ANTON VLASENKO *(Location: US/EU)*

Summary of Qualifications

Data and Machine Learning Architect/Engineer with 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 utilized well-performing predictive algorithms in domains such as NLP, LLM, CV, and Time Series.

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.

Performs equally well in academic and engineering environments.

Passionate about bringing an evolving engineering approach to the Data Science discipline.

Technical Skills

- Programming languages and technologies: Python, SQL, Terraform, TypeScript (for cloud resource management with CloudFormation/CDK), Scala, Groovy, Argo Workflows DSL, CI/CD languages, Docker.
- Operating Systems: Mac, Linux.
- Programming practices: Agile Development, Continuous Integration and Deployment (CI/CD), Test-Driven Development (TDD), Design Patterns, Data Modeling, DRY, ACID.
- Frameworks and Libraries: TensorFlow, PyTorch, Pandas, NumPy, Scikit-Learn, Flask, FastAPI, HuggingFace (Transformers like BERT, GPT), Ray, Celery, RabbitMQ, Airflow, Kubeflow, Argo Workflows, MLflow, PySpark, AWS MWAA, AWS Glue, AWS ECS, Hive, Hadoop, Presto, 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), GCP (BigQuery, Dataflow, Pub/Sub, Cloud Storage, Cloud SQL, Cloud Spanner, Dataproc, AI Platform, Vertex AI, Cloud Composer, Cloud Data Fusion, Cloud Functions, Cloud Run, Kubernetes Engine, Firestore, Stackdriver, IAM).
- Environments and Tools: Jupyter Notebook, Visual Studio Code, Git, Docker, Terraform, Helm, Kustomize, ArgoCD.
- CI/CD & Build: Jenkins, Maven, GitHub Actions, GitHub Pipelines, CircleCI, Bitbucket Pipelines.
- UML Modeling: Use Case Diagram, Activity Diagram, Solution Architecture.
- Source Control: GitHub, Git, Bitbucket.
- IDEs: Visual Studio Code, PyCharm, Sublime Text.
- Databases: MySQL, PostgreSQL, MongoDB, Amazon Redshift, AWS DynamoDB, AWS RDS, MongoDB Atlas, Apache Iceberg.

Professional Experience

NCube *https://ncube.com/* **03.2025 - Present**

Position: Senior Data Engineer, Data Engineering Lead, Solutions Architect

Project: Data platform modernization and new data services for a European last-mile delivery and on-demand logistics platform. Re-platforming legacy ETL onto a Kubernetes-native, event-driven data stack on AWS with ArgoCD-managed GitOps deployments and a unified lakehouse for analytics, operations and ML use cases.

Responsibilities included:

- Designing and building net-new data services from scratch - ingestion connectors, streaming enrichment jobs, batch transformation pipelines, and analytics-ready datasets serving operations, finance, and ML teams.
- Architecting end-to-end data flows across the full AWS data ecosystem: S3 (data lake / Iceberg tables), Kinesis Data Streams and MSK (Kafka) for real-time event ingestion, Glue (catalog, crawlers, jobs), EMR and EMR Serverless for Spark workloads, Athena and Redshift for SQL analytics, Lambda and Step Functions for orchestration glue, DynamoDB for low-latency lookups, RDS Aurora and DocumentDB for OLTP, Lake Formation for fine-grained access, and SageMaker for downstream ML.
- Operating the entire data platform on Amazon EKS - packaging Spark jobs, Airflow workers, Argo Workflows, custom microservices and data APIs as Helm charts, with autoscaling via Karpenter and resource governance via namespaces and quotas.
- Implementing GitOps delivery with ArgoCD - every data service, DAG, dbt project, and infrastructure change is deployed declaratively from Git, with progressive sync waves, drift detection, and automated rollback on failed health checks.
- Building real-time streaming pipelines (Kafka to Flink / Spark Structured Streaming to Iceberg / OpenSearch / Redshift) for delivery tracking, ETA prediction features, fleet telemetry, and operations dashboards.
- Standing up a dbt + Airflow + Argo Workflows transformation layer with data contracts, schema enforcement, automated tests, and lineage tracking via OpenLineage / DataHub.

- Owning the Solutions Architecture for new data products - writing ADRs, leading design reviews with engineering, product and infrastructure stakeholders, and evaluating build-vs-buy trade-offs (Confluent vs. MSK, Snowflake vs. Redshift, Airflow vs. Argo Workflows, Glue vs. EMR Serverless).
- Embedding data quality and observability - Great Expectations for validation, Prometheus / Grafana / CloudWatch for runtime metrics, structured logging shipped to OpenSearch, and alerting via PagerDuty.
- Driving Infrastructure-as-Code adoption with Terraform (multi-account AWS landing zones, VPCs, IAM, KMS, EKS, networking) and Helm/Kustomize for Kubernetes resources.
- Securing the data platform - IAM least-privilege, IRSA for pod-level AWS access, KMS envelope encryption, VPC endpoints, secrets via External Secrets Operator backed by AWS Secrets Manager, and GDPR-aware PII handling and tokenization.
- Building CI/CD pipelines (GitHub Actions, Bitbucket Pipelines) covering linting, unit / integration / data tests, container builds, image scanning (Trivy), Helm chart rendering, and ArgoCD syncs across dev / staging / prod.
- Reviewing code, guiding mid-level engineers, and pair programming on complex work.

Environment: AWS (S3, Kinesis, MSK / Kafka, Glue, EMR, EMR Serverless, Athena, Redshift, Lake Formation, Lambda, Step Functions, DynamoDB, Aurora, DocumentDB, SageMaker, IAM, KMS, Secrets Manager, CloudWatch, EventBridge), Kubernetes (EKS, Karpenter, Helm, Kustomize), ArgoCD, Argo Workflows, Apache Airflow, Apache Spark / PySpark, Apache Flink, Apache Iceberg, Apache Kafka, dbt, Great Expectations, OpenLineage, DataHub, Terraform, Docker, GitHub Actions, Bitbucket Pipelines, Trivy, Prometheus, Grafana, OpenSearch, PagerDuty, Python, SQL, Scala.

Exadel *www.exadel.com* **09.2021 - 02.2025**

Position: Senior Data Engineer, Data Engineering Lead, Solutions Architect

Project: Automation software for gathering and governing data of clinical institutions. Amsterdam HQ.

Responsibilities included:

- Identifying business needs and creating model design.
- Creating project mapping and architecture; designing and implementing custom solutions covering specific needs of an enterprise.
- Reviewing code, guiding mid-level engineers, and pair programming on complex work.
- Configuring and maintaining infrastructure for data pipelines with Airflow, Argo Workflows, Spark, Kubernetes, Postgres, Object Storages.
- Set up automated CI/CD pipelines for microservices (Bitbucket, Jenkins).

Environment: AWS (CodeCommit, CodeBuild, AMI, EC2, VPC, S3, EFS, Lambda, Python SDK, Organizations, STS, SSM, DynamoDB, ECR, EKS, CloudWatch, CloudTrail, CloudFormation), Azure (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).

Position: Senior Data Engineer

Project: Full cycle of automated automotive repair (SaaS, on-site consulting, R&D; department).

Responsibilities included:

- Gathering requirements of the custom solutions, delivering proposed architectures constrained by the technical landscape of an organization (AWS, GCP, Azure).
- Implementing and leading development efforts of data transformation pipelines at scale (~50 TB of data monthly).
- Reviewing code, guiding mid-level engineers, and pair programming on complex work.
- Troubleshooting and adjusting pipelines / environments upon need. Took care of Helm / HULL deployments.
- Built CI/CD pipelines according to the established corporate standards.

Environment: PySpark, Airflow, Kubernetes, Argo Workflows, Hive, Hadoop, AWS EMR, AWS storages, AWS Kinesis, MWAA.

SoftServe **12.2019 - 08.2021**

Position: Machine Learning Operations and Data Engineer

Project: Singapore enterprise banking system. Development of DAGs and machine learning pipelines at scale.

Responsibilities included:

- Deploying ML/DL models at scale (cloud and on-prem) combined with distributed training / data gathering.
- Set up automated CI/CD pipelines for microservices and ML/DL models.
- Developing Kubeflow pipelines for continuous retraining and ML/DL model inference.
- Participating in designing cloud applications (creating Cloud services architecture).
- Reviewing code, guiding mid-level engineers, and pair programming on complex work.

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

Xperra **10.2018 - 08.2019**

Position: Machine Learning Engineer

Project: The cornerstone of the business is to provide insights out of a 25 years' worth database of physical mail pieces which US citizens get daily in mailboxes; with the help of deep and machine learning techniques the team was able to uncover patterns in successful mail piece design, text usage, target audience, etc.

Responsibilities included:

- Analysis of large amounts of text and handwritten data by embracing Natural Language Processing.
- Knowledge sharing, training.
- Requirements analysis and clarification.
- Performing Image Recognition of Direct Mail Pieces.
- Reviewing code, guiding mid-level engineers, and pair programming on complex work.
- Product support and documentation maintenance.

Environment: Python, NLTK, SpaCy, KNIME, TensorFlow, Keras.

Prosapient **07.2017 - 10.2018**

Position: Data Engineer

Project: Web-based platform to connect top-tier consultants with clients (companies).

Responsibilities included:

- Calculating Churn, Retention, CAC, LTV, User Behavior.
- Segmentation and clusterization of User Base.
- Developing a web platform.
- Designing and implementing data pipeline within existing infrastructure.
- Reviewing code, guiding mid-level engineers, and pair programming on complex work.
- Improving text recognition technology.

Environment: Django, Redis, MongoDB, Docker, Postgres, NLTK / SpaCy.

DataRobot **01.2017 - 07.2017**

Position: Data Analyst

Project: Machine Learning Automation Application - a fully automated machine learning application which follows leading standards in Data Science, Machine Learning and MLOps.

Responsibilities included:

- Cleaning and shaping data sets.
- Analyzing data and uncovering insights.
- Solving Classification and Regression problems (KNN, SVM, Naive Bayes, Custom ensemble models).
- Marketing Flows Automation.
- Reviewing code, guiding mid-level engineers, and pair programming on complex work.

Environment: Pandas, NumPy, R, Salesforce, Apex, Scikit-Learn.

Education

Zaporizhzhia National Technical University **2016**
Master's Degree (Computer Systems and Network Engineering).

Certifications

Associate Data Practitioner, GCP **2019**
L4-DE: Data Engineering in KNIME Analytics Platform **2018**
Professional Cloud Database Engineer, GCP **2021**