Python
- Design and develop data pipelines (DAGs). Automation tests.

Technologies: AirFlow, Snowflake, Python3, AWS S3, Hadoop, Spark, GitLab CI/CD, Kubernetes, LDAP.

Data Science Engineer, Audience Segmentation

2020

Overview:

Develop of recommendation system, user analysis, bot searching, and analytics report

Responsibilities:

- Building models
- Analytics
- Python Development & Automation

Technologies: Python3.7, Google AI platform, Big Query, Data storage, AI platform automation tool, NumPy, TensorFlow, Pandas, Google Data Studio, Dask, SQL, GitLab CI/CD

Data Science Engineer, Steel Defect Detection research

2019

Overview: Building a model for image classification and segmentation of damaged surfaces on steel sheets.

Responsibilities:

- Building models
- Data analytics
- Implementation in Python

Technologies: Python3.7, GC compute engine, TensorFlow, PyTorch, Google Data Studio, Matplotlib, OpenCV, Keras

Automation Engineer, Google Cloud AI platform, automation

2019

Overview: Goggle Cloud resources CI/CD implementation for Data Science pipelines, that allows to scale resources, train and predict custom models as specific jobs and save results.

Responsibilities:

- CI/CD implementation for data science pipelines (GCP environment)
- Developing and implementing train/prediction models

Technologies: Python3.7, AI platform, Big Querry, Data storage, GitLab CI/CD, StackDriver, GC API, NumPy, TensorFlow, OpenCV

Data Science Engineer, Data import

2018

Overview: Developed a data processing algorithm to automate the process of analysis, cleaning, transformation, and merging damaged data.

Responsibilities:

- Engineering a data processing algorithm

Technologies: Python3.5, NumPy, pandas, Google cloud, Matplotlib

Data Engineer, Apple stocks forecasting research

2018

Overview: A created model for apple stocks forecasting based on previous data

Responsibilities:

- Built a forecasting model

Technologies: Python3.6, Keras, LSTM, NumPy, pandas.

Software Engineer, Football Data Scraping

2017

Overview: Developed tool for automatically football data scraping from 50+ webpages, created a statistics report.

Responsibilities:

- Developing a scraping tool, design, and implementation
- Developing statistics reports

Technologies: Python3.6, pandas, grab, Selenium.

Software Engineer, Live-score bots

2017

Overview: Developed bots for different messengers that gives advanced information about football teams, based on parsed Livescore data.

Responsibilities:

- Developing bots
- Livescore data scraping and messaging

Technologies: Python3.6, pandas, NumPy, GC SQL, Livescore API, messengers APIs.

EDUCATION & CERTIFICATES

- **AWS Certified Data Analytics.**2022
- **Cybersecurity (Bachelor) – National Technical University**2019
- **Data Engineer and Data Science career tracks – DataCamp**
  2018 – 06/2019
- **Data Science fundamentals – Data Science Camp, Ukraine**
  2018 – 2019
- **Specialist in Software Development – Technical National University**2015 – 2019