



Using Apache Airflow in Regulated Environments

Christian Förnges

3.0

Learn from Deutsche Bank: Using Apache Airflow in Regulated Environments

Operating within the stringent regulatory landscape of Corporate Banking, Deutsche Bank relies heavily on robust data orchestration. This session explores how Deutsche Bank's Corporate Bank leverages Apache Airflow across diverse environments, including both on-premises infrastructure and cloud platforms. Discover their approach to managing critical data & analytics workflows, encompassing areas like regulatory reporting, data integration and complex data processing pipelines. Gain insights into the architectural patterns and operational best practices employed to ensure compliance, security, and scalability when running Airflow at scale in a highly regulated, hybrid setting.

What is the environment we operate in?

- Biggest Euro clearing bank
- Focused on corporations and financial institutes
- 35+ countries with different regulatory requirements
- 50 payment systems from checks to instant payments
- 2 Million accounts
- 50 Million payments daily

Our regulatory challenges

- Completeness, Accuracy, Timeliness, ...
- Must be able to prove all the above
- Local requirements
 - Data must be stored in the country
 - Data can not be stored on public cloud
 - Retention policies
 - Access policies
 - Many more

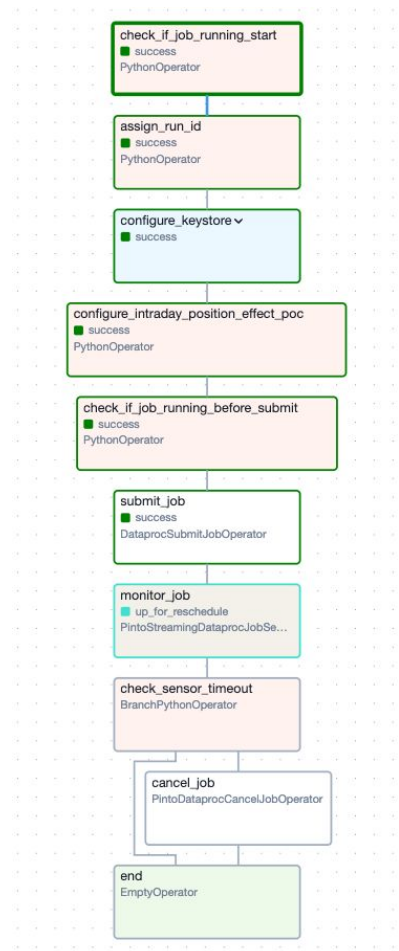
How do we use Airflow to address those challenges?

- Single Interface to all data pipelines
 - Batch
 - Streaming
- 4-eye support access
- Runner for odd jobs
 - Manage our BigQuery database schemas
 - Run a Spark job in test
 - Create a search index on a 250Tb table in production



Intraday Pipeline

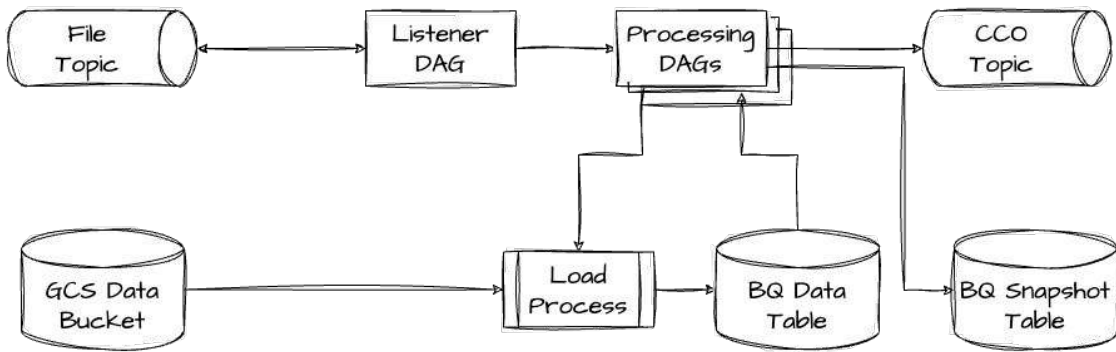
Maintaining the continuous operation
of our intraday pipeline





EOD Loading

Use AirFlow Run IDs to allow exactly once load



Questions?

christian.foernges@db.com