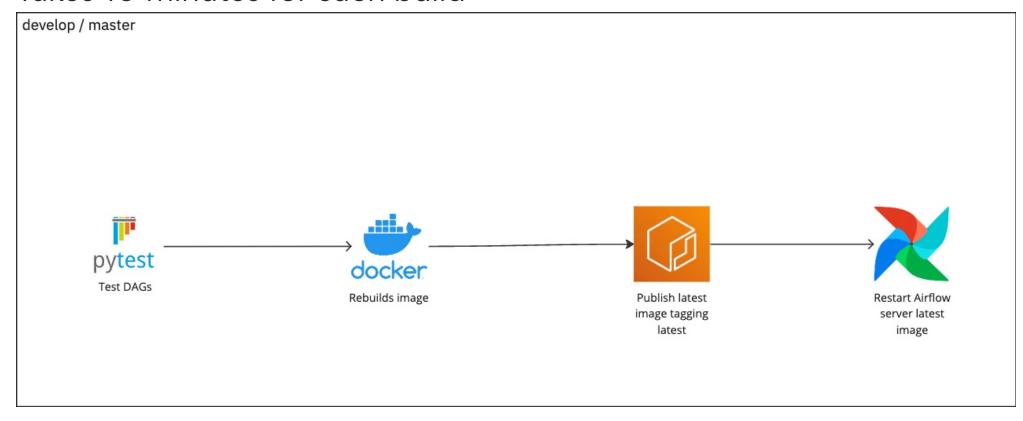


Agenda

- # CI/CD
 - Present
 - Planned
- Future of Airflow
- ? Questions

** Present CI/CD

- Builds a new image for any changes
- Publishes this image ~500 MB each time on updated changes
- Takes 10 minutes for each build

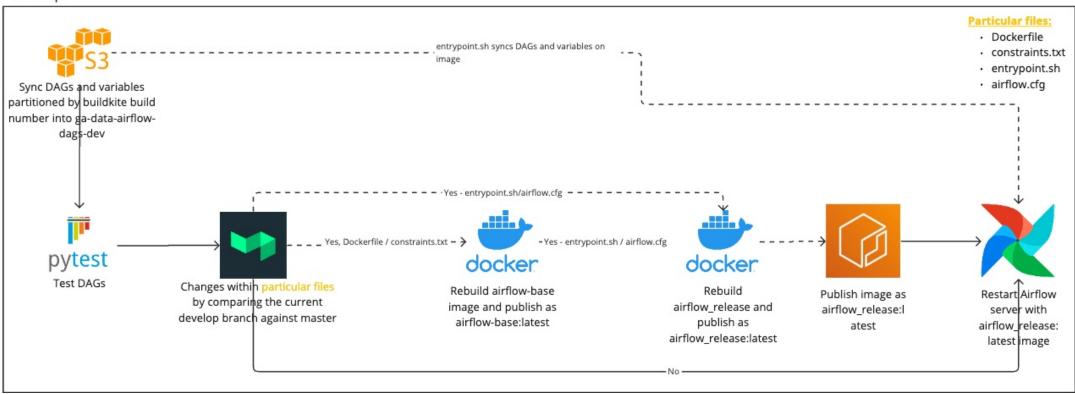


© Planned CI/CD

- Goals:
 - Reduce the time to build by 50%
- Achieved by:
 - i. Sync DAGs / variables into ga-data-airflow-dags-{env} S3 bucket.
 - ii. Splitting the Dockerfile into a multi stage build
 - airflow_base : for requirements
 - airflow_release: for entrypoint.sh / airflow.cfg changes
 - iii. Update entrypoint.sh using aws s3 sync to sync DAGs back
- Other changes:
 - Using SEEK's open source software plugins (seek-oss/docker-ecr-publish)

@ Planned CI/CD - develop

develop



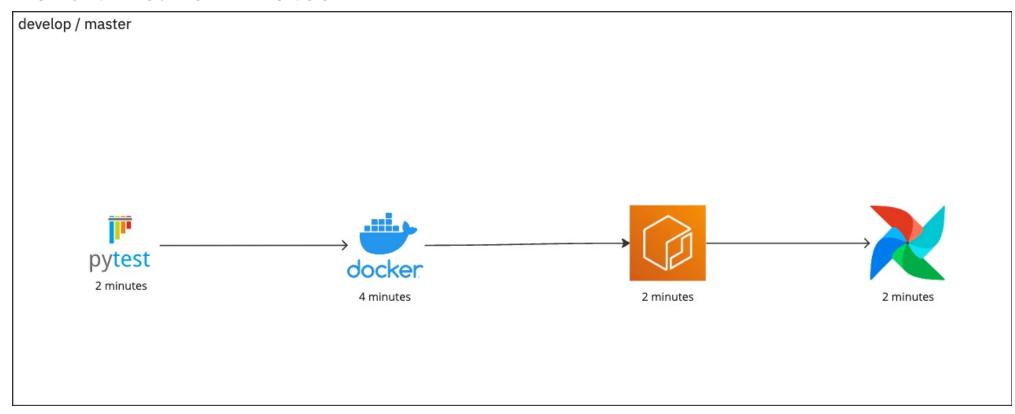
Planned CI/CD - master

Difference between develop and master is that each PR is a squash and merge.

master Particular files: Dockerfile entrypoint.sh syncs DAGs and variables on image · constraints.txt entrypoint.sh · airflow.cfg Sync DAGs and variables partitioned by buildkite build number into ga-data-airflowdags-prod - Yes - entrypoint.sh / airflow.cfg Yes, Dockerfile / constraints.txt - > - Yes - entrypoint.sh / airflow.cfg · -> pytest docker docker Test DAGs Changes within particular files Rebuild airflow-base Rebuild Restart Airflow Publish image as by comparing the current image and publish as airflow_release and airflow release:l server with commit and previous commit airflow-base:latest publish as airflow release: atest airflow release:latest latest image Assumes squash and merge PR

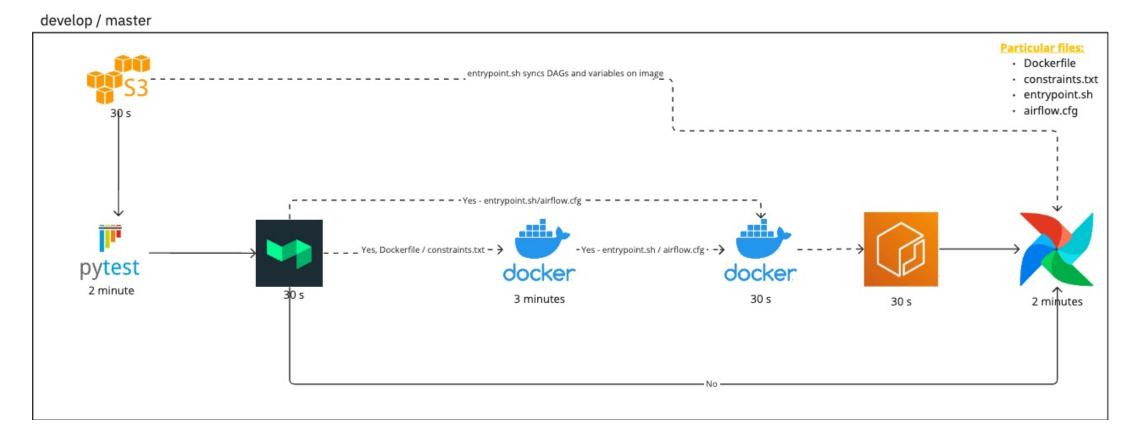
Timing differences - Present CI/CD

• Build time: 10 minutes



Timing differences - Planned CI/CD

- Build times:
 - 5 minutes for DAG / variable changes (most of the time ~95%)
 - 10 minutes for rebuilding image



? Questions