DMITRII CHERKASHIN

Data engineer · Montevideo, Uruguay · remote · open to consulting

mit.cherkashin@gmail.com linkedin.com/in/dimstunt github.com/dimstunt

dimstunt.de

ABOUT

Data engineer with nine years of experience building data platforms — from Data Vault warehouses in banking and ride-hailing to an AWS lakehouse in health-tech. I design architectures, build CDC and batch pipelines, and own platforms end to end: reliability, data quality, access control and cost. I have hired and led a data team and mentored engineers. Degree in Mathematical Physics from Lomonosov Moscow State University.

EXPERIENCE

WELLTORY — Data Engineer — Remote · 01/2025 – now

ACHIEVEMENTS

- Forked and maintain an abandoned open-source Kafka Connect DynamoDB Streams connector in production (12 tables, 300+ shards); cut its checkpoint write load **~5.7×** with zero throttling
- Scaled the CDC pipeline through an **80×** surge in data volume after a major product launch
- Retired the legacy data lake after migrating its history into Iceberg

RESPONSIBILITIES

- Own the company's data platform end to end: availability, quality, access speed and cost of data for analytics and health calculations
- Designed and built a data lakehouse on AWS from scratch: S3 + Apache Iceberg, Glue, Athena, Airflow
- CDC replication from Postgres and DynamoDB via Kafka into the lakehouse
- Data catalog and lineage in OpenMetadata; column-level access control with Lake Formation tags (PII/PHI); GDPR export and erasure pipelines

US MUSIC PUBLISHER — Data Engineer (contract) — Remote · 2024

- Replication from IBM AS/400 into ClickHouse with data-quality reconciliation and recovery tooling

ANABAR — Senior Data Engineer — Remote · 02/2023 – 06/2024

ACHIEVEMENTS

- Reduced data processing time from **17 to 7 hours**, raising SLA compliance from **70% to 95%**
- Cut ClickHouse storage usage **by 40%**
- Improved p90 query response time from **20 seconds to under 1 second**
- Mentored engineers and gave internal tech talks on data engineering

RESPONSIBILITIES

- Designed and implemented data processing pipelines with Airflow, Trino and ClickHouse in Yandex Cloud
- Took part in the migration from a monolith to FastAPI microservices

AGRO.CLUB — Lead Data Engineer — Moscow · 09/2021 – 11/2022

ACHIEVEMENTS

- Accelerated business report generation from **1.5 hours to 1–2 minutes**

RESPONSIBILITIES

- Hired and managed the DWH team from scratch; set the strategy, processes and deadlines for the DWH division
- Designed the Data Warehouse architecture with Data Vault
- Built ETL pipelines from MongoDB, Postgres and APIs into the warehouse with Airflow
- Built data synchronization between partner ERP systems and the PostgreSQL backend, keeping partner data consistent
- Introduced data quality checks and data governance

CITYMOBIL Senior Data Engineer — Moscow · 06/2020 – 09/2021

ACHIEVEMENTS

- Designed self-service replication for “one-click” table onboarding

RESPONSIBILITIES

- Built ETL pipelines from MySQL, ClickHouse, Kafka and APIs into the warehouse with Airflow
- Implemented online CDC replication MySQL → Kafka → Exasol
- Designed and implemented the Data Vault DDS loader with Airflow DAGs and operators

MAIL.RU GROUP Senior System Analyst DWH — Moscow · 11/2018 – 06/2020

- Created the DWH data model for Delivery Club and the logical data model for Citymobil from scratch
- Built e-commerce user portrait reports in Tableau; supported Spark processing scripts
- Reduced time from task creation to production reporting from **3 weeks to 1 week**

TINKOFF BANK System Analyst DWH — Moscow · 10/2017 – 11/2018

- Designed the integration receiving checks from the national payment card system
- Designed and prototyped Data Vault data models and data marts in SAP PowerDesigner
- Worked with Greenplum, Vertica, SAS DIS, Informatica PC and SAP BO; wrote BRDs and FSDs

SKILLS

LANGUAGES	Python, SQL, Scala
LAKEHOUSE & DWH	Apache Iceberg, AWS (S3, Glue, Athena, Lake Formation), ClickHouse, Trino, Greenplum, Vertica, Exasol, Data Vault
DATABASES	PostgreSQL, MySQL, MongoDB, DynamoDB
STREAMING & CDC	Kafka, Kafka Connect, Debezium, KSQL, Faust, NiFi
PROCESSING	Airflow, Spark, dbt, Hive
GOVERNANCE	OpenMetadata, data modeling, data quality, GDPR
BI	Metabase, Superset, Tableau, Redash
INFRA	Docker, Kubernetes, Terraform, Grafana, Sentry, GitLab CI

EDUCATION

Lomonosov Moscow State University · Bachelor, 2018
Mathematical Physics, Faculty of Computational Mathematics and Cybernetics

LANGUAGES

Russian — native · English — B2, professional working

INTERESTS

Running · Climbing · Snowboarding · Surfing