APACHE AIRFLOW FULL MIGRATION

Optimove

INTRODUCTION

The project was done for Optimove's Data Engineering group. Optimove is a leading AI-driven CRM platform specializing in optimizing marketing efforts. Optimove relies heavily on Apache Airflow for it's business operations, Airflow is an open-source platform used for orchestrating complex workflows and data pipelines. The current insfactrutre was built on top of version 1 of Airflow within Google Cloud Platform, specifically on Composer 1.
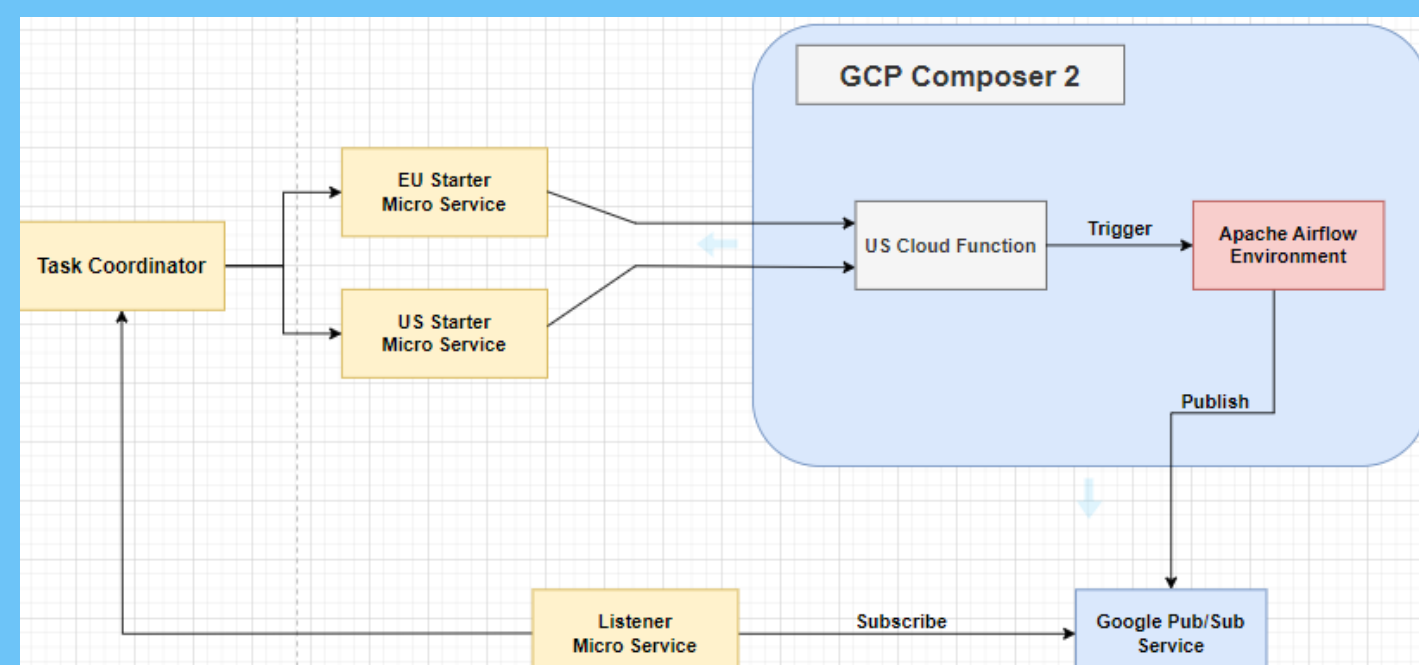
MOTIVATION

The motivation behind this project stems from the desire to enhance four key factors crucial to Optimove's data orchestration infrastructure:

- Performance Optimization- Airflow 2 brings architectural improvements for more efficient task execution, aiming to reduce overall processing times and enhance workflow responsiveness.
- Advanced Functionality- With new features and enhanced UI in Airflow 2, Optimove can better manage and monitor workflows, leading to streamlined operations and improved UX
- Long Term Support - Migration to Airflow 2 & Composer 2 ensures access to ongoing maintanencee and reducing vulnerabilities

METHODOLOGY

Our methodology commenced with a focus on infrastructure development, which included creating a new Airlfow environment and configuring a new GCP Composer 2 environment. Additionally, we undertook the task of adjusting all existing DAGs to ensure compatibility with Airflow 2. Furthermore, we ensured that Optimove's internal scheduler was capable of triggering the correct DAGs within the new Airflow environment and Composer seutp.

WORKFLOW

- Setting the foundation - We created a new Airflow 2 environment using Docker Compose, initialized with essential services and configurations using shell scripts.
- First migration & adjustments - Existing DAGs were analyzed and adjusted accordingly to Airflow 2 syntax, and were migrated to the new environment.
- Ensure Reliability - Extensive QA testing validated code functionally and addressed edge cases, ensuring robustness in production scenarios
- Full automation & Cloud migration - Automation scripts for migration were deployed to modify DAGs for Airflow 2 compatibility, while Composer 2 was set up on GCP. Furthermore, CI/CD pipeliens facilitated DAG deployment, and Pub/Sub enabled real-time communiaction

CONCLUSION

The project has significantly improved Optimove's workflow management by transitioning to Airflow 2 along with GCP Composer 2. These changes have enhanced effciency, reaalibility and adaptablilty to evolving business needs.

TECHNOLOGIES