



Pantech e Learning
DIGITAL LEARNING SIMPLIFIED

Amazon Web Services

MLOps with AWS

Masterclass



Machine Learning Operations with AWS

Day -13



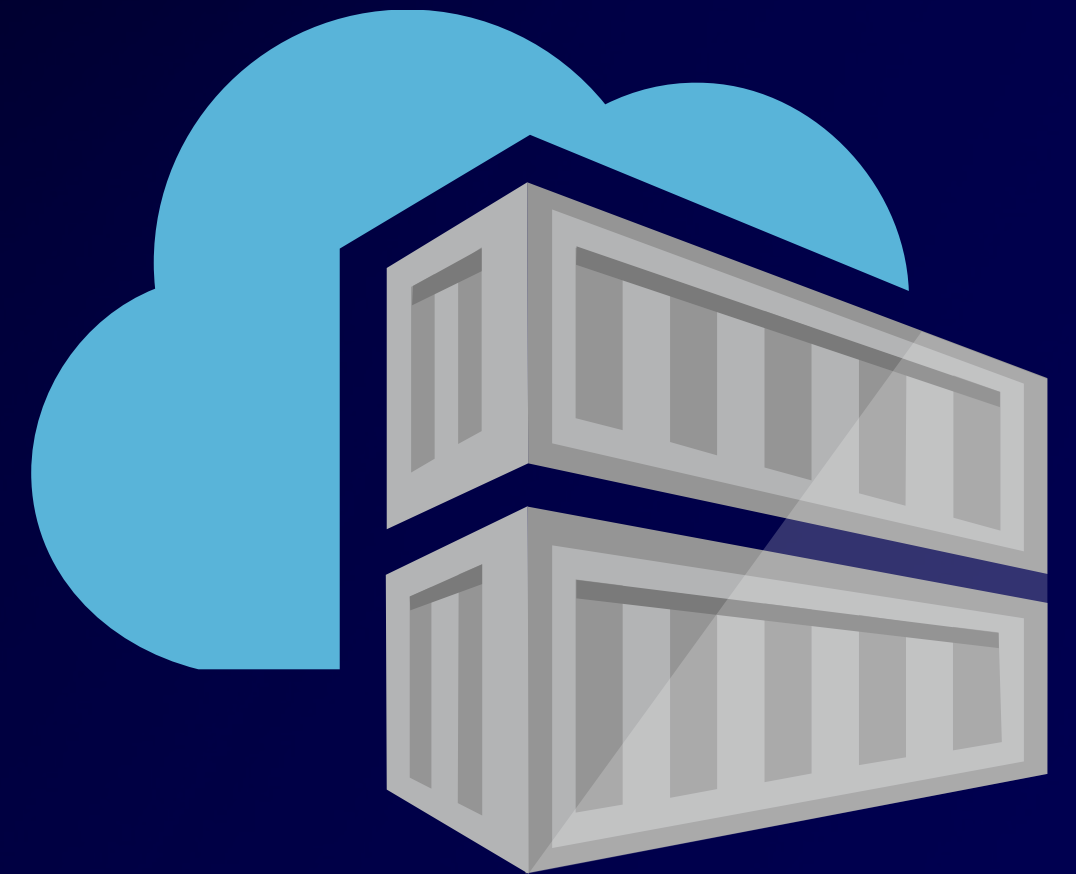
Amazon Web Services
MLOps with AWS

Masterclass



Pantech e Learning
DIGITAL LEARNING SIMPLIFIED


Sagemaker Container Processor



Sagemaker Container Processor

- SageMaker's Scikit-Learn Container is a pre-built environment provided by Amazon SageMaker that allows you to use the Scikit-Learn library for machine learning preprocessing tasks within the SageMaker ecosystem.
- It provides a convenient way to perform data preprocessing, feature engineering, and transformation steps on large datasets using Scikit-Learn's familiar APIs.

Configuring the notebook



```
import boto3

import sagemaker

from sagemaker import get_execution_role

region = boto3.session.Session().region_name

role = get_execution_role()
```

Sagemaker SKLearn Processor



```
from sagemaker.sklearn.processing import SKLearnProcessor

sklearn_processor = SKLearnProcessor/framework_version='0.20.0',

role=role,

instance_type='ml.m5.xlarge',

instance_count=1)
```

**Create the preprocessing code in
preprocessing.py**



Run the container

```
from sagemaker.processing import ProcessingInput,
ProcessingOutput

sklearn_processor.run(code='preprocessing.py',
                       inputs=[ProcessingInput(
                           source=input_data,
                           destination='/opt/ml/processing/input')],
                       outputs=[ProcessingOutput(output_name='train_data',
                                                  source='/opt/ml/processing/train',
                                                  destination='s3://slytherins-test/'),
                                ProcessingOutput(output_name='test_data',
                                                  source='/opt/ml/processing/test',
                                                  destination='s3://slytherins-test/')],
                       arguments=['--train-test-split-ratio', '0.1']
                       )
```


Getting information about the job



```
preprocessing_job_description = sklearn_processor.jobs[-1].describe()  
  
output_config = preprocessing_job_description['ProcessingOutputConfig']
```

Thank you



Pantech e Learning
DIGITAL LEARNING SIMPLIFIED