Best Technology Talent For
Companies Around The World

ncube

Dmytro

Senior Data Engineer

Summary

Senior Data Engineer with 5+ years of experience building large-scale batch data pipelines, cloud analytics platforms and cost-efficient storage architectures. Led migrations, modernized ETL workflows and optimized Spark workloads for high-volume business use cases. Strong focus on data quality, reproducibility, orchestration and IaC, partnering closely with product, ML and analytics teams to deliver reliable, insight-driven data systems.

TECHNICAL SKILLS:

- **Languages**: Python, SQL, Scala, Bash
- **Cloud Platforms**: AWS, GCP, Databricks
- **Technologies**: Apache Spark, PySpark, Apache Airflow, PostgreSQL, dbt, Apache Iceberg, Docker, Terraform, Git

Professional Career

Senior Data Engineer

April 2025 – Present | Ukraine

Project info:

Led data engineering initiatives for cloud cost-optimization and reporting, focusing on scalable data pipelines, storage optimization, and automated analytics delivery for business stakeholders. Built and optimized ETL flows on AWS, migrated historical data to low-cost storage, and implemented reporting pipelines.

Responsibilities:

- Migrated ~300+ GB of historical data from MongoDB to S3 Parquet, reducing long-term storage costs and ensuring rollback safety via quarantine retention and validation checks
- Defined Parquet partitioning strategy based on date ranges and access patterns, improving query efficiency and reducing read operations for reporting workflows
- Built ETL pipelines using AWS Glue (PySpark) to ingest and transform data from MongoDB, RDS, and S3; optimized Spark execution through caching, column pruning, and selective reads
- Implemented automated reporting engine supporting both on-demand and scheduled jobs, orchestrated via Step Functions and triggered through SQS/Lambda
- Reduced report generation latency through optimized joins, partitioning, and dataset filtering, improving execution times and lowering compute cost
- Set up centralized logging and monitoring via CloudWatch for Glue, Lambda, and Step Functions, enabling proactive failure detection
- Provisioned cloud infrastructure with Terraform (S3 buckets, Glue jobs, IAM, Lambda), collaborating with DevOps to integrate into shared IaC standards and promote to higher environments
- Leveraged S3 lifecycle policies to transition cold data to lower-cost tiers, contributing to long-term cost optimization.

Software and technologies:

Python, SQL, AWS, Terraform, PostgreSQL, MongoDB, Apache Spark, Apache Parquet.

Senior Data Engineer

May 2024 – March 2025 | **Ukraine**

Project info:

Retail Pricing & Analytics Platform.

Contributed to the migration of data workflows from BigQuery to Databricks to unify analytics and machine-learning pipelines under a single data platform. Modernized ETL processes, streamlined business logic, collaborated with stakeholders to validate transformations, and automated orchestration to reduce manual effort and latency.

Responsibilities:

- Re-implemented GCP BigQuery data pipelines on Databricks, porting SQL logic to Spark DataFrame API and restructuring transformations to align with Databricks architecture
- Consolidated transformations and removed legacy logic, reducing pipeline complexity
- Collaborated with data consumers (data scientists, ML engineers, analysts) to validate business logic changes and ensure alignment with downstream modeling and reporting needs

- Migrated and optimized workflows processing tens of GBs per pipeline (~50–100GB), achieving comparable or faster execution time vs BigQuery baselines
- Applied data quality checks (record counts, consistency, schema alignment) to validate legacy vs new pipeline outputs
- Built custom Airflow sensor to automatically detect DWH data readiness and trigger downstream pipelines, eliminating manual coordination and reducing processing delay
- Coordinated schema and transformation changes via structured review process, stakeholder syncs, and ticketing workflows, ensuring traceable and approved modifications.

Software and technologies:

Python, SQL, Apache Spark, Apache Airflow, GCP (BigQuery), Databricks, Docker.

Big Data Engineer

August 2021 – April 2024 | **Ukraine**

Project info:

Global Tech Client – Mobile Analytics.

Delivered large-scale batch data processing pipelines supporting analytics and data science for a high-traffic consumer platform in the mobile/entertainment domain. Focused on Spark-based transformations, orchestration modernization, metadata migration, and data quality enforcement to ensure reliable and timely data delivery across distributed teams.

Responsibilities:

- Built and scaled Spark pipelines processing datasets from several GB to 100+ GB, running daily and multiple times per day to meet analytical and ML data needs
- Implemented aggregation, join, and window-based transformations using both Spark SQL and DataFrame API, ensuring predictable performance at scale
- Optimized batch workloads via broadcast joins, caching strategies, and column-level pruning; applied bucketing for high-cardinality join keys, reducing shuffle overhead and improving performance by several-fold
- Migrated orchestration from Azkaban to Airflow, rewriting DAGs and validating outputs against legacy pipelines
- Applied internal data quality framework for schema checks, null/duplicate thresholds, and statistical validations
- Transferred metadata definitions from Hive Metastore to internal YAML-based configuration system and CSV-in-GitHub storage, enabling version control and PR-based approval workflow
- Wrote scripts for master-data table migration, ensuring schema compatibility and accurate historical transfer

- Partnered with analysts, data scientists, and engineers to align transformations with evolving analytics and downstream needs
- Improved data readiness SLAs by reducing Spark job execution time and accelerating pipeline rollout through Airflow adoption.

Software and technologies:

Python, Scala, SQL, Apache Spark, Apache Iceberg, Apache Airflow, AWS, Docker, Kubernetes.

Python Developer

July 2019 – May 2021 | **Ukraine**

Project info:

Freelance project.

Responsibilities:

- Delivered backend services and automation tools for small business clients, collaborating with a distributed freelance team.
- Focused on Python-based APIs, data processing scripts and deployment of lightweight web applications.

Software and technologies:

Python, Django, Flask, PostgreSQL, AWS, Pandas, OpenCV, scikit-learn, PyTorch, Docker.

Software Engineer Intern

July 2018 – August 2018 | **Ukraine**

Project info:

Internal R&D Project.

Responsibilities:

- Completed an engineering internship focused on full-stack development and cloud API integration
- Built an internal prototype system that analyzed user emotions via computer vision and generated personalized music suggestions, gaining hands-on experience in backend development, external API usage, and end-to-end application flow.

Software and technologies:

Python, Django, PostgreSQL, Microsoft Azure (Face API).

Education

2018 – 2022

Ivan Franko National University of Lviv

Bachelor's degree, Computer Science

Certifications

- Databricks Certified Data Engineer Associate 2025
- Astronomer Certification for Apache Airflow Fundamentals 2024
- Amazon Web Services Cloud Practitioner 2023.

Foreign Languages Skills

English: Upper-Intermediate

Ukrainian: Native.