

# Implementation details - Model Pipelines

## Introduction

This page contains implementation details for "Cashflow2" MVP.

Currently, the DD Workspace is set up in the TEST environment. So the flow that was agreed upon was **Test DD Workspace → TST OPS (pre-prod) → QA OPS (prod)**

## What configurations were done in order to deploy the model?

### Build agent configuration

1. Configure the build agent with different Azure Databricks connection profiles using the service user's PAT token.
2. These profiles are passed as a secret variable in the pipelines.

### Data access configurations

1. Configure SAS Token (for now) as a variable to access the data from ADLS. This ADLS is in the DD workspace. This is currently okay (agreed with the architects) as we don't have Immuta and its policies in place.
2. SAS token configuration won't be needed later as we will configure the cluster with "Credential Pass-through".

### CMR configurations

1. It was agreed that the model registration will happen in CMR (Central Model Repository). So this was configured using PAT token as below:
  - a. TST DD → TST CMR
  - b. TST OPS → TST CMR
  - c. QA OPS → TST CMR

## Where can I find the repository and pipeline details?

A repository was created in Azure Repo named - **cashflow2-mvp**

The pipelines were created as YAML files and were placed as part of the repository.

*Naming convention followed for pipelines*

`<mvp-name>-<type>-pipeline`

Pipeline Type	Name	Purpose
<i>PR pipeline</i>	cashflow2-pr-pipeline	Triggered as soon as a PR is raised against master branch
<i>CI pipeline</i>	cashflow2-ci-pipeline	Triggered as soon as the PR is merged to master
<i>Model Training Pipeline</i>	cashflow2-train-evaluate-pipeline	Triggered automatically after CI pipeline execution is done
<i>Model Release Pipeline</i>	cashflow2-release-pipeline	Triggered automatically after training pipeline execution is done

More details can be found in the table below:

<b>Azure Repo</b>	<a href="https://dev.azure.com/GBI-ODL/AnalyticsOps/_git/cashflow2-mvp">https://dev.azure.com/GBI-ODL/AnalyticsOps/_git/cashflow2-mvp</a>
<b>Azure Pipeline</b>	<a href="https://dev.azure.com/GBI-ODL/AnalyticsOps/_build?definitionScope=%5Ccashflow2">https://dev.azure.com/GBI-ODL/AnalyticsOps/_build?definitionScope=%5Ccashflow2</a>

## Next Steps

1. As soon as PROD DD Workspace is in place, the model should be migrated there.
2. CMR configurations should be done as well because a new CMR instance in PROD will be created.
3. Find out if there is a better way of configuring the Databricks connection profile because there will be many connections as we grow