

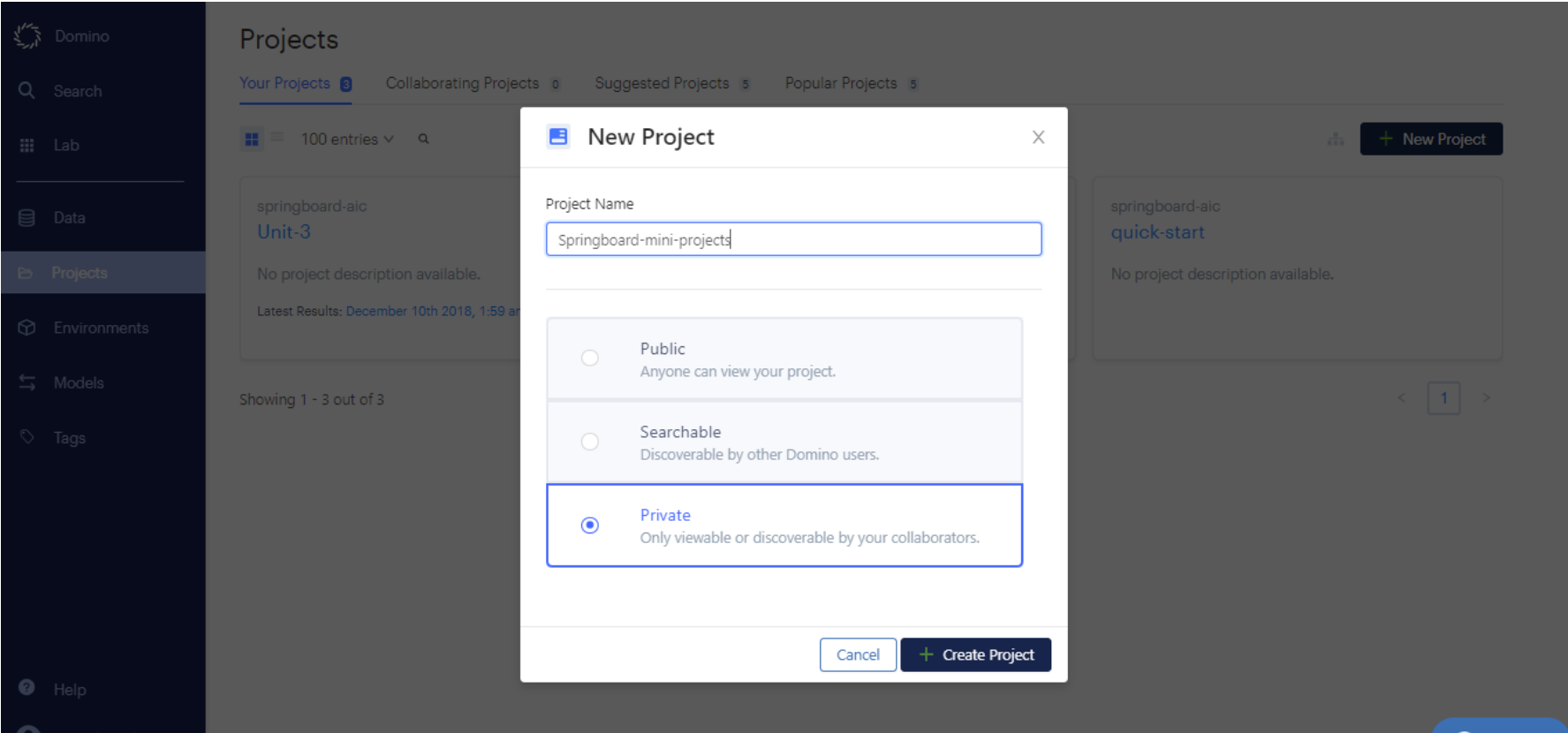
Domino Data Lab Setup Tutorial

- Dipanjan (DJ) Sarkar

Creating a Project

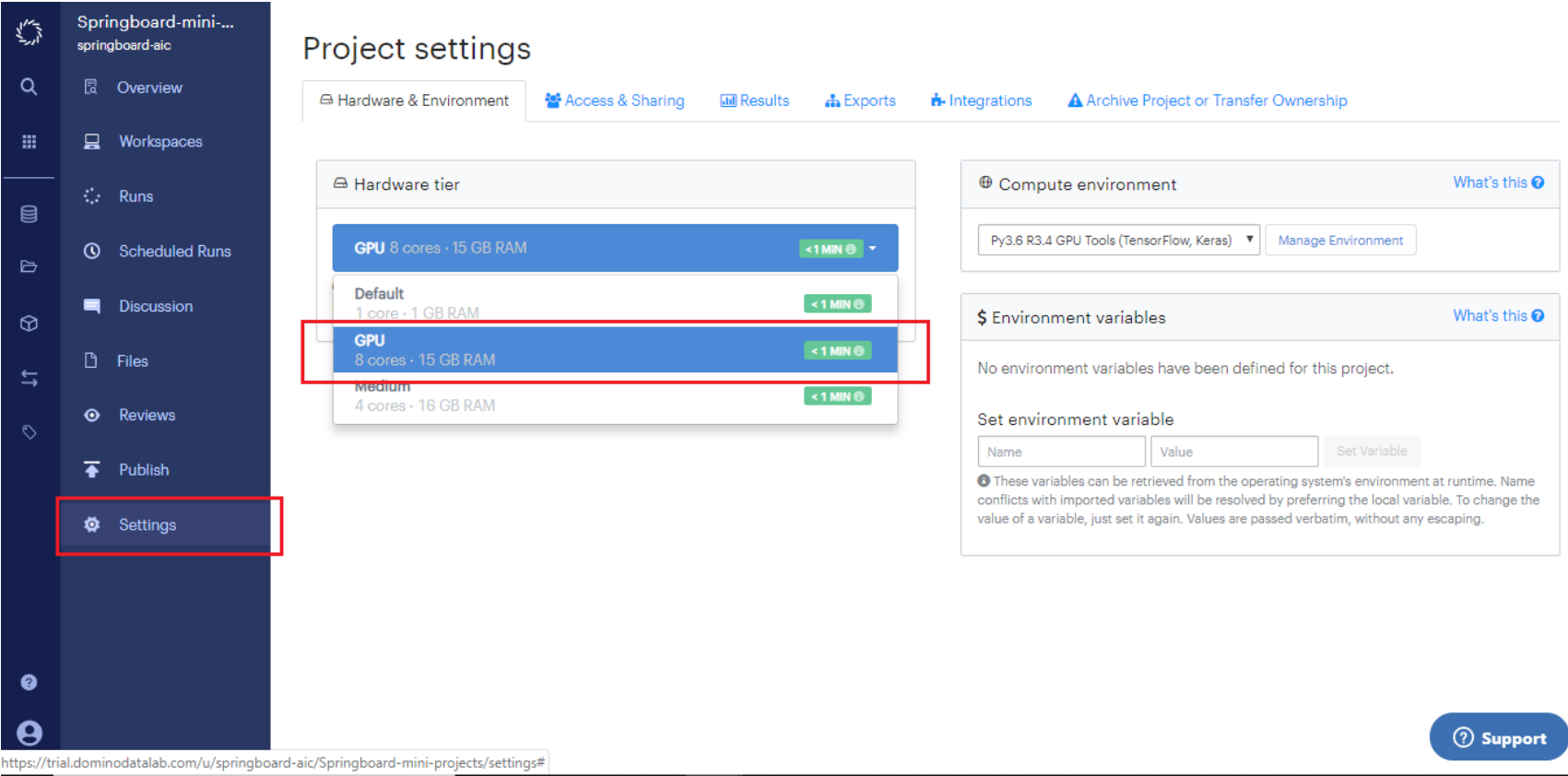
Make sure you have signed up for a Domino Data Lab account and head over to the following URL: <https://trial.dominodatalab.com/projects>

Create a new project as depicted in the following snapshot (this is where you will be working on your mini-projects on the cloud). You are free to give your project name based on your choice



Configuring your Project Settings

Head over to **'Settings'** and select the **hardware tier** of your choice (you get both CPU and GPU environments)



Also in ‘Settings’ select the **compute environment** of your choice (we recommend the **Py3.6 GPU Tools (tensorflow, keras) + Spark 2.2**)

Springboard-mini-...
springboard-aic

Overview

Workspaces

Runs

Scheduled Runs

Discussion

Files

Reviews

Publish

Settings

Project settings

Hardware & EnvironmentAccess & SharingResultsExportsIntegrationsArchive Project or Transfer Ownership

Hardware tier

GPU 8 cores · 15 GB RAM<1 MIN

Hardware is free during your trial.
Please [subscribe](#) for access to the most powerful machine tiers.

Compute environment

Py3.6 R3.4 GPU Tools (TensorFlow, Keras)Manage Environment

Owner

Test 1

Global

Domino Analytics Distribution py2.7 R3.4

Domino Analytics Distribution py3.6 R3.4

Intel Python

Py2.7 R3.4 GPU Tools (TensorFlow, Keras) + Spark 2.2.1[Local]

Py3.6 R3.4 GPU Tools (TensorFlow, Keras) + Spark 2.2.1[Local]

py3.6 R3.4 with LIME

Ubuntu16_DAD_Py2.7_R3.4

Name

Value

Set Variable

These variables can be retrieved from the operating system's environment at runtime. Name conflicts with imported variables will be resolved by preferring the local variable. To change the value of a variable, just set it again. Values are passed verbatim, without any escaping.

Support

Adding your Project Files

Head to the ‘Files’ section and you can create a new directory (for each of your mini-projects)

Springboard-mini-...
springboard-aic

Overview

Workspaces

Runs

Scheduled Runs

Discussion

Files

Reviews

Publish

Settings

LocalOther ProjectsGit Repositories

Springboard-mini-projects /

Extract Data SetAdd Directory

springboard-aic "Output from Run 5c0d6097c9e77c000766abd0" Run #3 b43b2ec on December 10, 2018 @ 01:59 am

Filter by Filename

Name

results

unit-3-data-wrangling-pandas

unit-3-sql-with-spark

.dominoignore

.dominoresults

Total size: 69.4 M

Add a New Folder

Folder Name

Folder names can only consist of letters, numbers, underscores, and hyphens; and they must start with a letter or a number.

CancelCreate Folder

Size

Modified

5.7 K -

64.7 M -

4.6 M -

529 B November 28th 2018, 2:5...

313 B November 28th 2018, 2:5...

Now you can upload all the necessary files for the corresponding mini-project (notebooks, datasets) in the respective directory (sample shown below)

Springboard-mini-...
springboard-aic

Overview

Workspaces

Runs

Scheduled Runs

Discussion

Files

Reviews

Publish

Settings

LocalOther ProjectsGit Repositories

Springboard-mini-projects / unit-3-data-wrangling-pandas /

Extract Data SetUpload

springboard-aic "Output from Run 5c0d6097c9e77c000766abd0" Run #3 b43b2ec on December 10, 2018 @ 01:59 am

DRAG & DROP

your files anywhere, or [browse for files](#)

UploadCancel

Files matching .dominoignore rules will not be uploaded.

File size may not exceed 200 MB and uploads are limited to 1000 files at a time. You can also use the [Domino CLI](#) to upload large volumes of files.

...

.ipynb_checkpoints

.dominokeep

cast.csv.bz2

Mini_Project_Data_Wrangling_Pandas.ipynb

251.6 K -

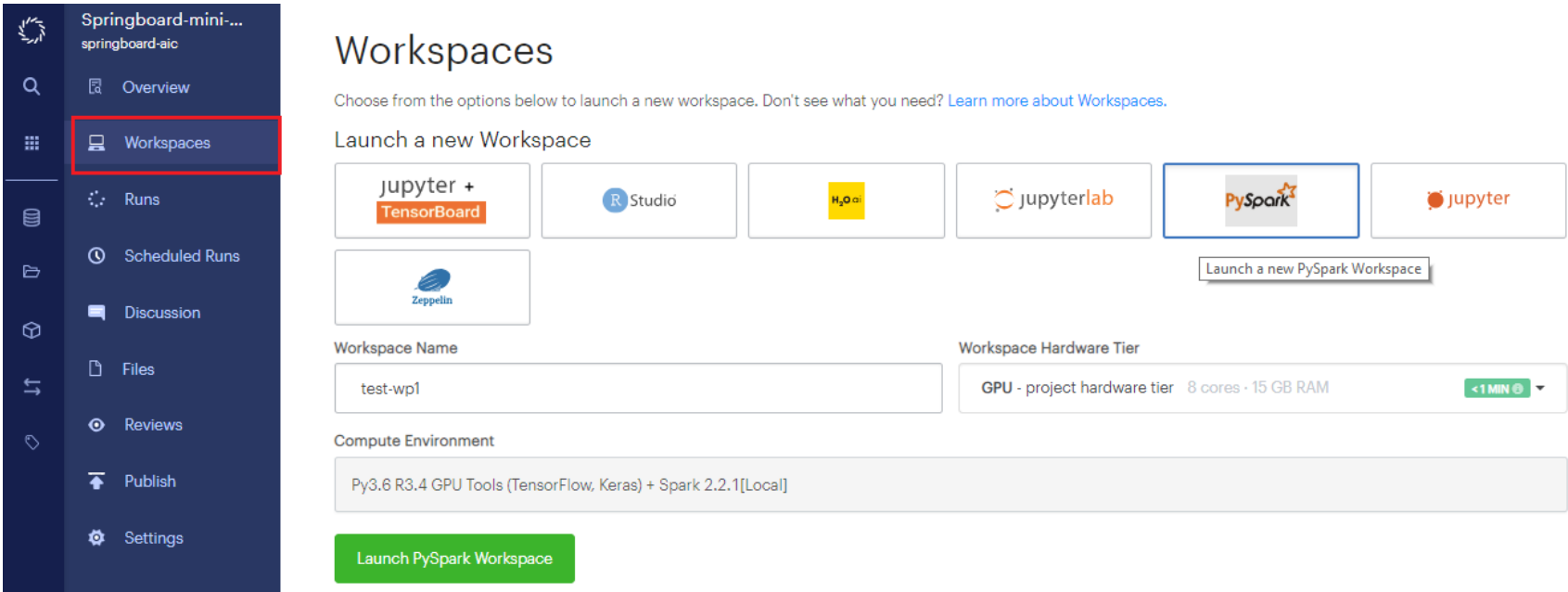
274 B December 2nd 2018, 8:2...

59.3 M December 2nd 2018, 8:4...

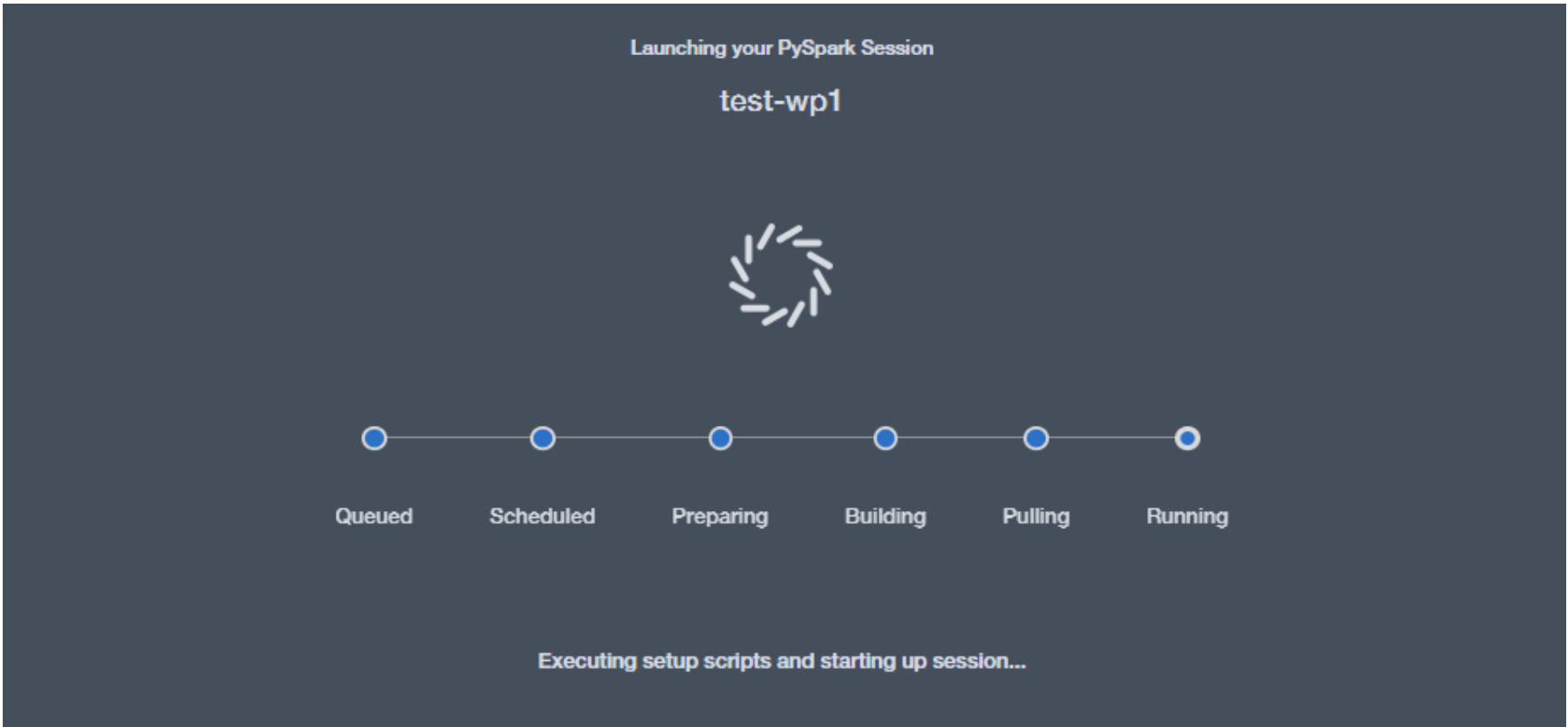
58.8 K December 10th 2018, 12:...

Launching a Workspace

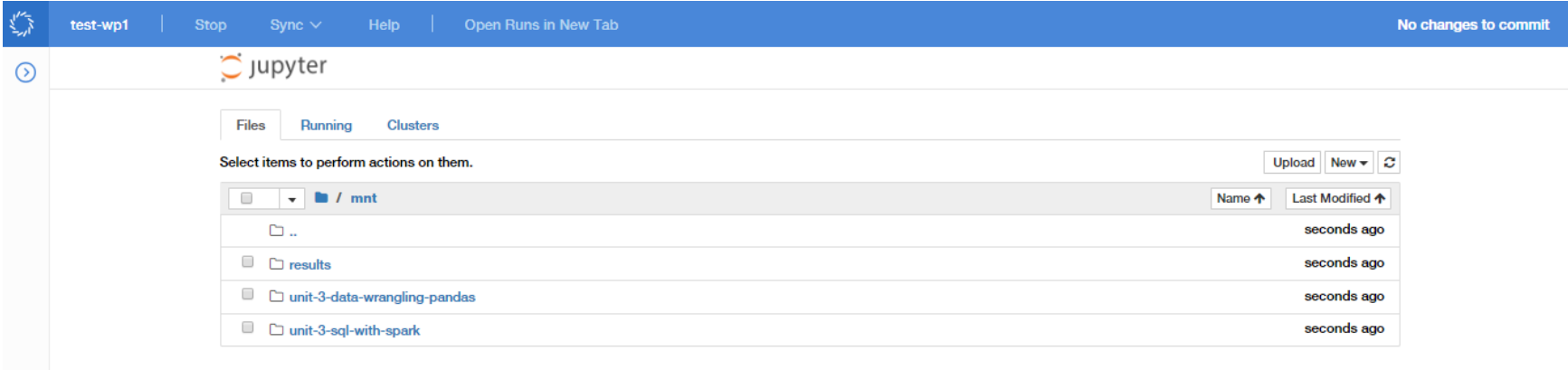
Head over to the `Workspaces` tab in your project, make sure to give a name to your Workspace, select the Hardware tier (or use the one you had selected earlier along with the compute environment) and Launch a new workspace! We recommend launching a PySpark Workspace where you get all the necessary Python libraries along with pre-configured Spark environment and context variables.



Domino will spin up a container based on the availability of instances (servers). Usually this can range from less than a minute to a few minutes.

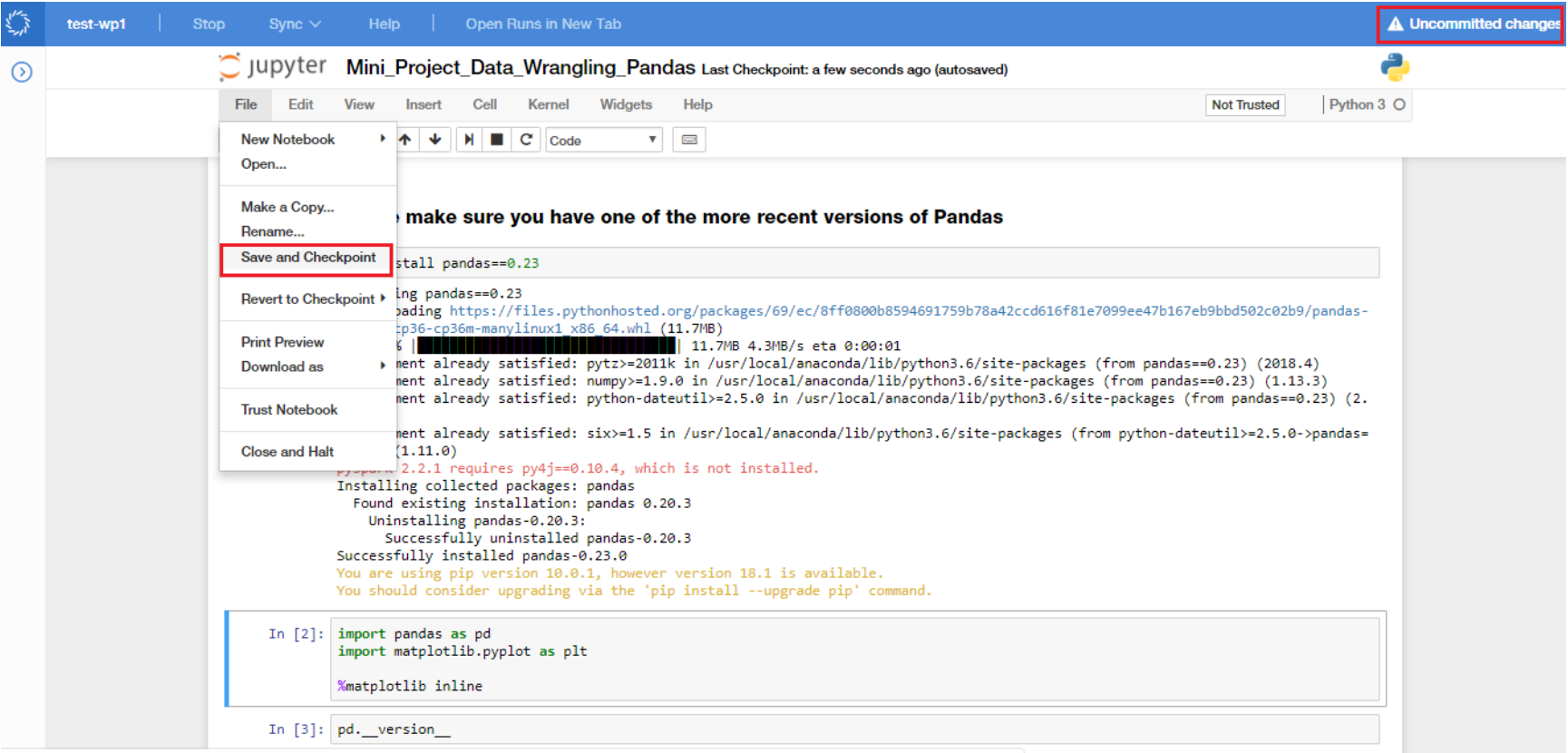


Once your workspace is up and running you should be able to see your `Files` directory which is typically present in ``/mnt`` as depicted in the following snapshot.

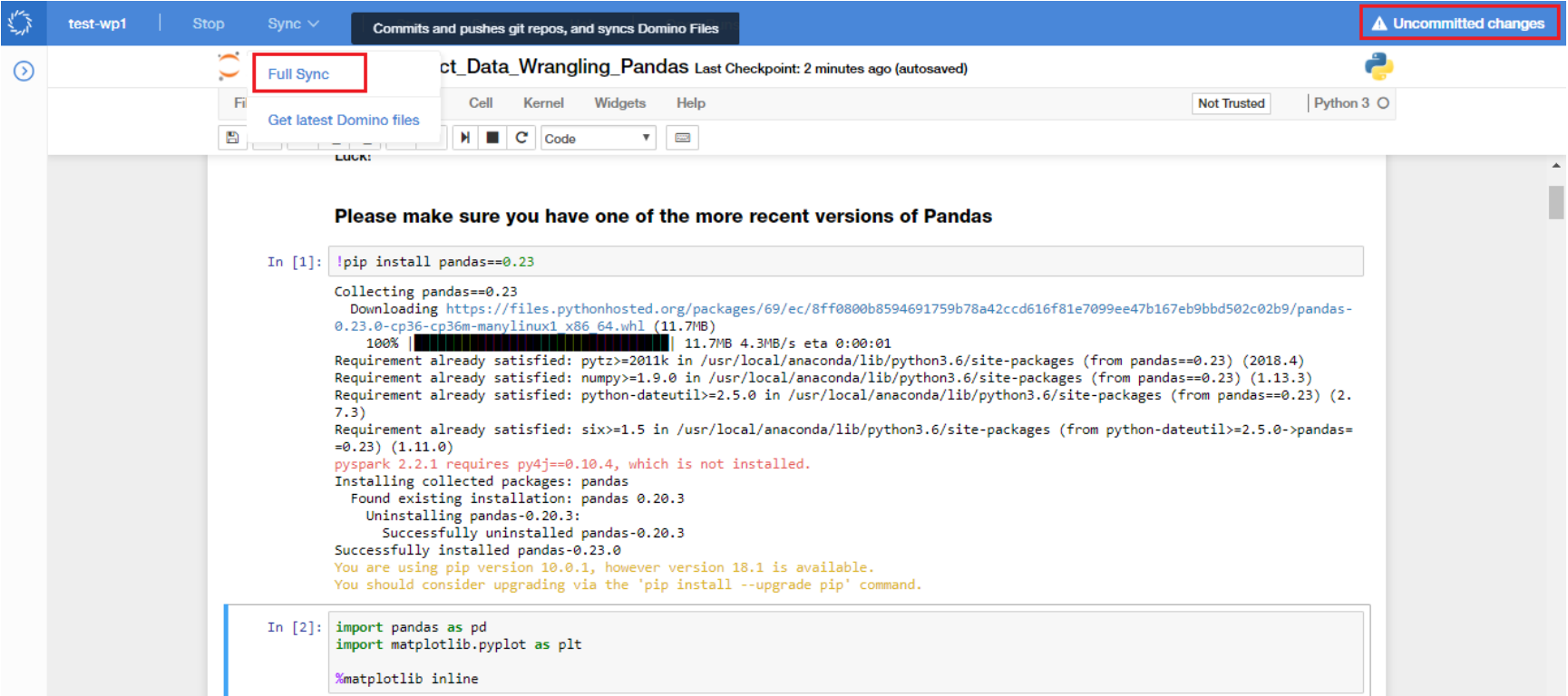


Working on your Project

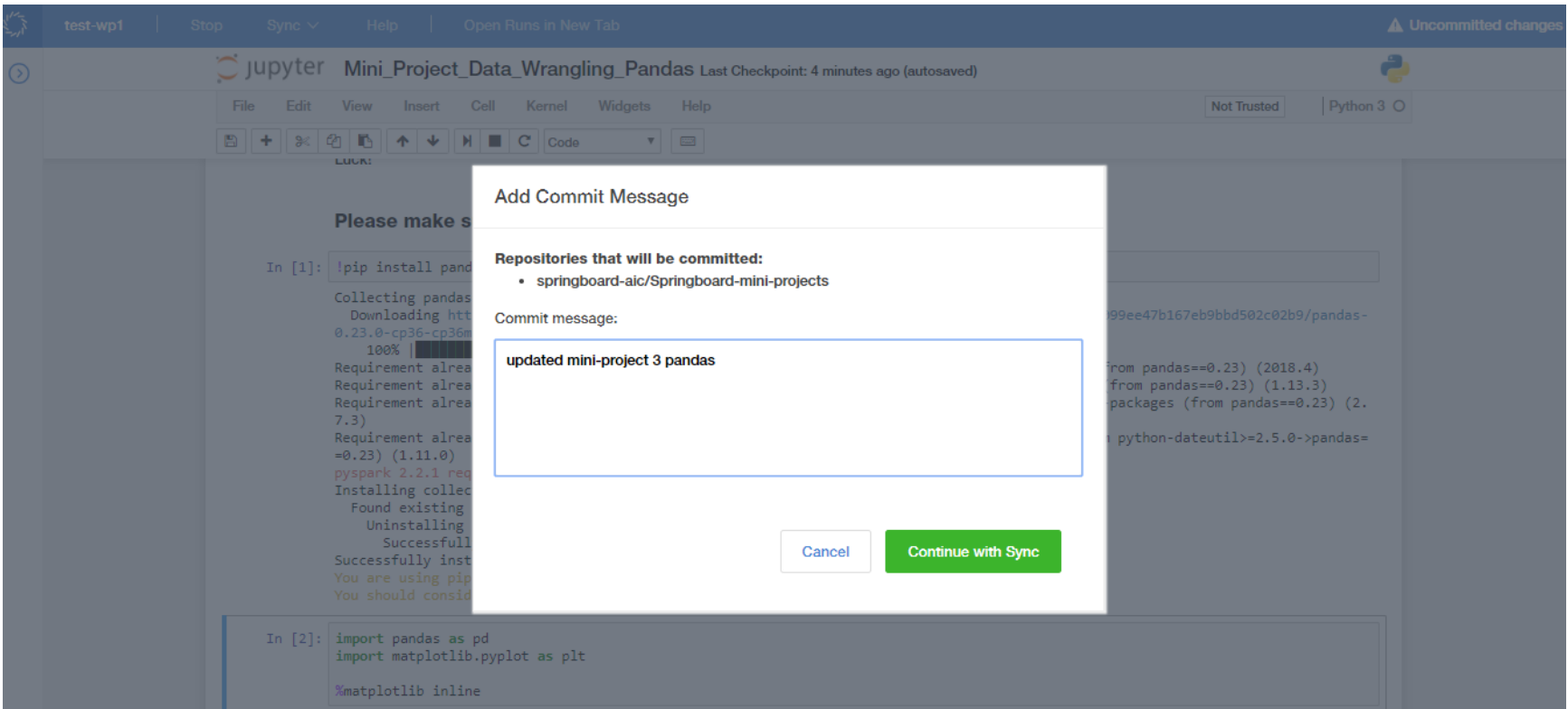
You can now open up any Jupyter notebook (including ones from your mini-projects) and start working on them. Once you make changes to your notebook, remember to save the notebook. Notice the **`Uncommitted changes`** label on the top right. Since you are working on a containerized environment, it is essential to also save the changes you make on the cloud to your storage directory (This is called Syncing in Domino)



Use the **`Full Sync`** option to sync all changes from the containerized instance to your storage (**`/mnt`**) in your project on Domino and also get any local changes from Domino to your running instance.



Add in a commit message which is useful to remember what you did last!



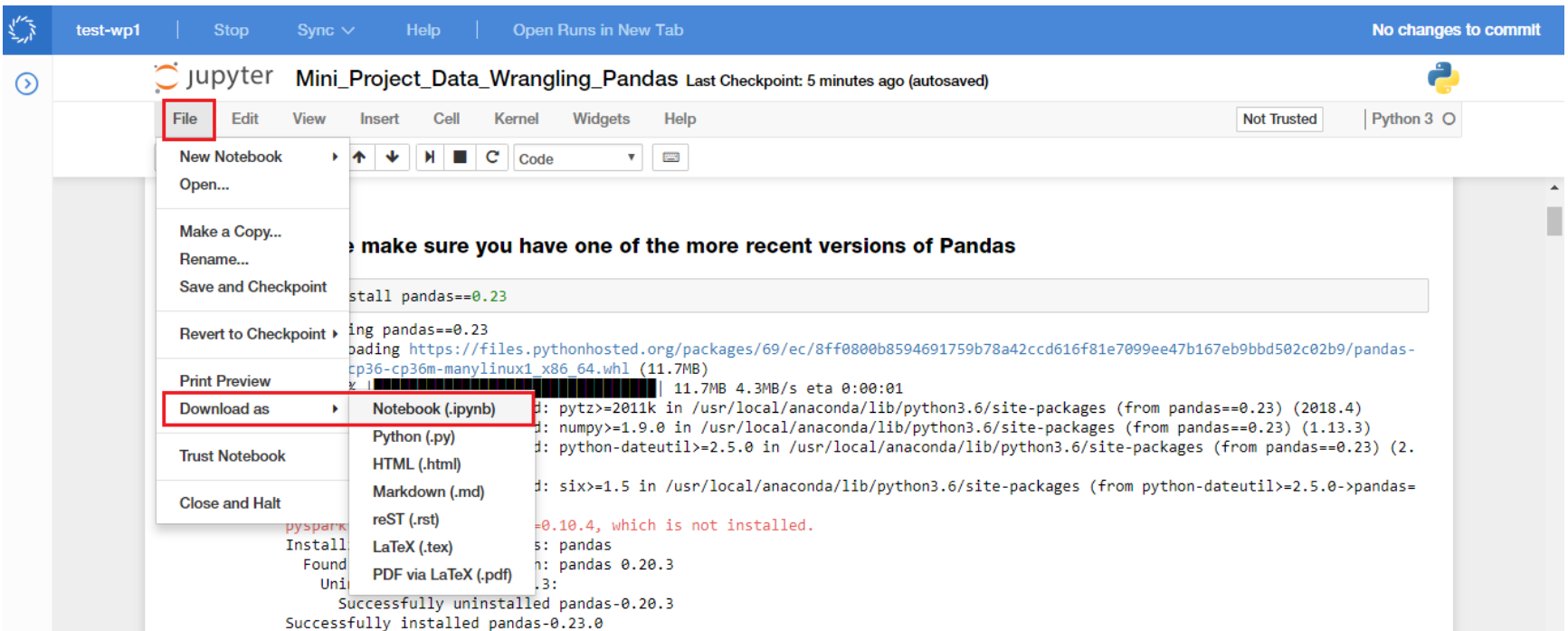
Once the syncing process is complete, you should see a `No changes to commit` label on the top right. Remember to save and commit frequently when using the Domino platform!



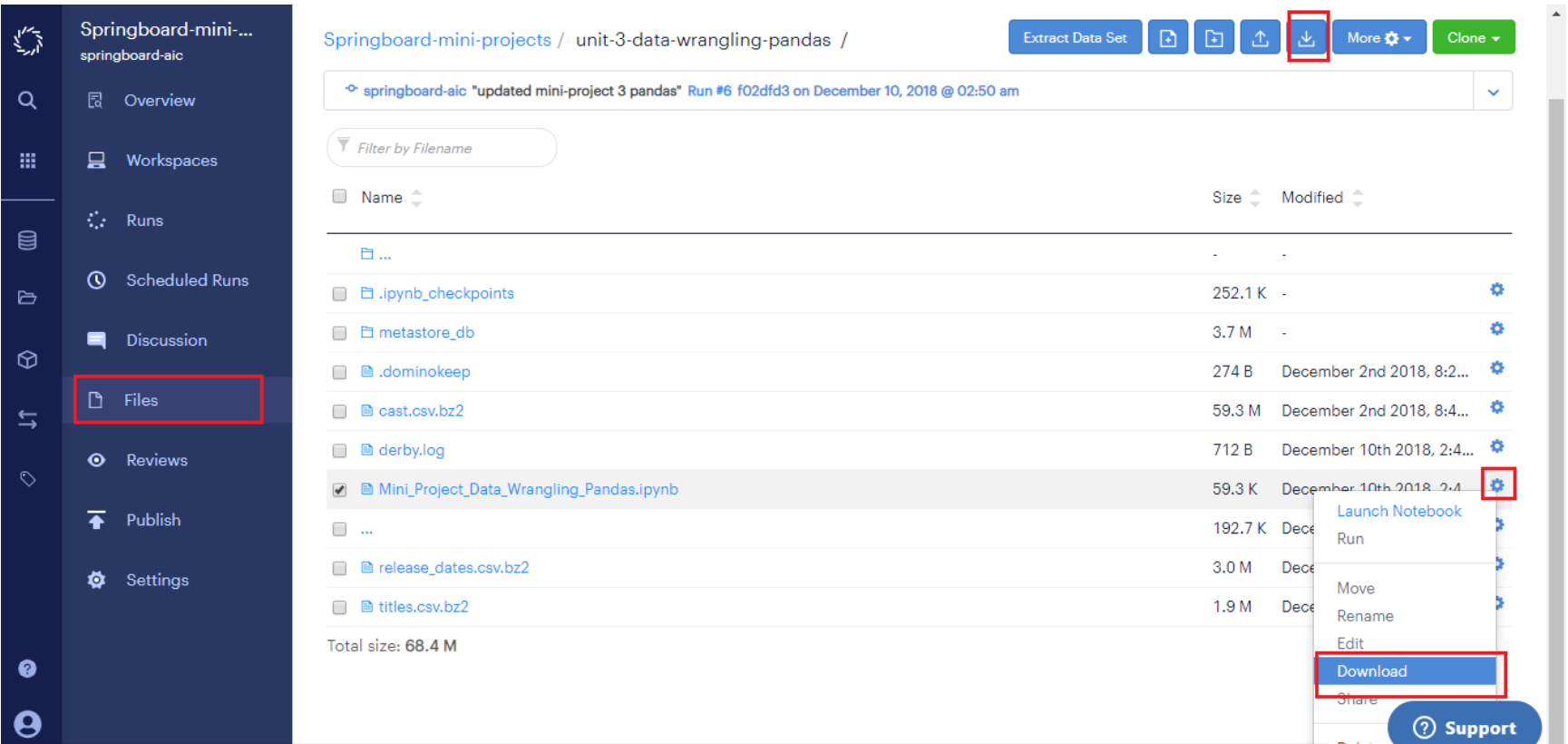
Download your work (for submissions)

Typically often you just need to submit your mini-project solutions as Jupyter notebooks in GitHub repositories. For this, once you are confident that your solutions are working well (or you can always revisit the previous steps and make changes and come back to this step). You can follow any of the following two methods to download your work to your system and push it to your GitHub repository (follow the GitHub tutorial in the curriculum pre-work in case you might need a refresher on how to add your projects on GitHub)

You can download the notebook directly from the Jupyterter interface!



You can also download the notebook from Domino’s Files section.



Hope this gives you a good introductory guide to using Domino Data Lab’s Data Science Platform!

Feel free to refer to their tutorials here for further detail: <https://support.dominodatalab.com/hc/en-us/articles/360016336612>