Table of Contents

[Use Case: 1](#_gjdgxs)

[High Level Steps 1](#_30j0zll)

[Prerequisite: 1](#_1fob9te)

[Jenkins related: 1](#_3znysh7)

[Bitbucket related: 1](#_2et92p0)

[Script related: 1](#_tyjcwt)

[Artifactory related 1](#_3dy6vkm)

[Dataiku related 1](#_1t3h5sf)

[Dataiku Project Details 2](#_4d34og8)

[Dataset Information: 2](#_2s8eyo1)

[Git Setup: 2](#_17dp8vu)

[Steps involved in solving the use cases 4](#_3rdcrjn)

[Useful Links: 12](#_26in1rg)

# Scope

This document showcases the steps to deploy recipe or any other file to the Dataiku automation node.

The Dataiku project bundle should use the Dataiku Design Studio to create the project bundle so that deployer can be used to deploy/re-deploy the bundle again on the Dataiku automation node.

# Use Case:

Deploy the code recipe/python notebooks changes done in design node to the automation node by only updating the changes without bundling the whole project and deploying the same project using the automation node.

# High Level Steps

**Steps involved to accomplish are as follows:**

Step 1: Performing changes in the recipe file and Python Notebook

Step 2: Pushing the changes to bit bucket.

Step 3: Trigger the pipeline for repo changes.

Step 4: Pipeline will run and is responsible for package and publishing the changes to

Artifactory.

Step 5: Fetch the updated files as .tar file, extract and place it in Dataiku.

Step 6: Run the scenarios and look for changes in automation node.

# Prerequisite:

## Jenkins related:

1. Jenkins Slave Node needs to include Python library ‘requests’,

## Bitbucket related:

1. Create Access Token from Dataiku to Bitbucket so that recipe files can be pushed from Dataiku Git repository to Bitbucket repository,

## Script related:

1. Authenticating the Artifactory repository using Auth code in the Python script.

## Artifactory related

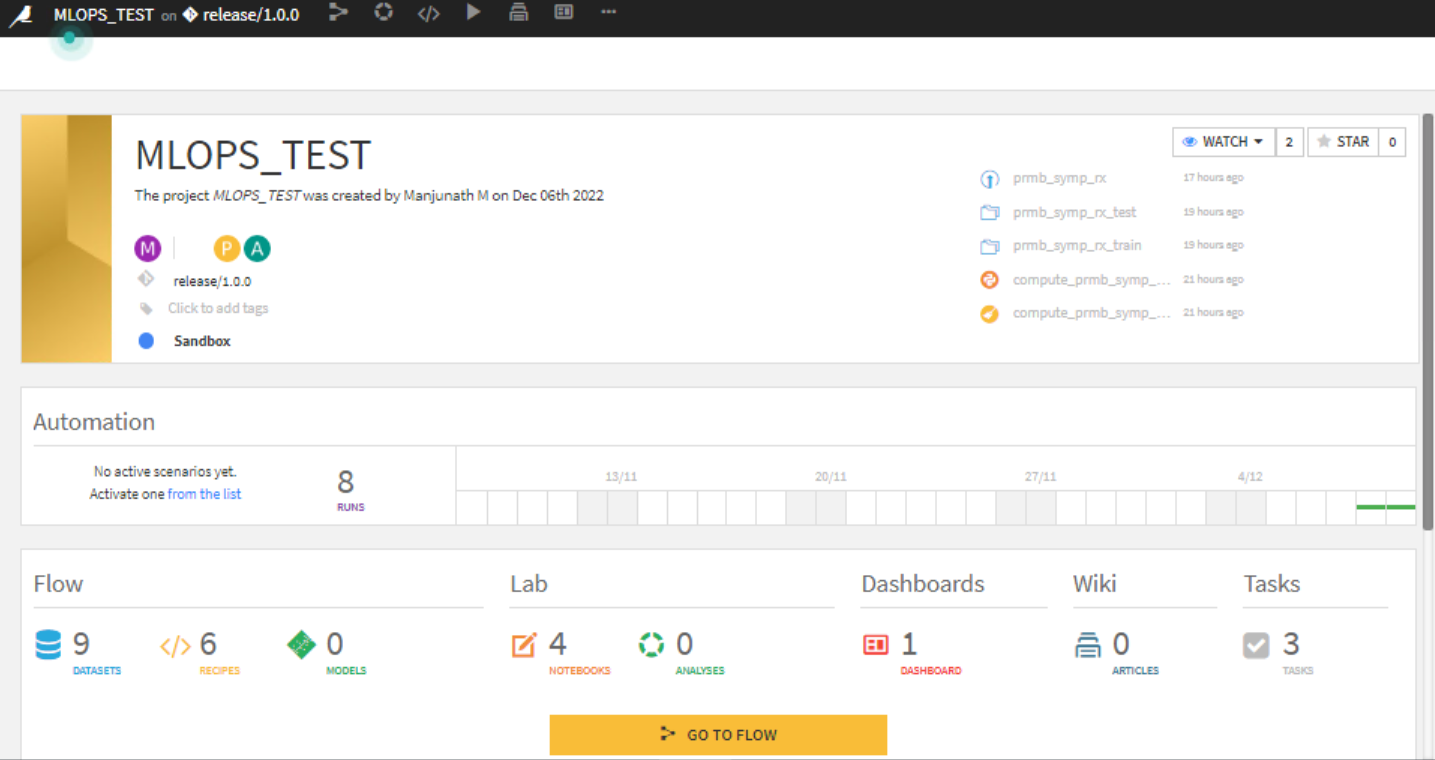
1. Creating the auth token in Artifactory.

## Dataiku related

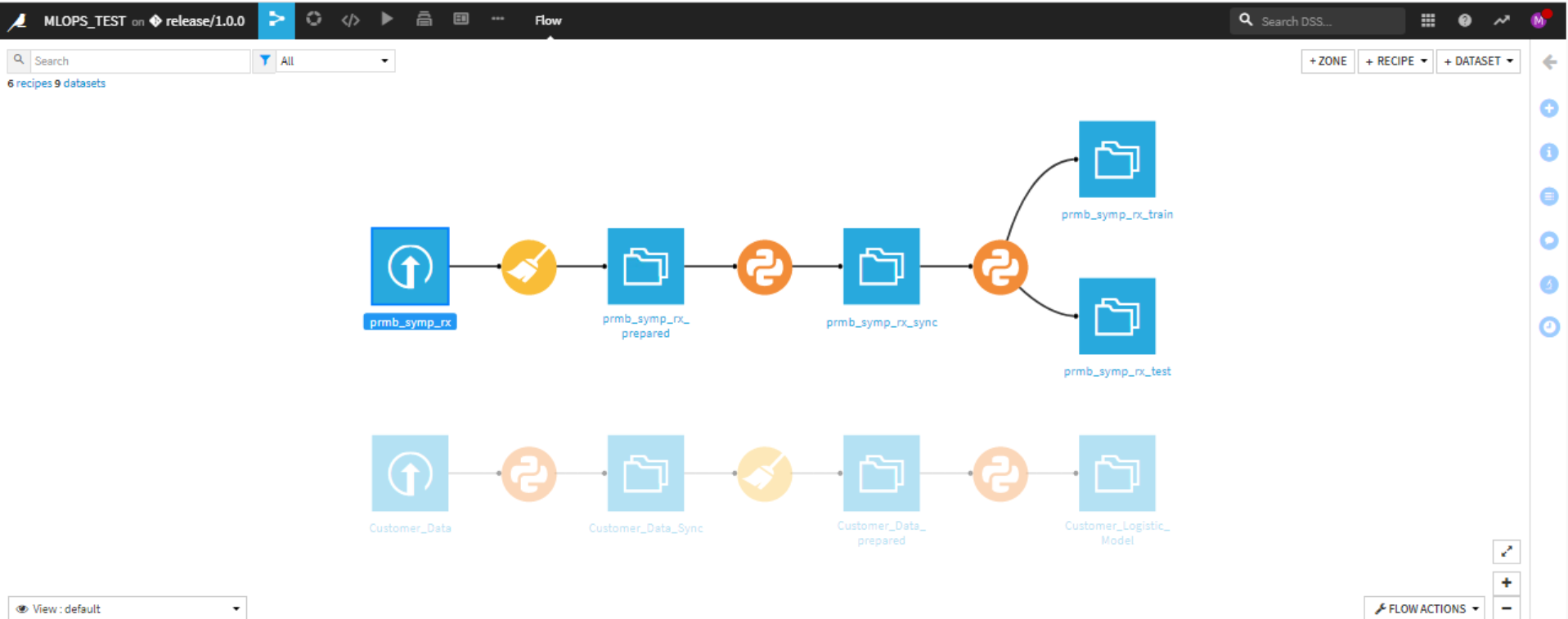
1. Bundles must be created in Dataiku Design Studio (Design Node),
2. Bundles must be deployed on Automation node, so Automation node IP address is required for the respective environment,
3. Creating the Dataiku infrastructure for Project Deployment.

# Dataiku Project Details

The project which we are using to solve the use case is **MLOPS\_TEST**



## Dataset Information:



The dataset from JET contains 20000 rows and 41 columns.

* Prepare recipe is used and added steps to prepare the dataset for python code recipe as input.
* Sync is used to maintain synchronization between the datasets.
* Python code recipe is used here in our case to split the dataset into train and test data.

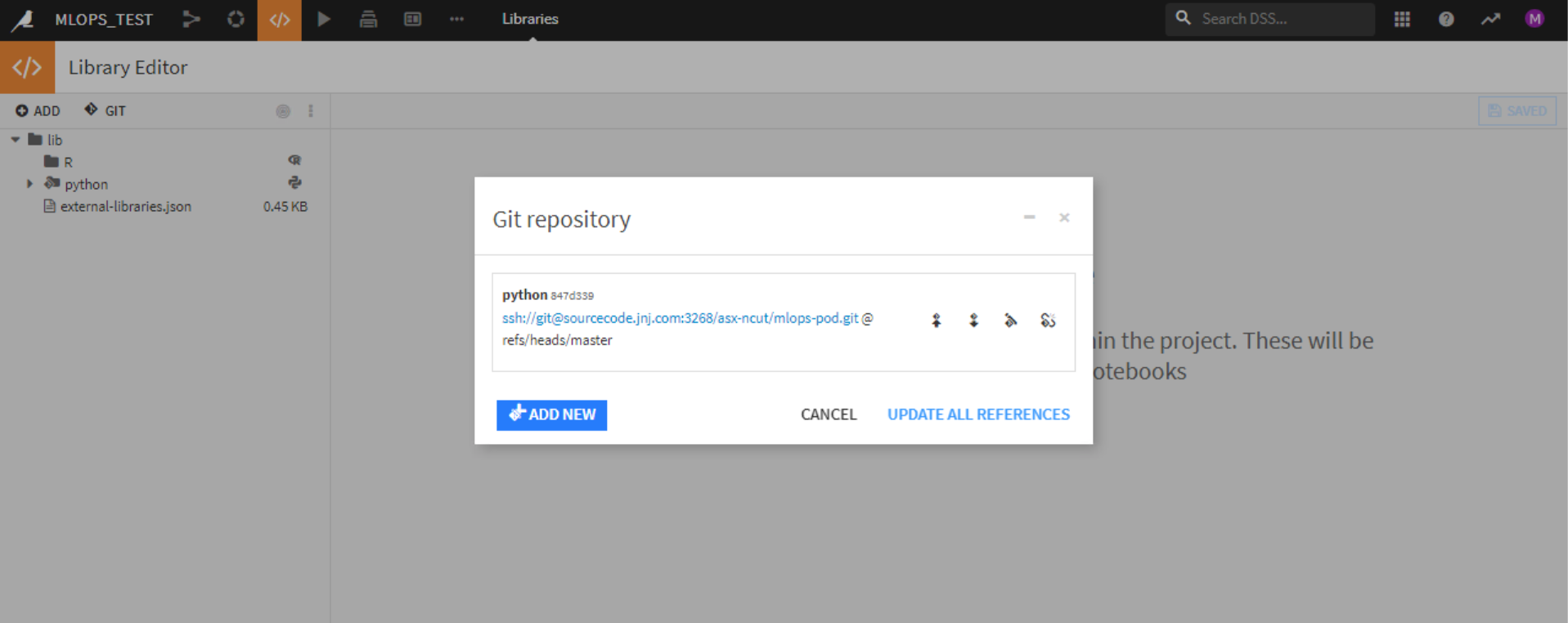
## Git Setup:

Remote GIT is setup to bitbucket version control

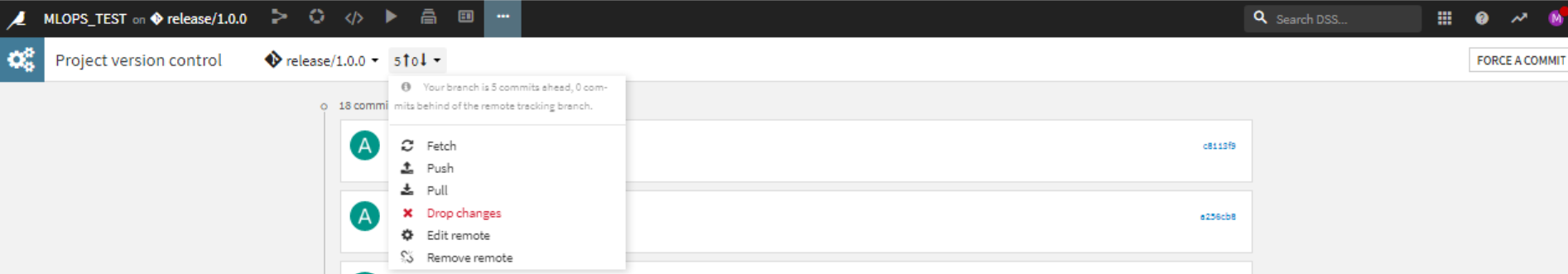
Added a remote git in version control

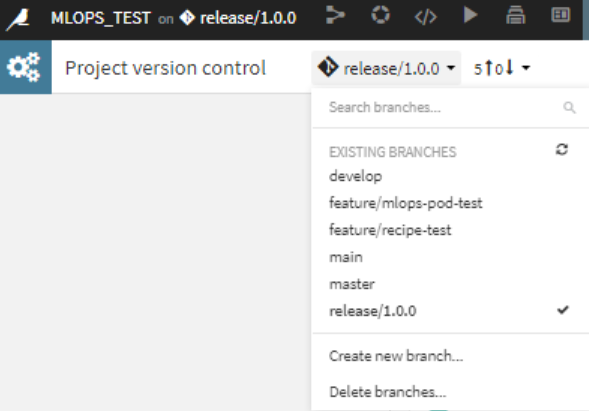


Linked the bitbucket repo to the library



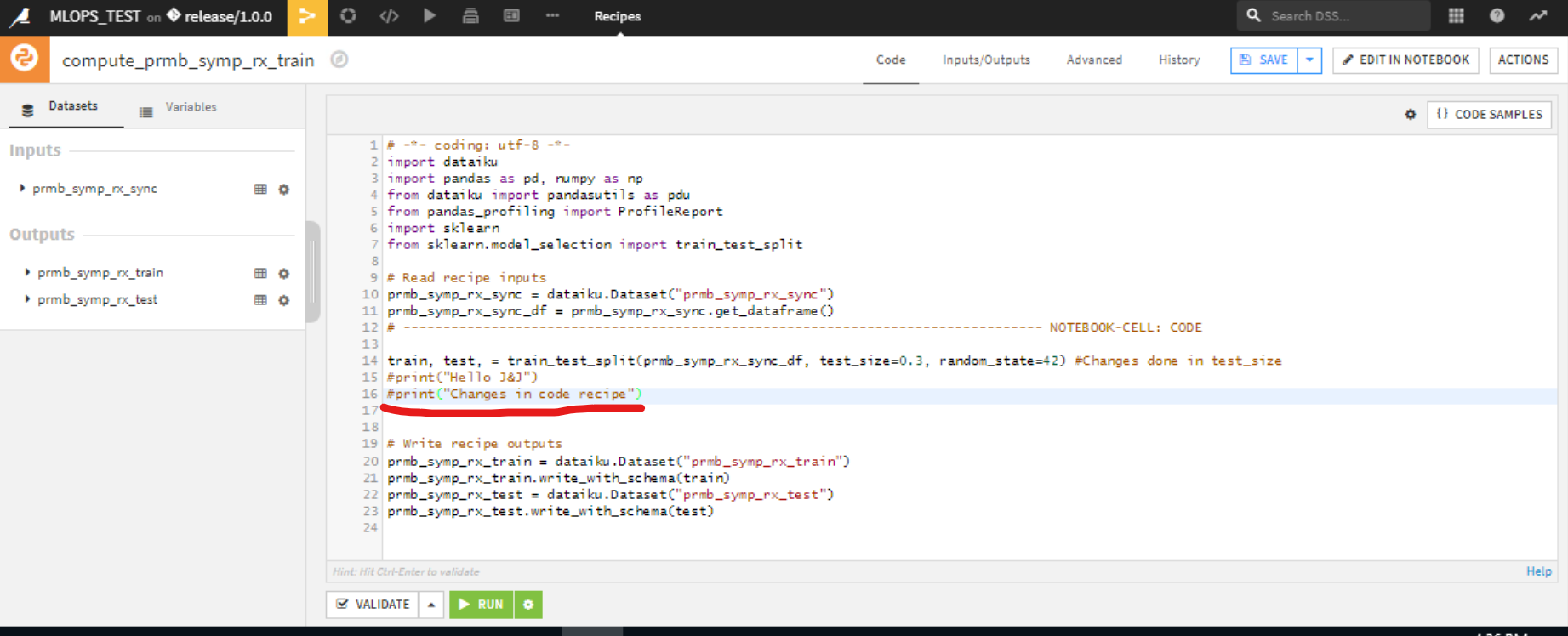
We can now see the completed setup of configuration of bitbucket.



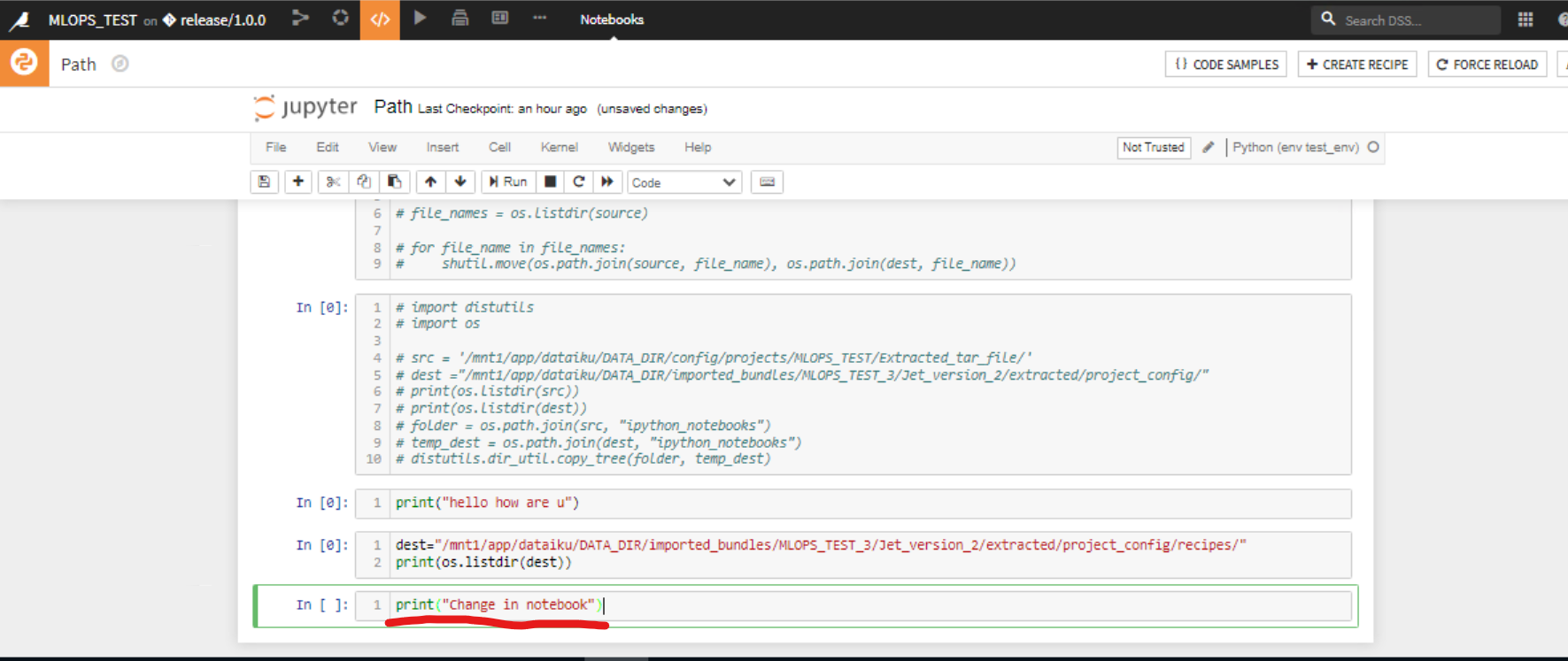


# Detailed steps involved in solving the use cases

**Step 1: Performing changes in the recipe file and Python Notebook.**

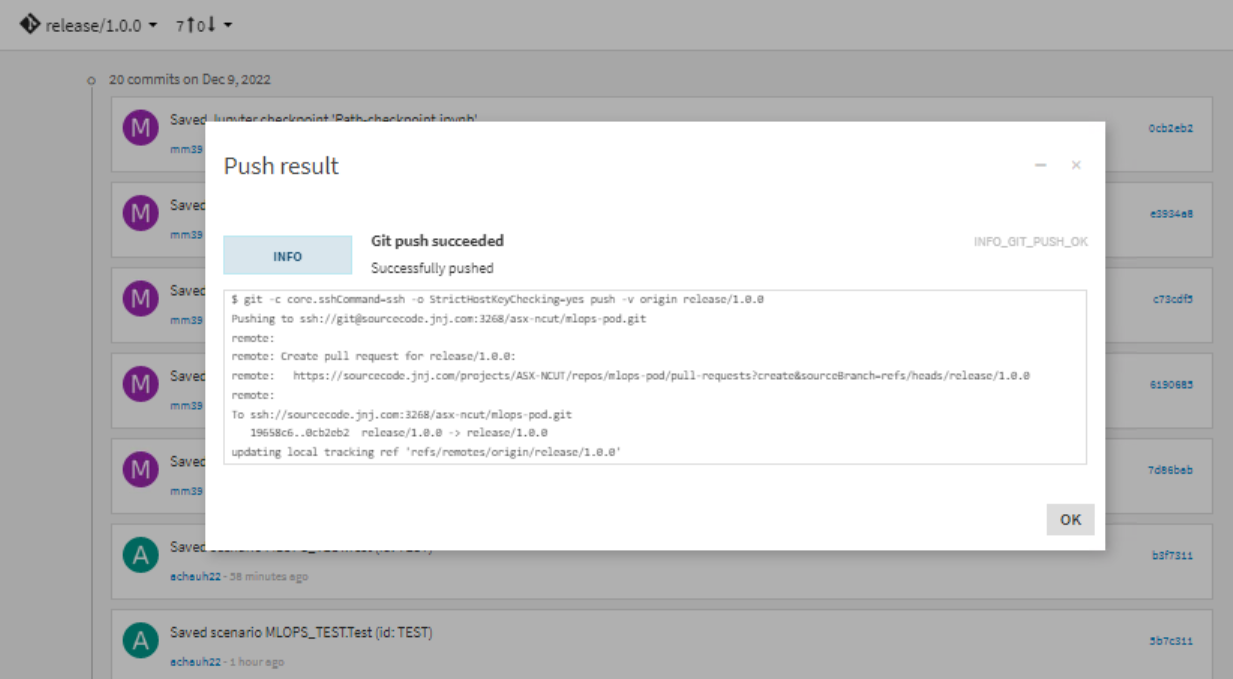


1.1 Making Changes in Code recipes.

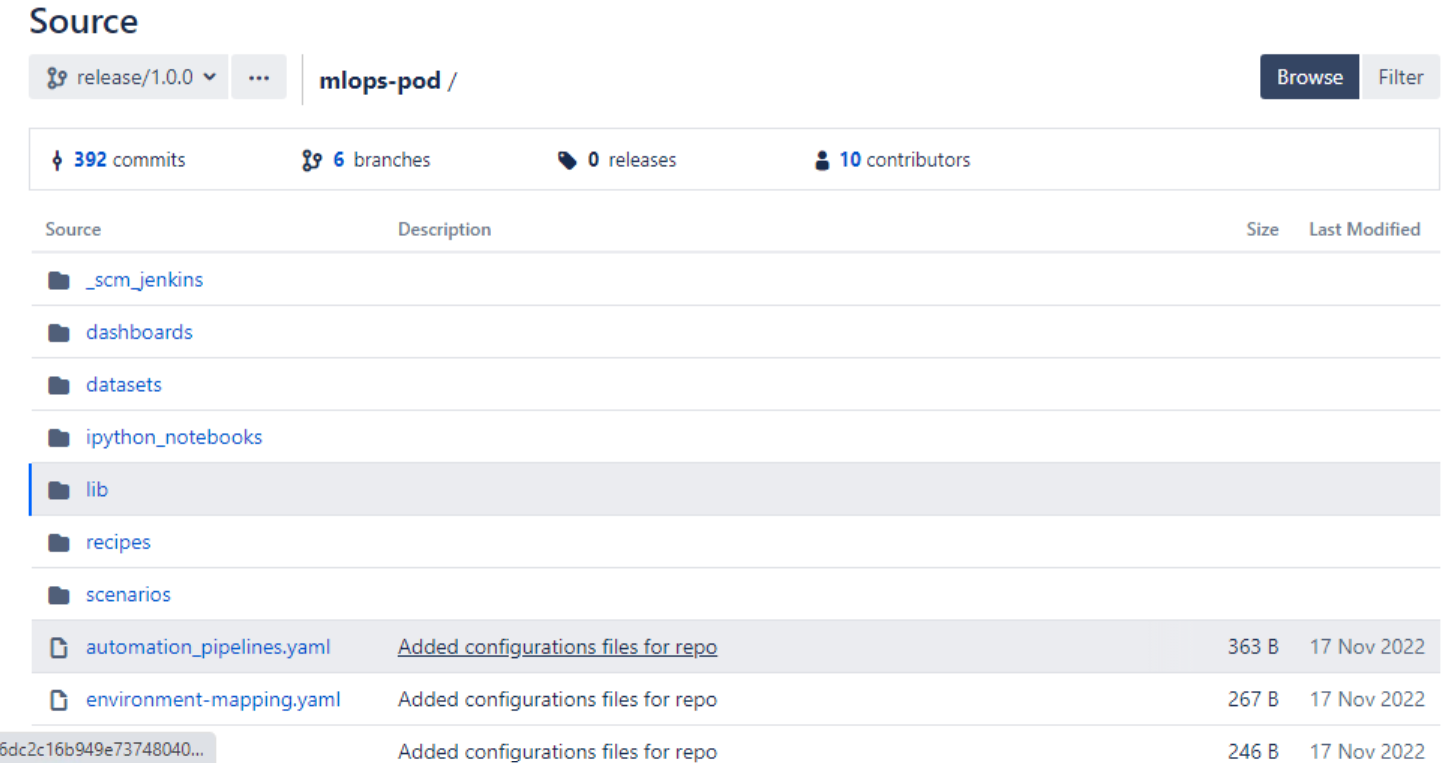


1.2 Changes done in notebooks.

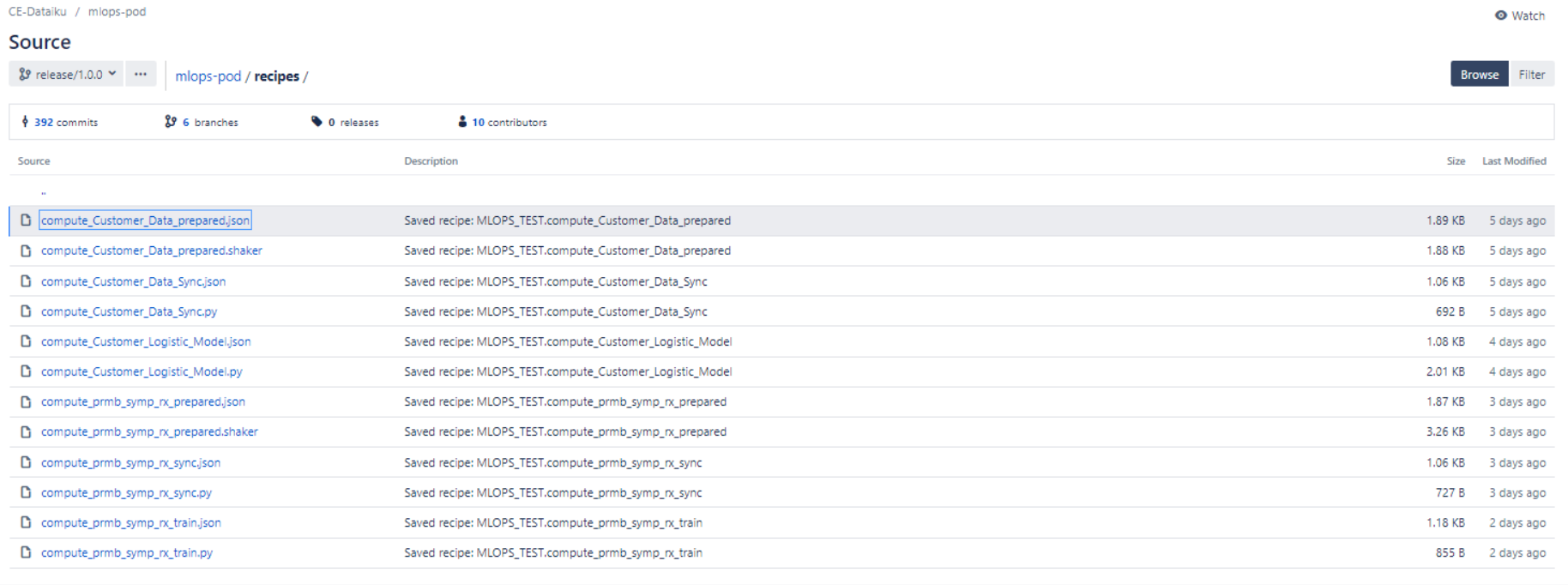
**Step 2: Pushing the changes from Dataiku Git Repo to Bitbucket Git Repository and Push Got successful**.



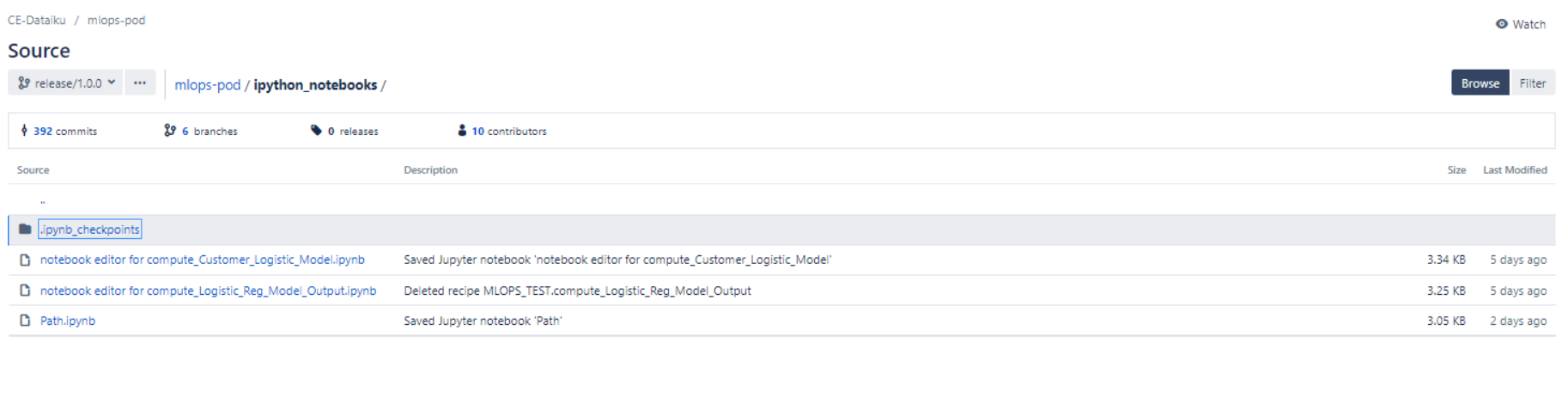
2.1 Git Push Successful



2.2 Folder Structure of our Bitbucket Repository.

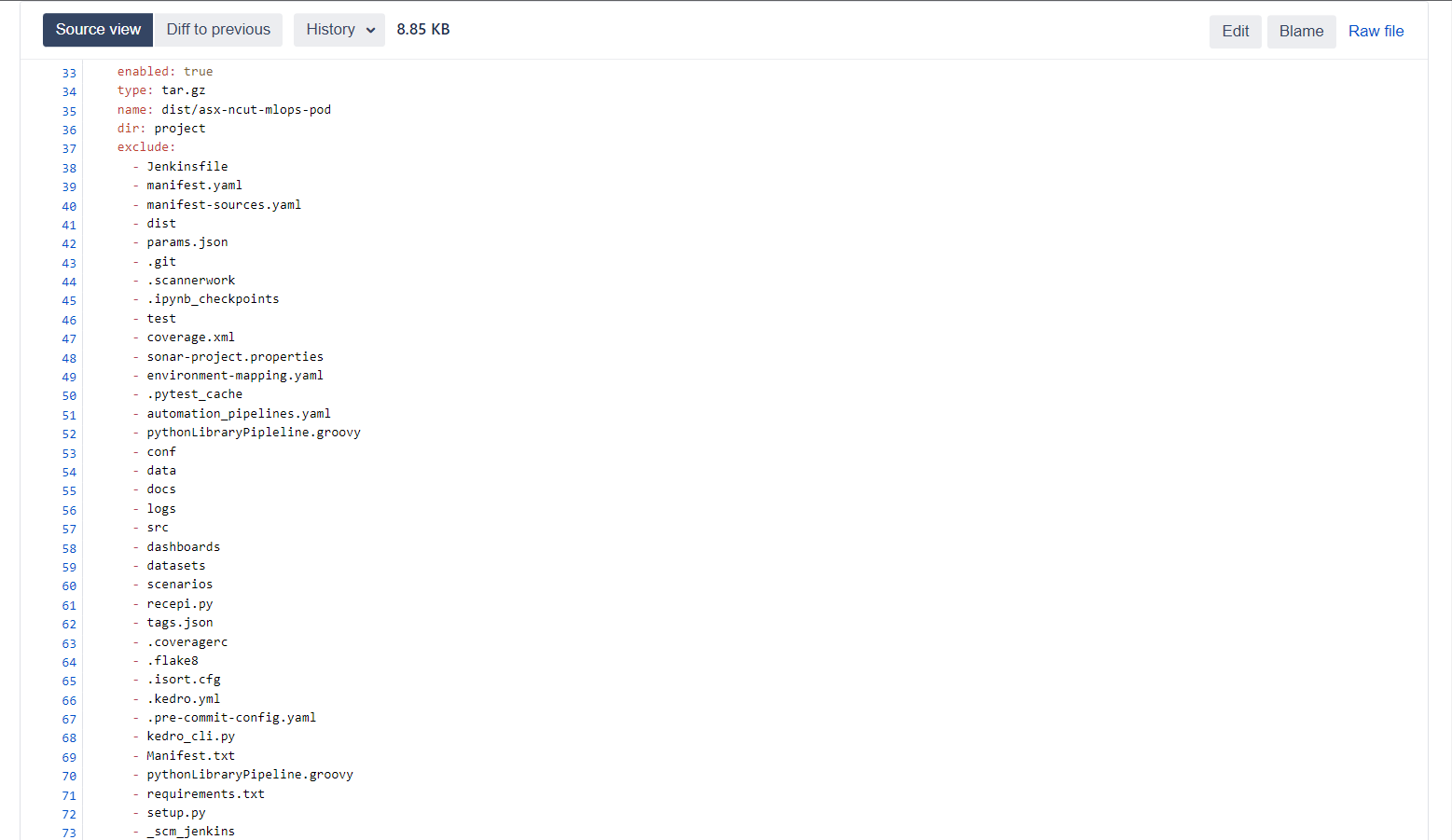


2.1 After pushing the changes, we can see updated code recipes files in recipes folder.

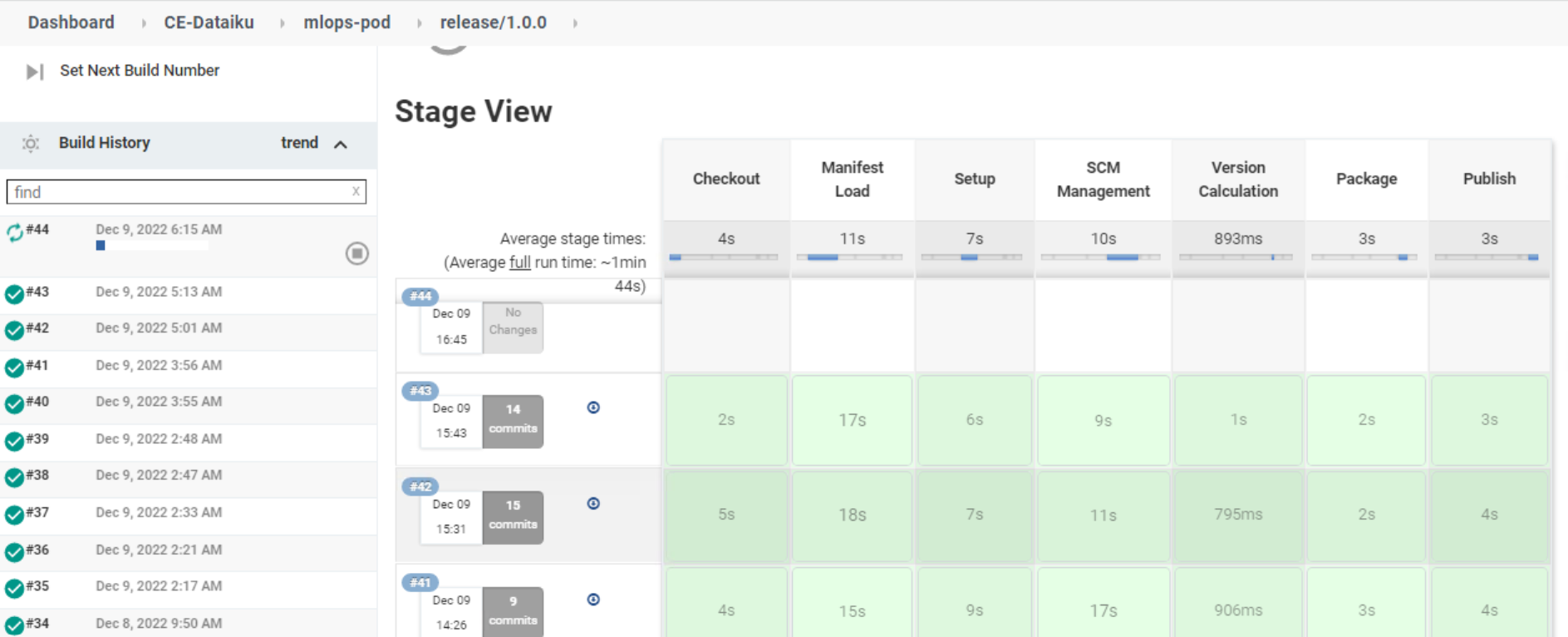


2.2 After pushing the changes, we can see updated notebooks in ipyhton\_notebooks folder.

**Step 3: Triggering the pipeline automatically for repo changes.**

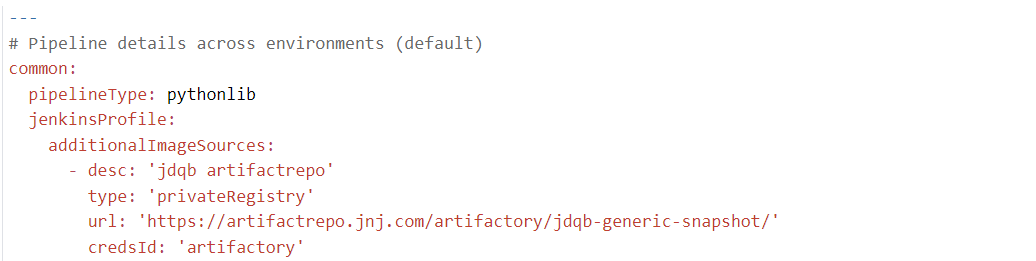


3.1 We have configured the **manifest.yaml** file and only included the required folder like code recipe and python notebook, other than that all folder and files are excluded.

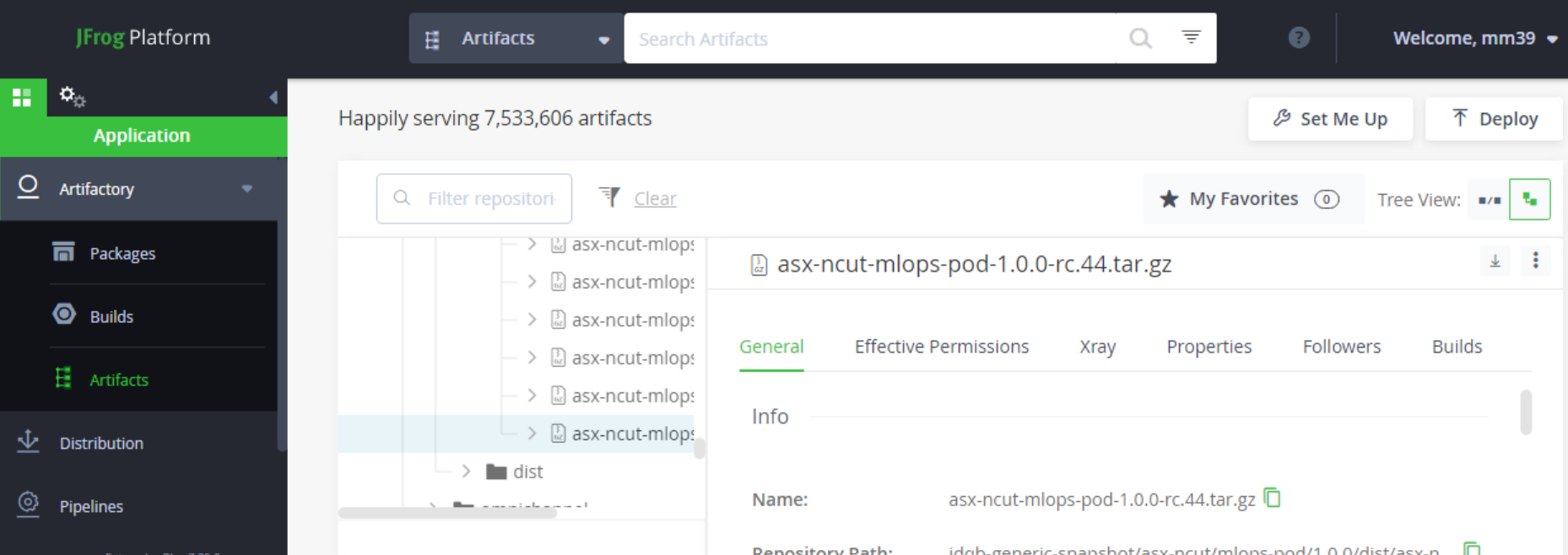


3.2 Pipeline will get triggered automatically once there are any changes pushed to bitbucket.

**Step 4: Pipeline will run and is responsible for package and publishing the changes to Artifactory.**

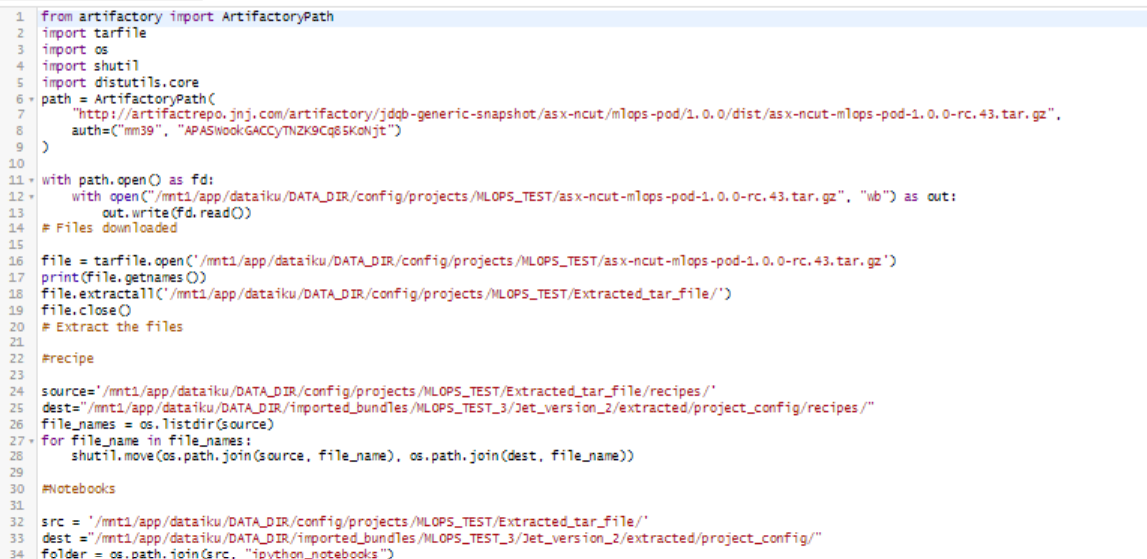
****

**4.1** URL mentioned in above snapshot(manifest.yaml) is the path for our Artifactory.



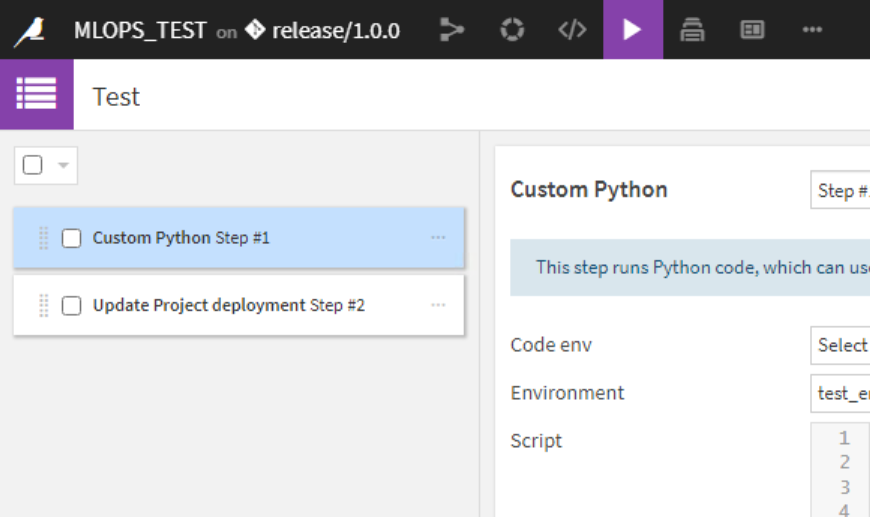
**4.2** After pipeline execution, we can see the latest .tar file stored in the path of Artifactory.

**Step 5: Fetch the updated files as .tar file, extract and place it in Dataiku.**



5.1 The above scripts will authenticate, download and extract the code recipe and python notebook to Dataiku automation node.

**Step 6: Run the scenarios and look for changes in automation node.**

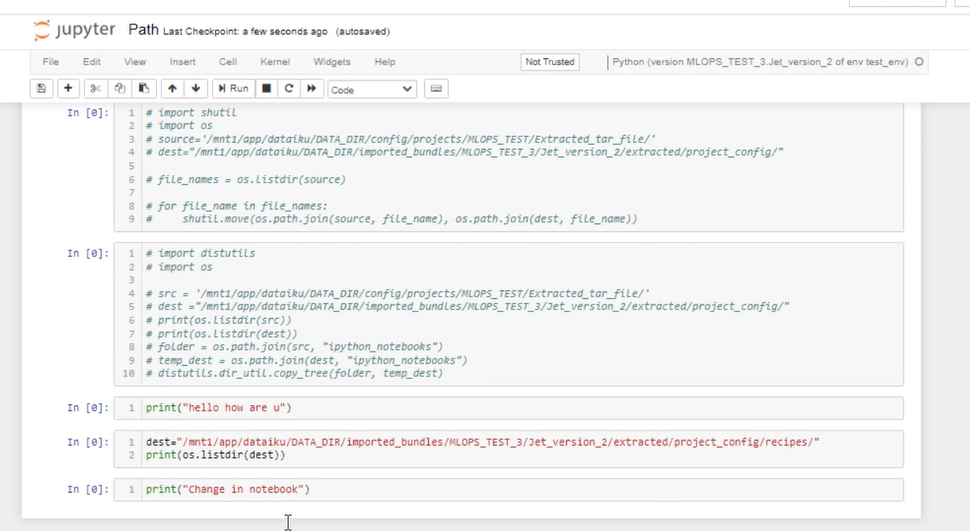


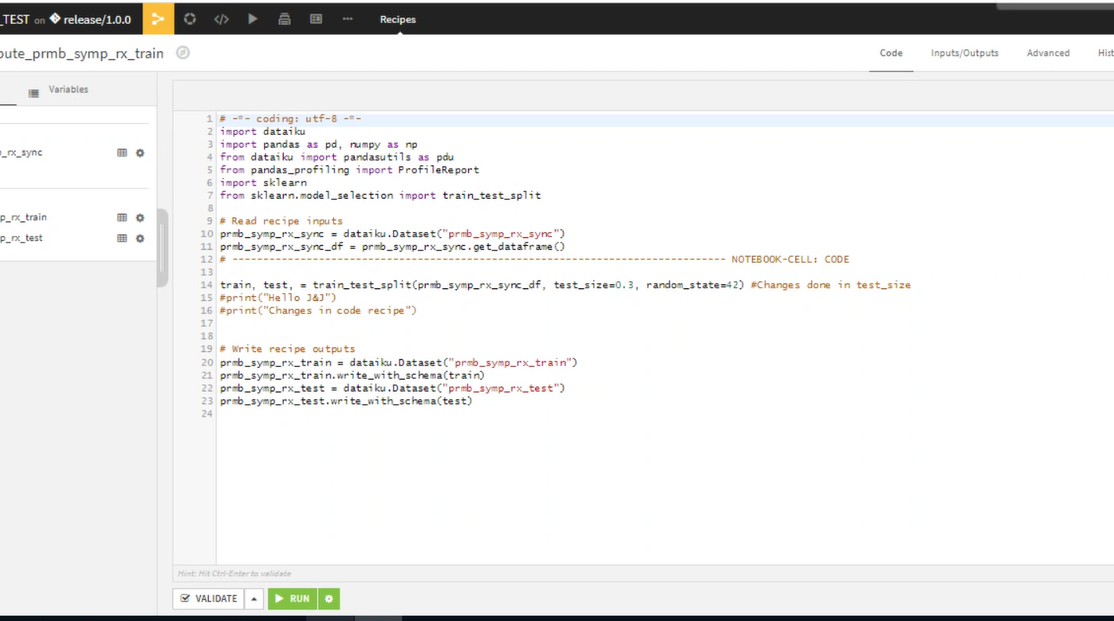
The scenario has 2 steps:

1. Run custom python script to authenticate with Artifactory, download the .tar file, extract and place the changes in the Dataiku automation node.
2. Update the project deployment as a step to avoid manual update of bundle.

**Results:**

We can now see the changes done for recipe file and python notebook





# Useful Links:

1. https://sourcecode.jnj.com/scm/asx-ncut/mlops-pod.git

2. https://artifactrepo.jnj.com/ui/repos/tree/General/eat-docker

3. https://jenkins.eat.jnj.com/jdqb-dev/

4. https://cejetdataikudev.jnj.com:11000/home/