Branch and directory

If a different branch or directory is being used, review the prefect.yaml file for the branch. Also if the orchestrate.py file is in a nested directory for example week3/3.4/orchestrate the deployment needs to have that. "prefect deploy week3/3.4/orchestrate.py:main_flow -n taxi1 -p zoompool" - Quinn Avila

Problem: Working from a nested directory failed when the prefect.yaml and deployment.yaml are not in the root and the work dir is specified in the prefect.yaml file
Solution: Activate project with Prefect init from the root

Deployment Data Files

For the initial workflow deployment to work properly (Q1, Q2) Data files should be included in your github repo, and the deployed flow should know the file path relative to the root repository. In my case, I had to include the nested week3 folder I'm using.

```
@flow
def main_flow(
    train_path: str = "./week3/data/green_tripdata_2021-01.parquet",
    val_path: str = "./week3/data/green_tripdata_2021-02.parquet",
) -> None:
```

Deprecation Warning: reg:linear Deprecated in favor of reg:squarederror

Problem: On best_params we have used "objective": "reg:linear". If you run the XGBoost model, a warning message is displayed stating that the reg:linear objective function is deprecated and should be replaced with reg:squarederror.

Solution: Replace to "objective": "reg:squarederror" or even

import warnings
warnings.filterwarnings('ignore')

To hide all warnings

https://discuss.xgboost.ai/t/xgboost-objective/789/4

Inability to Intentionally Trigger Flow Failure for Email Notification

Problem: Send an email when flow fails. I could not make my flow fail intentionally.
Solution: One strategy I implemented consisted of validating either train and validation path so if one of these doesn't exist the system sends an email.

By Erick Calderin

FileNotFoundError on a flow run

Problem: You see FileNotFoundError: [Errno 2] No such file or directory on a flow run.

Solution: Prefect fetches your code from the remote repository. Anything that's not on the remote repository, it won't find. If you're getting an error stating that the file where your flow is defined can't be found, you likely haven't committed the file to the remote repository yet.

By Daniel Meleras

FileNotFound do git push

Problem: FileNotFound on a flow run.

Solution: Just adding to the previous FAQ. git commit and git push

By Quinn Avila

FileNotFoundError on a flow run (2)

Problem: You see FileNotFoundError: [Errno 2] No such file or directory on a flow run.

It also depends in which point of your filesystem you launch the deployment command.

The relative path in the orchestrate.py file are relative to the point of launch of the deployment command, NOT relative to the position of the file in the filesystem.

By Giordano Pignagnoli

FileNotFoundError for ./data/* during running the deployment of orchestration.py

Problem: You see FileNotFoundError: [Errno 2] No such file or directory for ./data/* when you want to run a deployment on Prefect

Solution: 1st Change your data input path (i.e., train_path, val_path) in the orchestration.py script to absolute paths. 2nd Use an absolute path when running 'prefect deploy' (e.g., prefect deploy $PWD/03-orchestration/orchestrate.py:main_flow …).

By Valerie Gumhalter, June 2023

Interaction free Prefect Cloud login

Problem: How to login to prefect cloud without using the interactive login

Solution: You need to create an api-key in prefect cloud in the “app.prefect.cloud/my/api-keys” screen. Then copy/save the key and use it to login with the following cli command.
`prefect cloud login --key <your_api_key>` --workspace “<your_handle>/<your_workspace_name>”

By Zharko Cekovski

Prefect + Google Cloud Storage

Problem: Uploading and downloading data to and from Google Cloud Storage
Solution: Follow the instructions in https://github.com/DataTalksClub/mlops-zoomcamp/tree/main/03-orchestration#notes under “Prefect local deployment using Google Cloud Storage as data storage”

By Unai Garay

Prefect + Google Cloud Run

Problem: Running Prefect with Google Cloud Run
Solution: Follow this tutorial

By Unai Garay

Exiting with Status Code 1 for Deployment

Problem: 'C:\Users\Chang' is not recognized as an internal or external command,
operable program or batch file. Exiting with status code 1. This occurs during the deployment.
Solution: When deploying Prefect locally, special characters such as “ “ between folder names are not allowed. Use a different directory stored in Python. I used a separate account that does not have space on my username.

By Sam Lim

Inconsistent Prefect Server UI Visibility due to Profile Configuration Conflict

Problem: When I was running `python orchestrate.py` my runs weren’t showing up in my local prefect server UI. The reason, as it turned it out, is that despite setting `PREFECT_API_URL = "http://127.0.0.1:4200/api"`, I still had `PREFECT_API_KEY` set in my `~/.prefect/profiles.toml` file and so it was causing some really weird behaviour (this is because I had been using

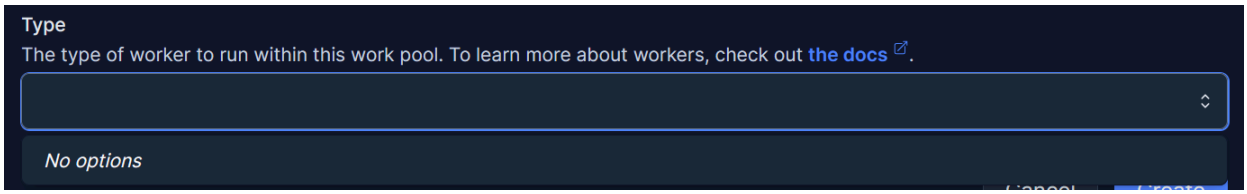
app.prefect.cloud during the data eng zoomcamp). I just deleted that line from profiles.toml and everything worked as it should.

Solution: Ideally you will have multiple profiles - one pointing to your local server if you want to use that and one pointing to a Cloud workspace. You can create a second profile with `prefect profile create my_cloud_profile` and then `prefect profile use my_cloud_profile`. Then set the API Key and API URL as needed for that profile (`prefect cloud login` should do the trick).

Added by Ming Jun, Asked by waleed, Answered by Jeff Hale

Type of Worker not available while creating the worker pool from the UI

Problem: The dropdown in the Type of Worker while creating the worker pool from the UI is empty. What is the cause of the problem?



ImportError: cannot import name 'SecretField' from 'pydantic'

pip intstall pydantic==1.10.0

Pydantic changed to v2 and older Prefect versions are not compatible.

Module 3 2022 (old)

Blank screen when running `prefect orion start`

Problem: you have a Windows computer, and when you run Prefect, the UI is blank.

Check the solution here:
https://github.com/DataTalksClub/mlops-zoomcamp/blob/main/03-orchestration/windows.md

(updated by subodh.chhabra@outlook.com)

Clean irrelevant files in repo

If one needs to remove unnecessary files (e.g. from last week), and would like to pull again, but while doing it is getting the below:

error: Your local changes to the following files would be overwritten by merge:
02-experiment-tracking/homework/hpo.py

Then, one should run the following first:

git clean -fdx (see here)

And then can “git pull” to pull the latest changes from origin github repo

Added by Daniel Hen (daniel8hen@gmail.com)

Conda environment didn’t load the correct python

Conda environment couldn’t load the correct python, even though I had specified python=3.9 when creating it.

Solution-

Sometimes conda can pick the python from the “base” conda environment when visual studio starts and automatically activates any conda environment other than the base environment.

Run `conda deactivate` twice and then activate the required conda environment.

Check python version by running `python -V`

Added by Oshan Modi

Cron expression

How to generate it?

Convenient cron expression generator

https://crontab.guru/

Added by Sergio Zemskov

Cronjobs, timezone and CronSchedule

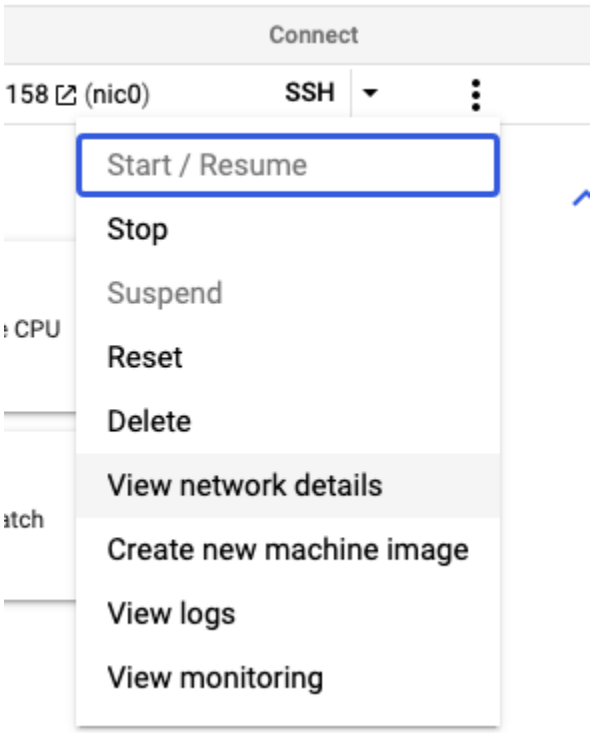
Cron jobs always run on local time and as part of the CronSchedule class, the timezone has to be provided. Prefect internally uses pendulum library for managing datetime stamps. The list of valid timezones are available at https://en.wikipedia.org/wiki/List_of_tz_database_time_zones. This is crucial while deploying in remote machines which might be in different timezones, especially if the execution and data servers are not colocated.

Added by Senthilkumar Gopal

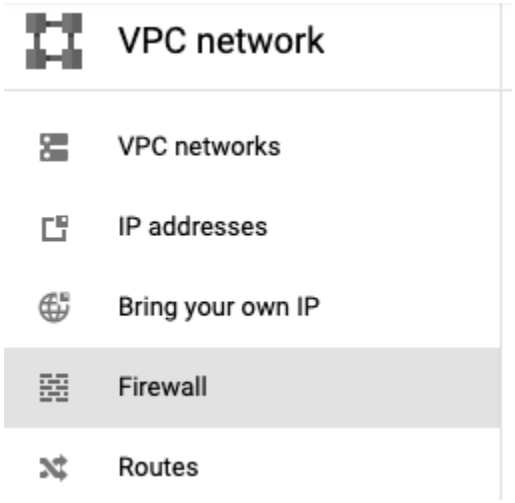
Enabling port 4200 access on GCP instance

Solution - Enabling access to a port on gcp requires setting up a new firewall rule. You can do so by following these steps -

1. Go to “View network settings” by clicking on the three dots next to gcp instance.

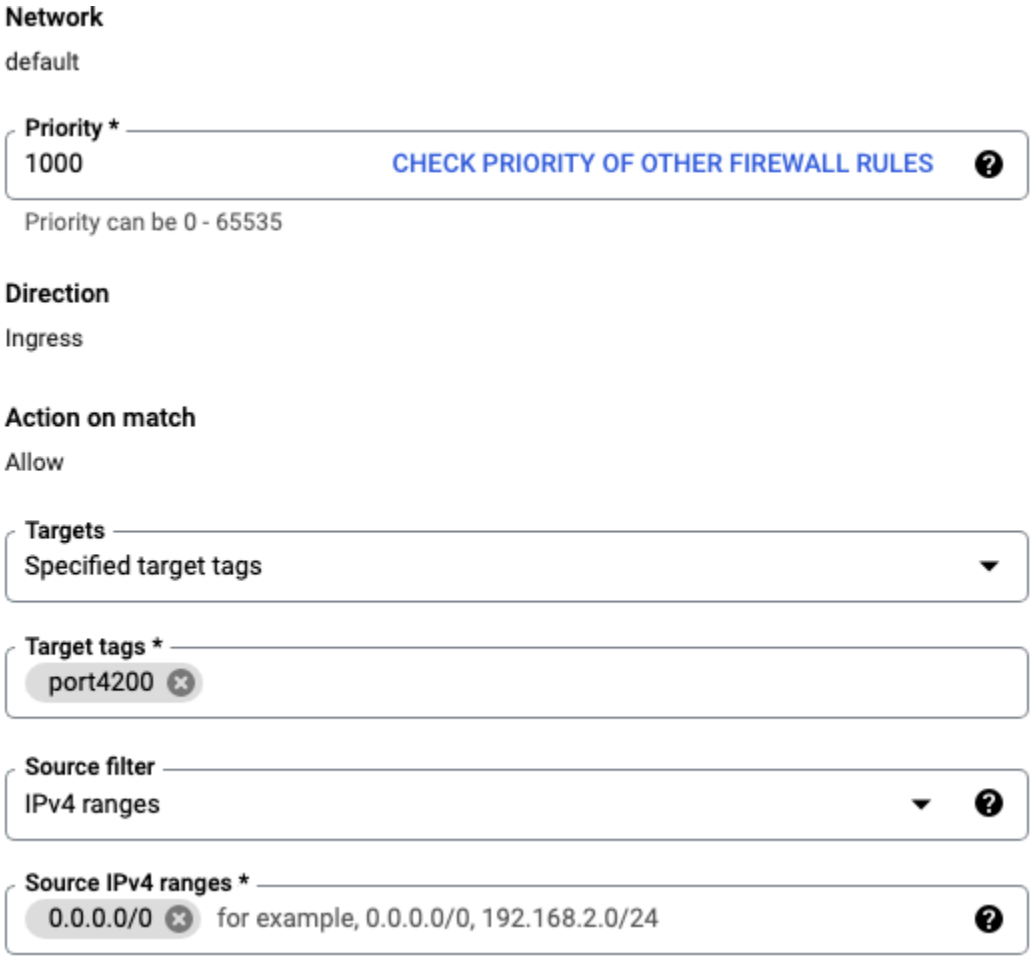


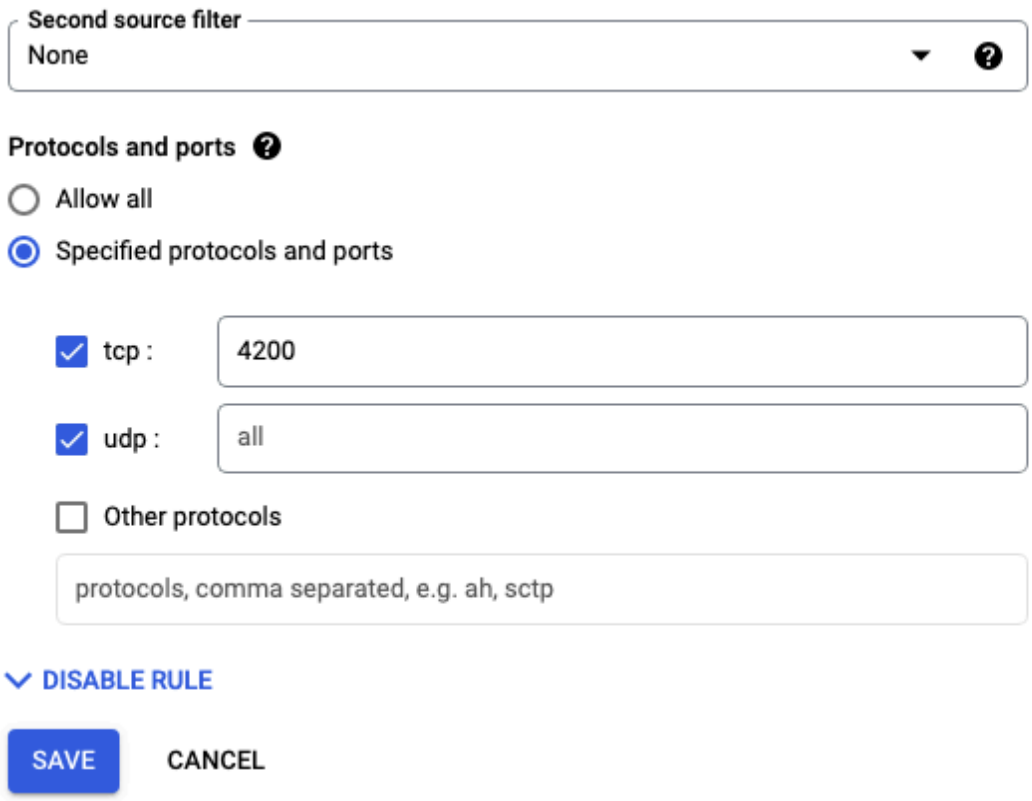
2. Select “Firewall” from left pane and then select “Create firewall rule”



3. Enter a rule name and a target tag (this would be used to implement this rule for your gcp instance).

4. Fill and save the rest of the details as per the screenshots below.





5. Return to your VM instance and click edit. Change firewall settings to accept ‘http’ and ‘https’ traffic.
6. In the network tags field, enter the target tag name you used in step 3.

Added by [Oshan Modi]

Error in optimizing function

2022/05/31 04:17:11 WARNING mlflow.utils.autologging_utils: Encountered unexpected error during sklearn autologging: The following failures occurred while performing one or more logging operations: [MlflowException('Failed to perform one or more operations on the run with ID 952bc0a59b7d4f8e95a3e06d992ebe42. Failed operations: [MlflowException("Changing param values is not allowed. Params were already logged=\'[{\'key\': \'bootstrap\', \'old_value\': None, \'new_value\': \'True\'}, {\'key\': \'ccp_alpha\', \'old_value\': None, \'new_value\': \'0.0\'}, {\'key\': \'criterion\', \'old_value\': None, \'new_value\': \'squared_error\'}, {\'key\': \'max_features\', \'old_value\': None, \'new_value\': \'auto\'}, {\'key\': \'max_leaf_nodes\', \'old_value\': None, \'new_value\': \'None\'}, {\'key\': \'max_samples\', \'old_value\': None, \'new_value\': \'None\'}, {\'key\': \'min_impurity_decrease\', \'old_value\': None, \'new_value\': \'0.0\'}, {\'key\': \'min_weight_fraction_leaf\', \'old_value\': None, \'new_value\': \'0.0\'}, {\'key\': \'n_jobs\', \'old_value\': None, \'new_value\': \'None\'}, {\'key\': \'oob_score\', \'old_value\': None, \'new_value\': \'False\'}, {\'key\': \'verbose\', \'old_value\': None, \'new_value\': \'0\'}, {\'key\': \'warm_start\', \'old_value\': None, \'new_value\': \'False\'}]\' for run ID=\'952bc0a59b7d4f8e95a3e06d992ebe42\'.")]')]

Solution - could not find a solution. I will appreciate it if you can please help.

– Need to use sequential task runner., this error points to duplicate parameter being logged on same run in MLflow, caused by default task runner, as concurrent tasks get executed

Updated by Subodh Chhabra(subodhchhabra@outlook.com)

Error when importing prefect modules

Error raise: AttributeError: 'NoneType' object has no attribute 'message_types_by_name'

The problem is related to the protobuf module. After updating from 3.17.2 to 3.20.0 it worked.

Added by Marcos J.

Flow Location Error with DeploymentSpec

I am getting an error when trying to create a deployment (see screenshot).

```
]: !prefect deployment create "homework_prefect_flow_q4.py"

Loading deployment specifications from python script at
'homework_prefect_flow_q4.py'...
Script at 'homework_prefect_flow_q4.py' encountered an exception
Traceback (most recent call last):
  File "C:\Users\Anna
V\Documents\GitHub\mlops_zoomcamp\homework\week3\homework_prefect_flow_q4.py",
line 122, in <module>
    DeploymentSpec(
  File
"C:\ProgramData\Anaconda3\envs\mlops\lib\site-packages\prefect\deployments.py",
line 133, in __init__
    super().__init__(**data)
  File "pydantic\main.py", line 341, in pydantic.main.BaseModel.__init__
pydantic.error_wrappers.ValidationError: 1 validation error for DeploymentSpec
flow_location
  str type expected (type=type_error.str)

Failed to load specifications from 'homework_prefect_flow_q4.py'
```

My Deployment Spec looks like this (see screenshot) and the name of the function is main_flow.

```
DeploymentSpec(
    flow_location=main_flow,
    name="scheduled-deployment-q4",
    flow_runner=SubprocessFlowRunner(),
    schedule=CronSchedule(cron="[redacted]")
)
```

Solution description : use the `flow` parameter if referencing the function, use `flow_location` is referencing a script (hence a string)

Added by Anna Vasylytsya

How to decode a cron expression easily

Given a cron expression. How to interpret it?

A cron expression made of five fields which represent the time to execute a given action (instruction or command)

Time is decoded using the following table:

```
# ┌──────────────── minute (0 - 59)
# │ ┌──────────────── hour (0 - 23)
# │ │ ┌──────────────── day of the month (1 - 31)
# │ │ │ ┌──────────────── month (1 - 12)
# │ │ │ │ ┌──────────────── day of the week (0 - 6) (Sunday to Saturday;
# │ │ │ │ │                                    7 is also Sunday on some systems)
# │ │ │ │ │
# │ │ │ │ │
# * * * * * <command to execute>
```

Added by Mario Recalde

Logging info in Prefect

Problem: How to log the info in prefect?
Solution:

```
from prefect import logging
logger = logging.get_run_logger("prefect")
logger.info(‘’)
```

Added by Low Kim Hoe

No runs found

After running prefect work-queue preview <queue_id>
there is an error: No runs found - try increasing how far into the future you preview with the --hours flag

Just increase the observed time horizon via adding an extension of the command:
prefect work-queue preview <queue_id> --hours 24*15
Note: <queue_id> is an ID of your queue, which you copied from Prefect Dashboard / Work Queue

Added by Ekaterina Krutman

OverFlow Error: Files too large when reading parquet

Reinstall Pyarrow with pip in the environment

pip install pyarrow

You can specify additional parameter in read_parquet function
read_parquet(path,engine='pyarrow')

Added by Ridwan Amure

How do we pass arguments into main when using prefect deployment start <program name>

Added by Raghav Menon

Parametrizing Prefect flows with date

To get an equivalent of `{{ data_interval_start }}` from Airflow, but in Prefect, you need to use the context:

```
from prefect import flow, task
from prefect.context import get_run_context

@flow
def hello_flow(date: datetime = None):
    if date is None:
        ctx = get_run_context()
        date = ctx.flow_run.expected_start_time

    result = hello_task(date)
```

Permission denied for storage folder

I am getting this error when I run the creation of the deployment (**prefect deployment create prefect_deploy.py**). Previously I have created the storage folder as following (on the AWS instance) :

- prefect storage create
- Option : 3
- Path : /ubuntu/.prefect
- Name : Local

```
(base) ubuntu@ip-172-31-1-189:~/workplace/mlops-zoomcamp/week3-orchestration$ prefect deployment creat
e prefect_deployment.py
Loading deployment specifications from python script at
'prefect_deployment.py'...
/home/ubuntu/anaconda3/lib/python3.9/site-packages/prefect/deployments.py:247: UserWarning: You have c
onfigured local storage, this deployment will only be usable from the current machine..
  warnings.warn(
Creating deployment 'model_training' for flow 'main'...
Deploying flow script from '/home/ubuntu/workplace/mlops-zoomcamp/
week3-orchestration/prefect_deployment.py' using Local Storage...
Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.9/pathlib.py", line
1323, in mkdir
    self._accessor.mkdir(self, mode)
FileNotFoundError: [Errno 2] No such file or directory:
'/ubuntu/.prefect'

During handling of the above exception, another exception
occurred:

Traceback (most recent call last):
  File "/home/ubuntu/anaconda3/lib/python3.9/site-packages/prefect
/cli/deployment.py", line 250, in create
    await spec.create_deployment(validate=False)
  File "/home/ubuntu/anaconda3/lib/python3.9/site-packages/prefect
/client.py", line 95, in with_injected_client
    return await fn(*args, **kwargs)
  File "/home/ubuntu/anaconda3/lib/python3.9/site-packages/prefect
/deployments.py", line 302, in create_deployment
    storage_token = await self.flow_storage.write(flow_bytes)
  File "/home/ubuntu/anaconda3/lib/python3.9/site-packages/prefect
/blocks/storage.py", line 239, in write
    storage_dir.mkdir(parents=True, exist_ok=True)
  File "/home/ubuntu/anaconda3/lib/python3.9/pathlib.py", line
1327, in mkdir
    self.parent.mkdir(parents=True, exist_ok=True)
  File "/home/ubuntu/anaconda3/lib/python3.9/pathlib.py", line
1323, in mkdir
    self._accessor.mkdir(self, mode)
PermissionError: [Errno 13] Permission denied: '/ubuntu'

Failed to create deployment 'main/model_training'
Failed to create 1 out of 1 deployments.
```

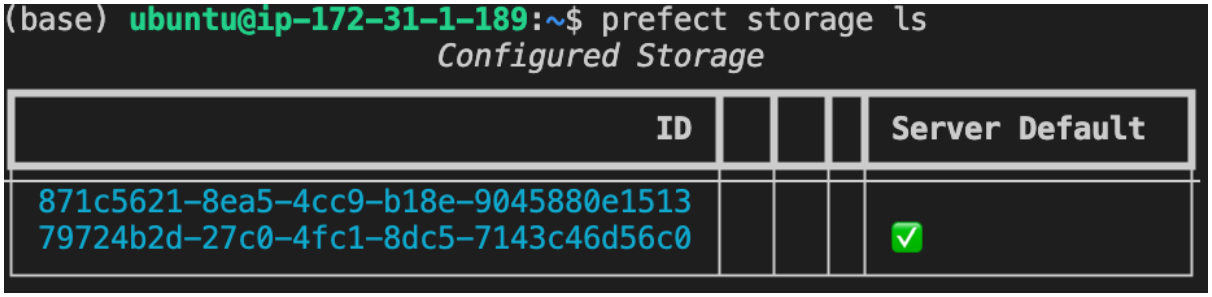
Solution description : Prefect allows us to create the folder (/ubuntu/) in this location but it is a location where we do not have permission.The solution is to create a new folder in a suitable location and change the prefect configuration to point to this new storage folder :

- Repeat the creation of the storage folder by choosing a suitable folder (/ubuntu/mlops-zoomcamp/week3)
- prefect storage ls (you can see now two folders)

```
(base) ubuntu@ip-172-31-1-189:~$ prefect storage ls
                    Configured Storage
```

ID				Server Default
871c5621-8ea5-4cc9-b18e-9045880e1513 79724b2d-27c0-4fc1-8dc5-7143c46d56c0		—		✅

- Change the default storage location : **prefect storage set-default** 79724b2d-27c0-4fc1-8dc5-7143c46d56c0
- Check that the change is taken into account



Now run again **prefect deployment create prefect_deploy.py** and the deployment is now created correctly.

Added by [Panayotis Papoutsis]

Prefect and MLflow on the same VM

In production, is it ok to host mlflow and prefect on the same vm?

Ideally, in production you will run them on kubernetes in different containers which may or may not be on the same node.

Added by Oluwatosin Olajide

Prefect Orion Installation on MacOS 10.14 or below, unable to find installation candidates for ray

Prefect Orion needs ray >= 1.9 if you are installing it on Python 3.9 or lower. The ray package is only available for MacOS 10.15 or above.

On MacOS 10.14 or below you could install Python 3.10 to avoid installing ray. Python 3.10 is compatible with MacOS 10.9 or above.

Running Prefect Flows using Prefect Cloud causes “400 bad request” Error

Make sure the latest pre-release version of Prefect is installed.

Added by Mohammed Ayoub Chettouh

Running Prefect Orion on a different port

Let's say you want to run prefect on port 4202.

This is what you need to do. On the server:

```
prefect config set PREFECT_ORION_UI_API_URL=http://localhost:4202/api
prefect orion start --port=4202
```

In the first line, you configure the URL for the UI - the UI will use this setting to fetch data from the backend (the /api part at the end is important)

The second line starts the server on port 4202

Now on the client, you need to configure the URL too:

```
prefect config set PREFECT_API_URL=http://127.0.0.1:4202/api
```

That's it - you're ready to go.

Script as task

If we have a script that preprocesses data and we have another script that trains the model. The processing script definitely has to run before the training script. Does prefect allow you to run a script instead of a function as a task? If it does, how do you configure the flow?

A whole script can be stuck in "@flow" and it will work off the shelf if you don't even need a task. But in the case you have multiple scripts, python affords you different methods of doing it. You can:

- Wrap the code in each script in a function, import it into a script and run it the way Kevin demonstrated in the video.
- Bring all the code into the same script and run it as Kevin demonstrated.
- Or you can create a task that will run each of the files in order.

Added by Oluwatosin Olajide

TypeError: cannot unpack non-iterable PrefectFuture object / How to connect isolated tasks in the Radar UI

Problem: Suppose a task returns a tuple of Python objects

```
@task
def process(train_path, valid_path):
     ...
    return X_train, y_train, X_valid, y_valid

@task
def add_features(X_train, y_train, X_valid, y_valid):
     ...
    return X_train, y_train, X_valid, y_valid

@flow
def main(train_path, valid_path):
    X_train, y_train, X_valid, y_valid = process(train_path, valid_path)
    X_train, y_train, X_valid, y_valid = add_features(X_train, y_train, X_valid,
y_valid)
    ...
```

Gives the above error. A solution in the lecture is to use `X_train, y_train, X_valid, y_valid = process(train_path, valid_path).result()`. However, note that by doing this we will get isolated tasks in the Radar UI as shown in the lectures where the three tasks have no dependency lines between them. This is because the lineage has been lost after converting to Python objects from a PrefectFuture object using the result method. Ideally, a future object has to be passed from task to task to track the task dependency. Thus, if we want radar lines we have to do:

```
@task
def process(train_path, valid_path):
     ...
    return X_train, y_train, X_valid, y_valid

@task
def add_features(future):
```

```
    X_train, y_train, X_valid, y_valid = future
    # No need to add `.result()`, with @task decorator future already
    # converted to Python object inside task.
    …

    return X_train, y_train, X_valid, y_valid

@flow
def main(train_path, valid_path):
    future = process(train_path, valid_path)
    future = add_features(future)
    … # And so on
```

Added by Ron Medina

ValueError: the greenlet library is required to use this function

ValueError: the greenlet library is required to use this function. (mach-o file, but is an incompatible architecture (have (x86_64), need (arm64e)))

This is a problem on M1/M2 Macs.

Solution: For some reason, the greenlet library is not properly installed when installing Prefect via pip. Solve it by uninstalling the package with pip (pip uninstall greenlet), followed by installing it with conda instead (conda install greenlet). Assuming that you're using conda environments.

Added by Jakob Salomonsson

TypeError: Not supported type for data.<class 'xgboost.core.DMatrix'

This error is encountered when you try to load xgboost model registered in Model registry using the following code:

model = pyfunc.load_model(f"runs:/{each_model.run_id}/models_mlflow")

Solution: The xgboost model pulled from model registry needs to be loaded using following method:

```
model = xgboost.load_model(f"runs:/{model_run_id}/models_mlflow")
```

Added by Amogh N Kulkarni