Best Technology Talent For
Companies Around The World



ncube

Anton

Senior Data Engineer

Summary

Data Engineer with almost 10 years of experience designing, building, and maintaining data ingestion and transformation pipelines with tools such as Spark, NumPy, Airflow, Kubeflow, Argo Workflows, and cloud services. Built and utilised well-performing predictive algorithms in domains such as NLP, LLM, Computer Vision, and Time Series.

Strong hands-on Snowflake expertise across platform architecture, ingestion, SQL-based ELT, analytical modelling, query optimisation, access control, and cost management.

Solid theoretical and practical knowledge of key machine and deep learning principles and algorithms with particular attention on getting solutions to production and getting value out of them in a stable, fault-tolerant manner.

TECHNICAL SKILLS:

- Programming languages: Python, SQL, Terraform, TypeScript (for cloud resource management with CloudFormation/CDK), Scala, Groovy, Argo Workflows DSL, CI/CD languages, Docker
- Snowflake: warehouse architecture and administration, data ingestion, SQL transformations, incremental processing, analytical data modelling, performance tuning, role-based access control, monitoring, and cost optimisation
- Frameworks and Libraries: TensorFlow, PyTorch, Pandas, NumPy, Scikit-Learn, Flask, FastAPI, HuggingFace Transformers (BERT, GPT), Ray, Celery, RabbitMQ, Airflow, Kubeflow, Argo Workflows, MLflow, PySpark, AWS MWAA, AWS Glue, AWS ECS, Hive, Hadoop, Presto
- AWS: CodeCommit, CodeBuild, EC2, VPC, S3, EFS, Lambda, DynamoDB, ECR, EKS, EMR, Kinesis, MWAA, Glue, ECS, Step Functions, EventBridge, SQS, SNS, Aurora, QLDB, CloudWatch, CloudTrail, CloudFormation, IAM/STS/SSM/Organizations, Redshift

- GCP: BigQuery, Dataflow, Pub/Sub, Cloud Storage, Cloud SQL, Cloud Spanner, Dataproc, Vertex AI, AI Platform, Cloud Composer, Cloud Data Fusion, Cloud Functions, Cloud Run, GKE, Firestore, Stackdriver, IAM
- Azure: Azure DevOps (Pipelines, Releases, Artifacts, Repos, Boards), AKS, ACR, Storage Accounts, VNets, SPN
- Oracle Cloud: Autonomous Data Warehouse, Data Integration, Data Safe, Big Data Service, Object Storage, GoldenGate, Streaming Service, Cloud SQL, Analytics Cloud, NoSQL Database, Data Flow, Data Catalog
- Databases & Data Platforms: MySQL, PostgreSQL, MongoDB (Atlas), Amazon Redshift, Snowflake, AWS DynamoDB, AWS RDS, Apache Iceberg, SAP BW (source integration and data extraction)
- DevOps & Infra: Kubernetes, Helm/HULL, Kustomize, ArgoCD, Docker, Jenkins, Ansible, Terraform, GitHub Actions, CircleCI, Bitbucket Pipelines, Maven
- ML / AI: NLTK, SpaCy, MLflow, KNIME
- Practices: Agile/Scrum, CI/CD, TDD, Design Patterns, Data Modelling, DRY, ACID, MLOps, Solutions Architecture
- IDEs & Tools: VS Code, PyCharm, Jupyter Notebook, Git, GitHub, Bitbucket

Professional Career

Senior Data Engineer

March 2025 – Present | **Slovakia**

Project Info:

Data Platform: European last-mile delivery and logistics company. Building and maintaining the data platform for operational reporting and analytics.

Responsibilities:

- Built and maintained ingestion and SQL transformation pipelines in Snowflake for operational reporting and analytics

- Developed data models and incremental transformations with SQL, dbt, and PySpark; scheduled pipelines with Airflow

- Maintained Kubernetes deployments and CI/CD workflows for data services

- Troubleshot data quality and pipeline failures; tuned Snowflake queries and warehouse usage, and managed role-based access

- Gathered requirements, contributed to solution design, reviewed code, and supported other engineers

Software and technologies:

Python, SQL, PySpark, Airflow, Argo Workflows, AWS (S3, EMR, Redshift), Snowflake, Kafka, Kubernetes, Helm, ArgoCD, Terraform, dbt, GitHub Actions

Senior Data Engineer

September 2021 – February 2025 | **Ukraine**

Project 1 Info:

Snowflake Data Platform: Enterprise data pipelines in Snowflake: source integration, SQL transformations, analytical data models, incremental processing, validation, and reconciliation. Full query tuning, warehouse management, role-based access control, monitoring, and production support.

Responsibilities:

- Developed enterprise data pipelines in Snowflake, from source integration and SQL transformations to analytical data models, incremental processing, validation, and reconciliation
- Tuned Snowflake queries and warehouse usage; managed role-based access, monitored pipeline execution, and resolved production data and performance issues

Project 2 Info:

Automation software for gathering and governing data from clinical institutions, including enterprise source integrations and data migration. Amsterdam HQ.

Responsibilities:

- Analysed source data structures, existing models, and business rules to define target data mappings
- Designed target data models and pipelines; created project mapping and architecture for enterprise-specific requirements
- Integrated SAP BW as an upstream source, using existing BW data models and extracting datasets through ODBC/Python for downstream transformation and validation
- Reviewed code, guided mid-level engineers, and pair-programmed on complex work
- Configured and maintained infrastructure for data pipelines using Airflow, Argo Workflows, Spark, Kubernetes, Postgres, and Object Storage
- Set up automated CI/CD pipelines for microservices using Bitbucket and Jenkins.

Software and technologies:

SAP BW, ODBC, Python, AWS (CodeCommit, CodeBuild, AMI, EC2, VPC, S3, EFS, Lambda, Python SDK, Organizations, STS, SSM, DynamoDB, ECR, EKS, CloudWatch, CloudTrail, CloudFormation), Azure (DevOps: Pipelines, Releases, Artifacts, Repos, Boards, SPN, AKS, ACR,

Storage Accounts, VNets), Jenkins, Ansible, Kubernetes, Helm, HULL, PySpark, Argo Workflows, Hive, Hadoop, Oracle (Autonomous Data Warehouse, Data Integration, Data Safe, Big Data Service, Object Storage, GoldenGate, Streaming Service, Cloud SQL, Analytics Cloud, NoSQL Database, Data Flow, Data Catalog)

Project 3 Info:

Automotive SaaS — Full cycle automated automotive repair: SaaS product, on-site consulting, and R&D department. Data transformation at scale (~50 TB of data monthly) across AWS, GCP, and Azure.

Responsibilities:

- Gathered requirements for custom solutions and delivered proposed architectures constrained by the technical landscape of the organisation (AWS, GCP, Azure)
- Implemented and led development of data transformation pipelines at scale (~50 TB monthly).
- Reviewed code, guided mid-level engineers, and pair-programmed on complex work
- Troubleshot and adjusted pipelines and environments as needed; managed Helm/HULL deployments
- Built CI/CD pipelines according to established corporate standards

Software and technologies:

PySpark, Airflow, Kubernetes, Argo Workflows, Hive, Hadoop, AWS EMR, AWS S3, AWS Kinesis, MWAA

Machine Learning Operations and Data Engineer

December 2019 – August 2021 | **Ukraine**

Project Info:

Singapore Enterprise Banking System: Development of DAGs and machine learning pipelines at scale for a Singapore-based enterprise banking client, covering ML/DL model deployment, continuous retraining, and cloud application architecture.

Responsibilities:

- Deployed ML/DL models at scale (cloud and on-premises) combined with distributed training and data gathering
- Developed Kubeflow pipelines for continuous retraining and ML/DL model inference
- Set up automated CI/CD pipelines for microservices and ML/DL models
- Participated in designing cloud application architecture and creating Cloud services architecture
- Reviewed code, guided mid-level engineers, and pair-programmed on complex work

Software and technologies:

Apache Kafka, Apache Airflow, BigQuery, Grafana, Docker, AWS (Step Functions, Lambda, EventBridge, QLDB, Aurora, SQS, SNS), NLTK, SpaCy, KNIME, TensorFlow, Keras, Pandas, NumPy, Scikit-Learn, PyTorch, PySpark, Kubeflow, Ray

Machine Learning Engineer

October 2018 – August 2019 | **Ukraine**

Project Info:

Direct Mail Analytics. Uncovering patterns in 25 years of physical mail piece data: successful design, text usage, target audience, and response patterns using deep learning and machine learning on large-scale NLP and image datasets.

Responsibilities:

- Analysed large amounts of text and handwritten data using Natural Language Processing (NLP) techniques
- Performed image recognition of direct mail pieces using computer vision models
- Conducted requirements analysis and clarification with business stakeholders
- Performed code reviews, development, and product support; maintained documentation
- Shared knowledge and delivered training to team members

Software and technologies:

Python, NLTK, SpaCy, KNIME, TensorFlow, Keras

Data Engineer

July 2017 – October 2018 | **Ukraine**

Project Info:

Expert Network Platform. Web-based platform connecting top-tier consultants with enterprise clients, with analytics and segmentation capabilities for user behaviour, churn, and lifetime value.

Responsibilities:

- Calculated and tracked Churn, Retention, CAC, LTV, and User Behaviour metrics
- Performed segmentation and clustering of the user base for targeted engagement
- Designed and implemented data pipelines within the existing infrastructure
- Developed web platform features and improved text recognition technology
- Reviewed code, guided mid-level engineers, and pair-programmed on complex work

Software and technologies:

Django, Redis, MongoDB, Docker, PostgreSQL, NLTK, SpaCy

Data Analyst

January 2017 – July 2017 | **Ukraine**

Project Info:

ML Automation Application. Fully automated machine learning application following leading standards in Data Science, Machine Learning, and MLOps.

Responsibilities:

- Cleaned and shaped datasets for model training and evaluation
- Analysed data and uncovered insights to support business decision-making
- Solved classification and regression problems using KNN, SVM, Naive Bayes, and custom ensemble models
- Automated marketing flows to improve campaign targeting and efficiency
- Reviewed code, guided mid-level engineers, and pair-programmed on complex work

Software and technologies:

Pandas, NumPy, R, Scikit-Learn, Salesforce, Apex

Education

Zaporizhzhia National Technical University

Master's Degree in Computer Systems and Network Engineering

Certifications

- Associate Data Practitioner — GCP
- Professional Cloud Database Engineer — GCP
- L4-DE: Data Engineering in KNIME Analytics

Foreign Languages Skills

Ukrainian: Native

English: Upper-Intermediate