

---

# Illia V.

## DATA ENGINEER

### Education

Computer Science and Software Engineering

### Language proficiency

English – C1

Polish - B2

### Domains

Banking

Finance

### Data Engineer with 6+ years of experience.

Data Engineer with 6+ years of experience specializing in on-premises and cloud financial data platforms, open-source stack integration, and enterprise data engineering. Proven expertise in designing self-hosted data platforms from scratch, building distributed PySpark transformation pipelines, and orchestrating complex workflows via Apache Airflow. Strong background in modern Lakehouse architecture with Apache Iceberg, SQL-based transformations using dbt, and federated querying via Trino. Academic background in Computer Science with a consistent track record in regulated banking and fintech environments.

### Programming languages

Python, SQL.

### Data Engineering

Apache Spark, PySpark, dbt, Apache Airflow, Apache Iceberg, OpenMetadata.

### Cloud

AWS (S3, Glue, Lambda, IAM, SNS, RDS, Athena, Redshift, SQS, Secrets Manager, etc.).

### Databases

PostgreSQL, Trino, TimescaleDB.

### BI and Data Visualization

Power BI, OpenSearch, Apache Superset.

### Data Science

Pandas, NumPy.

### Data Modeling

Medallion Architecture, Dimensional Modeling, Entity-Relationship (ER) Modeling, Normalization /

Denormalization, Data Vault, Data Mesh.

### DevOps

Docker, Docker Compose, Bash scripting, Kubernetes, Terraform, GitHub Actions, Prometheus, Grafana.

### AI Tools

Gemini, Claude.

### Version control systems

Git, GitHub, GitLab.

## Projects

### BANKING INFRASTRUCTURE DATA PLATFORM

A self-hosted data platform built from scratch for a regulated financial institution, replacing managed cloud dependencies with a vendor-independent open-source stack. The system handles ingestion, transformation, and analytics across sensitive banking data with full governance, lineage tracking, and on-premises infrastructure orchestration.

### Project roles

Data Engineer

### Period

07.2023 - till now

### Responsibilities

- Deployed self-hosted Airflow on Kubernetes orchestrating complex DAG dependency graphs;
- Designed dynamic Airflow DAGs with custom retry policies and SLA monitoring;
- Developed custom Airflow operators and sensors for domain-specific banking workflows;
- Built PySpark pipelines on Kubernetes with partition tuning and skew handling;
- Implemented Iceberg table compaction, snapshot expiry, and hidden partitioning strategies;
- Designed dbt incremental models, macros, and custom generic tests for transformation layer;
- Configured Trino federated queries across Iceberg tables and PostgreSQL sources;
- Tuned Trino resource groups and connector configuration for query performance optimization;
- Designed PostgreSQL schemas with TimescaleDB

## FINANCIAL OPERATIONS MANAGEMENT PLATFORM

A financial operations management platform built to process and analyze massive digital service transactional volumes. The automated system orchestrates financial ledger records calculation and generates dynamic analytics data assets for corporate trend

- hypertables for time-series banking data;
- Provisioned self-hosted platform components via Terraform with modular infrastructure-as-code;
- Deployed platform components via ArgoCD with Helm charts and Kubernetes StatefulSets;
- Built GitLab CI/CD pipelines automating build, test, and environment promotion workflows;
- Integrated OpenMetadata for lineage tracking and governance across transformation layers;
- Configured OpenSearch indexes for log analytics and pipeline observability across infrastructure;
- Set up Prometheus and Grafana dashboards monitoring cluster health and pipeline SLAs;
- Exposed analytical datasets to business users through self-hosted Apache Superset dashboards;
- Enforced column-level access controls and audit logging across sensitive financial datasets.

### Environment

Python, SQL, Apache Spark, PySpark, Apache Airflow, dbt, Apache Iceberg, Trino, PostgreSQL, TimescaleDB, Pandas, NumPy, OpenMetadata, OpenSearch, Prometheus, Grafana, Apache Superset, Docker, Docker Compose, Kubernetes, Terraform, GitHub Actions, Bash scripting, Gemini, Claude, GitHub.

### Project roles

Data Engineer

### Period

07.2020 - 06.2023

### Responsibilities

- Designed Python ingestion scripts handling large transactional volumes from external APIs;
- Developed PySpark batch jobs aggregating financial transaction logs across staging environments;

reporting.

- Programmed Airflow workflows scheduling data validation routines and processing alerts;
- Configured AWS S3 data lakes with lifecycle rules managing corporate historical files;
- Applied AWS Glue catalogs managing metadata assets and automating data structure discovery;
- Deployed AWS Lambda microservices processing transactional inputs from external system endpoints;
- Managed cloud security policies via AWS IAM safeguarding sensitive corporate credentials;
- Distributed pipeline failure notifications to engineering teams using AWS SNS topics;
- Refactored legacy schemas into dimensional models inside AWS RDS improving reporting;
- Executed ad hoc queries via AWS Athena auditing historical tables in cold storage;
- Architected data warehouses within AWS Redshift centralizing ledger records for analytics;
- Designed Power BI dashboards visualizing financial trends and corporate reporting metrics.

**Environment**

Python, SQL, Apache Spark, PySpark, Apache Airflow, AWS (S3, Glue, Lambda, IAM, SNS, RDS, Athena, Redshift, SQS, Secrets Manager, etc.), Power BI, Pandas, NumPy, PostgreSQL, Docker, Docker Compose, Bash Scripting, GitLab.

Professional skills

| SKILLS                |        | EXPERIENCE IN YEARS | LAST USED |
|-----------------------|--------|---------------------|-----------|
| Programming languages | Python | 6                   | 2026      |

| SKILLS                    |                 | EXPERIENCE IN YEARS | LAST USED |
|---------------------------|-----------------|---------------------|-----------|
| Data Engineering          | SQL             | 6                   | 2026      |
|                           | Apache Spark    | 6                   | 2026      |
|                           | PySpark         | 6                   | 2026      |
|                           | Apache Airflow  | 6                   | 2026      |
|                           | dbt             | 3                   | 2026      |
|                           | Apache Iceberg  | 3                   | 2026      |
|                           | OpenMetadata    | 3                   | 2026      |
| Clouds                    | AWS             | 3                   | 2023      |
| Databases                 | PostgreSQL      | 6                   | 2026      |
|                           | TimescaleDB     | 3                   | 2026      |
|                           | Trino           | 3                   | 2026      |
| BI and Data Visualization | Power BI        | 3                   | 2023      |
|                           | Apache Superset | 3                   | 2026      |
|                           | OpenSearch      | 3                   | 2026      |
| DevOps                    | Docker          | 6                   | 2026      |
|                           | Docker Compose  | 6                   | 2026      |
|                           | Kubernetes      | 3                   | 2026      |
|                           | Bash scripting  | 6                   | 2026      |
|                           | Terraform       | 3                   | 2026      |
|                           | GitHub Actions  | 3                   | 2026      |
|                           | Prometheus      | 3                   | 2026      |
|                           | Grafana         | 3                   | 2026      |
| Data Science              | Pandas          | 6                   | 2026      |

| SKILLS                  |        | EXPERIENCE IN YEARS | LAST USED |
|-------------------------|--------|---------------------|-----------|
|                         | NumPy  | 6                   | 2026      |
|                         | Gemini | 2                   | 2026      |
|                         | Claude | 1                   | 2026      |
| Version control systems | Git    | 6                   | 2026      |
|                         | GitHub | 3                   | 2026      |
|                         | GitLab | 3                   | 2023      |