

SAEID SHAHRIARI

Data Engineer

Brussels, Belgium | saeid.shahriari@outlook.com | [LinkedIn](#) | [GitHub](#) | [Website](#)

PROFESSIONAL SUMMARY

Junior Data Engineer with hands-on experience designing, building, and operating batch and real-time data pipelines using Python and SQL. Proven ability to ensure data quality, automate workflows, and monitor pipeline reliability with Grafana. Built production-style data integrations with Apache Airflow, Kafka, Flink, and dbt. Experienced working cross-functionally with analysts and stakeholders to translate business requirements into scalable data solutions. Familiar with Power BI reporting and eager to expand expertise in Databricks and AWS-native data services.

TECHNICAL SKILLS

Programming	Python (Advanced Scripting, Pandas, PySpark, PyFlink), SQL (PostgreSQL, ClickHouse, Snowflake), Java, Bash
Data Quality & Transformation	dbt (tests, schema validation, staging/marts), data quality frameworks, idempotency patterns, schema enforcement
Orchestration & Automation	Apache Airflow (DAG design, retries, logging), Docker, Kubernetes, Terraform, GitLab CI/CD, Jenkins
Streaming & Processing	Apache Kafka, Apache Flink (SQL/Table API), Spark Structured Streaming, Debezium (CDC)
Cloud & Infrastructure	GCP (Cloud Run, BigQuery), AWS (Glue, Athena, Lambda, S3)
Datastores	PostgreSQL, Snowflake, ClickHouse, Redis, OpenSearch
Monitoring & Reporting	Grafana (dashboards & alerts), Power BI, Tableau
Learning	Databricks (Community Edition), AWS (deepening expertise)

WORK EXPERIENCE

Data Visualization / BI Automation – Tableau Governance | Orange Business

Brussels, Belgium · Jul 2025 – Aug 2025

- Built Python automation jobs on Google Cloud Run to audit Tableau Server content, enforce governance rules, and produce admin reports — ensuring data accuracy and access integrity across dashboards.
- Automated deployment pipelines with Terraform and GitLab CI/CD, reducing manual operational steps by ~40% and standardizing releases.
- Collaborated with BI stakeholders to validate dashboard data quality, troubleshoot performance issues, and document resolution procedures.
- Gathered and translated business requirements into data checks and quality metrics, ensuring pipeline outputs matched reporting needs.

Data Engineering Trainee | Data Engineering Masterclass

Remote · Nov 2024 – Dec 2025

- Designed and built end-to-end Apache Airflow DAGs to ingest data from APIs, CSV, and JSON sources; implemented retries, logging, and idempotency patterns to ensure reliable and repeatable pipeline execution.
- Improved data quality and pipeline reliability by validating schemas, monitoring pipeline behavior, and optimizing SQL transformations when data issues appeared.
- Contributed to real-time streaming pipelines using Kafka with Spark and Flink; assisted with debugging lag, schema mismatches, and consumer offset issues.

- Built Grafana dashboards and alert rules to monitor pipeline health, data quality metrics, and system lag — enabling faster incident triage and root-cause analysis.
- Used ClickHouse for high-speed analytical queries and optimized SQL transformations for large-scale datasets.

KEY PROJECTS

Modern ELT Pipeline — Snowflake, dbt & Airflow [GitHub](#) ■

Tech Stack: Snowflake · dbt · Apache Airflow · Docker · Python

- Built dbt models across staging → intermediate → marts layers to produce clean, reliable analytics tables with clearly defined data contracts.
- Enforced data quality with dbt schema tests (relationships, not-null, uniqueness) and generated documentation for maintainability.
- Scheduled and orchestrated dbt runs and tests with Airflow, automating the full transformation pipeline with built-in quality gates.

End-to-End Real-Time Food Delivery Platform [GitHub](#) ■

Tech Stack: Python · PostgreSQL · Apache Kafka · Apache Flink SQL · Redis · OpenSearch · Docker

- Designed and containerized a full real-time pipeline: PostgreSQL → Debezium (CDC) → Kafka → Flink SQL → Redis/OpenSearch.
- Implemented stateful Flink SQL aggregations to compute live KPIs: active orders, average delivery time, and cancellation rate.
- Added verification scripts and troubleshooting runbooks to reproduce failures, validate outputs, and speed up incident resolution.

Streaming Fraud Detection Pipeline [GitHub](#) ■

Tech Stack: Python · PostgreSQL · Kafka · Spark · Airflow · Grafana · Docker

- Implemented streaming ingestion (Kafka → Spark → PostgreSQL) with idempotent writes and replay-friendly processing for safe incident recovery.
- Orchestrated monitoring workflows using Airflow + Grafana to track pipeline lag, throughput, and data-quality signals with automated alerts.
- Performed root-cause analysis on pipeline failures by tracing logs, validating SQL logic, and documenting fixes to prevent recurrence.

EDUCATION

M.Sc. in Applied Computer Science | Vrije Universiteit Brussel (VUB)

Brussels, Belgium · 2023 – Present

Focus: Data Engineering, Machine Learning, Big Data, Distributed Systems. Key courses: Big Data Processing, AI Techniques, Advanced Databases, Cloud & Distributed Systems.