00_prereqs

As part of the `Prerequisites`, we will walk through the initial setup and configuration steps needed in your environment before we can proceed with the `Ingest` phase labs, including:

- Decide on a unique **Prefix** that will identify the databases you will create
- Deploy a new Machine Learning (ML) project
- Configure and deploy an Applied Machine Learning Prototype (AMP)

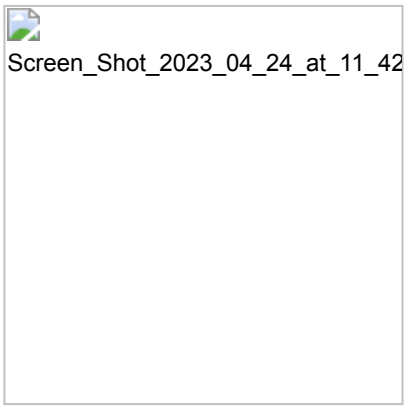
Cloudera Machine Learning (CML) Project

In this phase, we will be deploying an end-to-end machine learning project that will also be used in the 04_predict (04_predict.md) phases.

Lab 1: Configure and Deploy an AMP

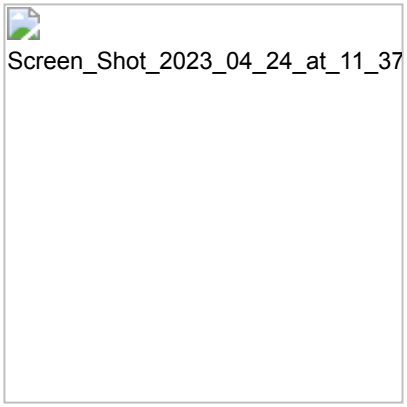
1. Open Cloudera AI
    - Select the `Machine Learning` tile on the CDP Home page

    > *Note: If you are not already at the CDP Home Page, click the bento menu icon in the top left corner, then click on* `Home`

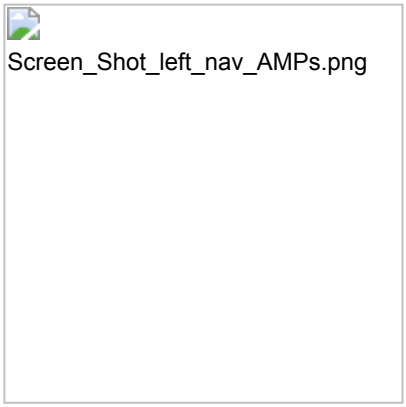
    

    - From the Machine Learning home page, click on the available Workspace (found under the `Workspace` column).

- Make note of the `Workspace/Environment` value listed as you will need to enter this value later on

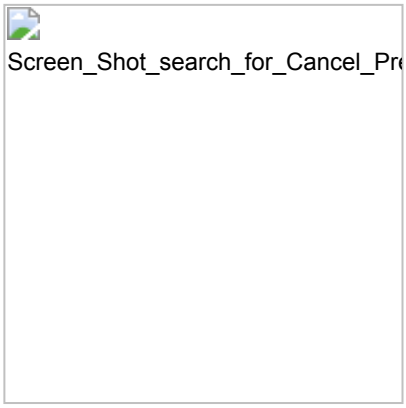

Screen_Shot_2023_04_24_at_11_37

2. Click `AMPs` in the left menu


Screen_Shot_left_nav_AMPs.png

3. You will see a catalog of available Machine Learning Prototypes

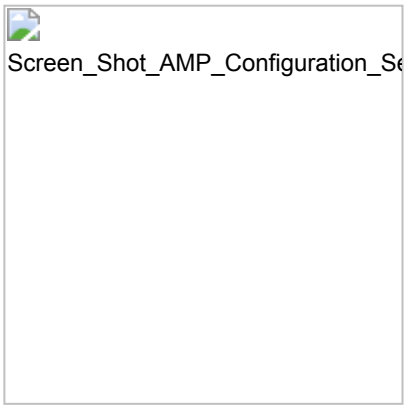
4. Search for the Canceled Flight Prediction prototype by entering `cancel` into the **Search AMPs** box and clicking the prototype tile


Screen_Shot_search_for_Cancel_Pr

5. Now click the `Configure & Deploy` button at the bottom-left

6. On the next screen, you are presented with **Environment Variable** values to fill in.

> *Note: For this step, you will need to choose a unique **prefix** that will identify the databases you will create and reference in the other labs (e.g. `evolve`)*



Fill out the form as noted below:

- **STORAGE_MODE:** `local`
- **SPARK_CONNECTION_NAME:** `<use the Workspace/Environment value we captured earlier>`
- **DW_DATABASE:** `<prefix>_airlines` (e.g. prefix = evolve)
- **DW_TABLE:** `flights`
- **USE_PREBUILT_MODEL:** `no`
- **Enable Spark:** `<click the toggle button to enable Spark>`
- Leave the rest of the fields with their default values.

7. Click the `Launch Project` button at the bottom-right

- It takes a few minutes to run the Jobs to build and deploy an end-to-end machine learning project
- CML will automatically execute the following 10 steps:

  `Step 1`: Job to install dependencies

  `Step 2`: Running the install dependencies job

  `Step 3`: Job to process the raw data files

  `Step 4`: Running job to process raw data files

  `Step 5`: Job to train the model

  `Step 6`: Run the model training job

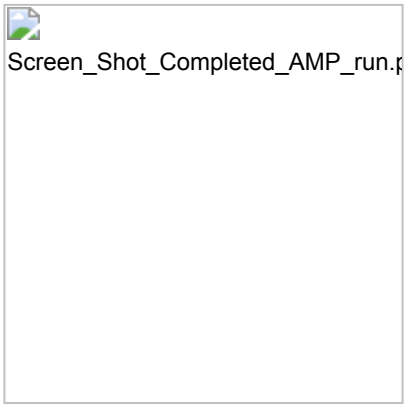
  `Step 7`: Create the flight delay prediction model API endpoint

  `Step 8`: Build the model

  `Step 9`: Deploy the model

  `Step 10`: Start the Application

- You can follow the executed step by clicking on the `View details` page to see the progress and what the prototype execution looks like in the background.
- All the steps above should be successful before proceeding to the next steps. It takes roughly 8 minutes for the prototype to be deployed. You should see a `Completed all steps` message above the executed steps.



Congratulations! You've completed your first lab!