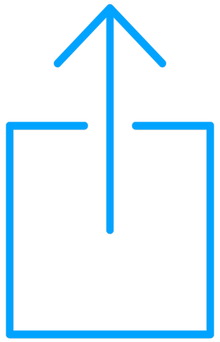
Build a Data Pipeline with Apache Airflow

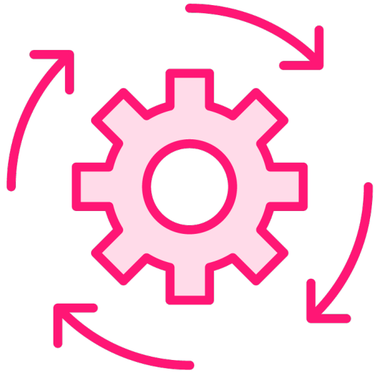
Julian Mateu
Software Engineer

@julian_mateu | www.jmateu.co.uk

ETL Pipelines



Extract



Transform

Load

Apache Airflow: Main Features

Monitor and retry

Scheduling

Scale

Python code

DAGs

Community

Summary

DAGs and tasks in an ETL pipeline

Retries, backfills, and monitoring with Airflow web UI

Do not use XComs to transfer datasets between tasks.