

Case Study 1: PySpark Data Pipeline Modernization

Client

Data Engineering Project (Freelance)

Challenge

The client needed to migrate legacy Ab Initio ETL logic into a modern, scalable solution while integrating with SQL Server and orchestration tools.

Solution

- Recreated ETL logic in PySpark (main.py).
- Built data ingestion & transformation workflow via JDBC → PySpark → SQL Server.
- Packaged solution with Docker (bundled dependencies + SQL driver).
- Deployed on OpenShift via Spark Operator + YAML configs.
- Orchestrated jobs with Apache Airflow (KubernetesPodOperator) for scheduling & automation.

Results

- Migration from legacy Ab Initio achieved successfully.
- Automated pipeline reduced manual interventions.
- Scalable, containerized architecture enabled faster deployment cycles.

Growtech Digital Solutions

DevOps as a Service — Faster Deployments · Scalable Infrastructure · Lower Cloud Costs

✉ growtechdigitalsolutions@gmail.com

🔗 LinkedIn: linkedin.com/company/growtech-digitsolutions

🌐 Website: <https://growtech25.github.io/growtech-site/>