AWS Documentation

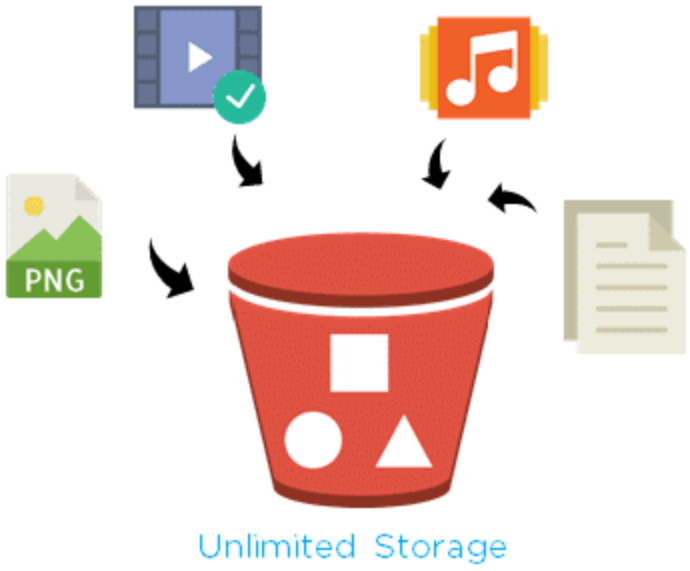
Author - Preeti Chatterjee
[This is created as a part of my Internship tenure at MinimumQue]

To understand the whole process-overview, we need to familiarise ourselves with a few terms and how they work. The list of such terms are:

- *Amazon S3*
- *Amazon Sagemaker*
- *Amazon Eventbridge*
- *AWS Lambda*

Let's first go through what each of these does:

- **Amazon s3:** Amazon S3 or Amazon Simple Storage Service is a service offered by Amazon Web Services (AWS) that provides object storage through a web service interface. Amazon S3 uses the same scalable storage infrastructure that Amazon.com uses to run its global e-commerce network. Amazon S3 can be employed to store any type of object, which allows for uses like storage for Internet applications, backup and recovery, disaster recovery, data archives, data lakes for analytics, and hybrid cloud storage.



Whatever training, testing data and model we train we save in sagemaker we'll also save in s3 bucket. Anytime we want to reference it, we can retrieve it because reading from s3 is good since it's scalable and we can hit any number of requests from this instance.

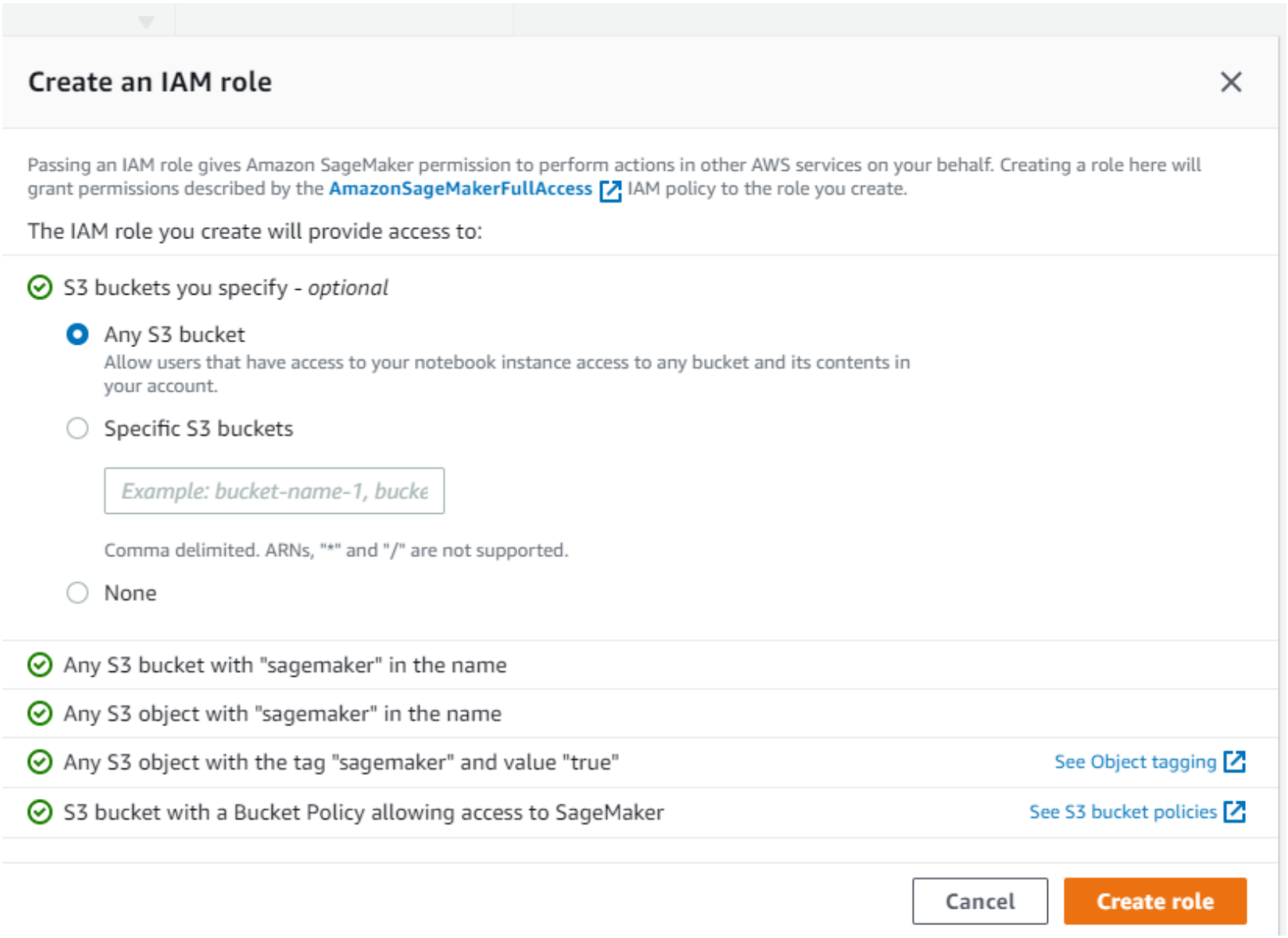
- **Amazon Sagemaker:** Amazon SageMaker is a cloud machine-learning platform that was launched in November 2017. SageMaker enables developers to create, train, and deploy machine-learning models in the cloud. SageMaker also enables developers to deploy ML models on embedded systems and edge-devices.

  Here we need to know about two terms:
  - ☐ *Notebook Instances*
  - ☐ *Lifecycle Configurations*

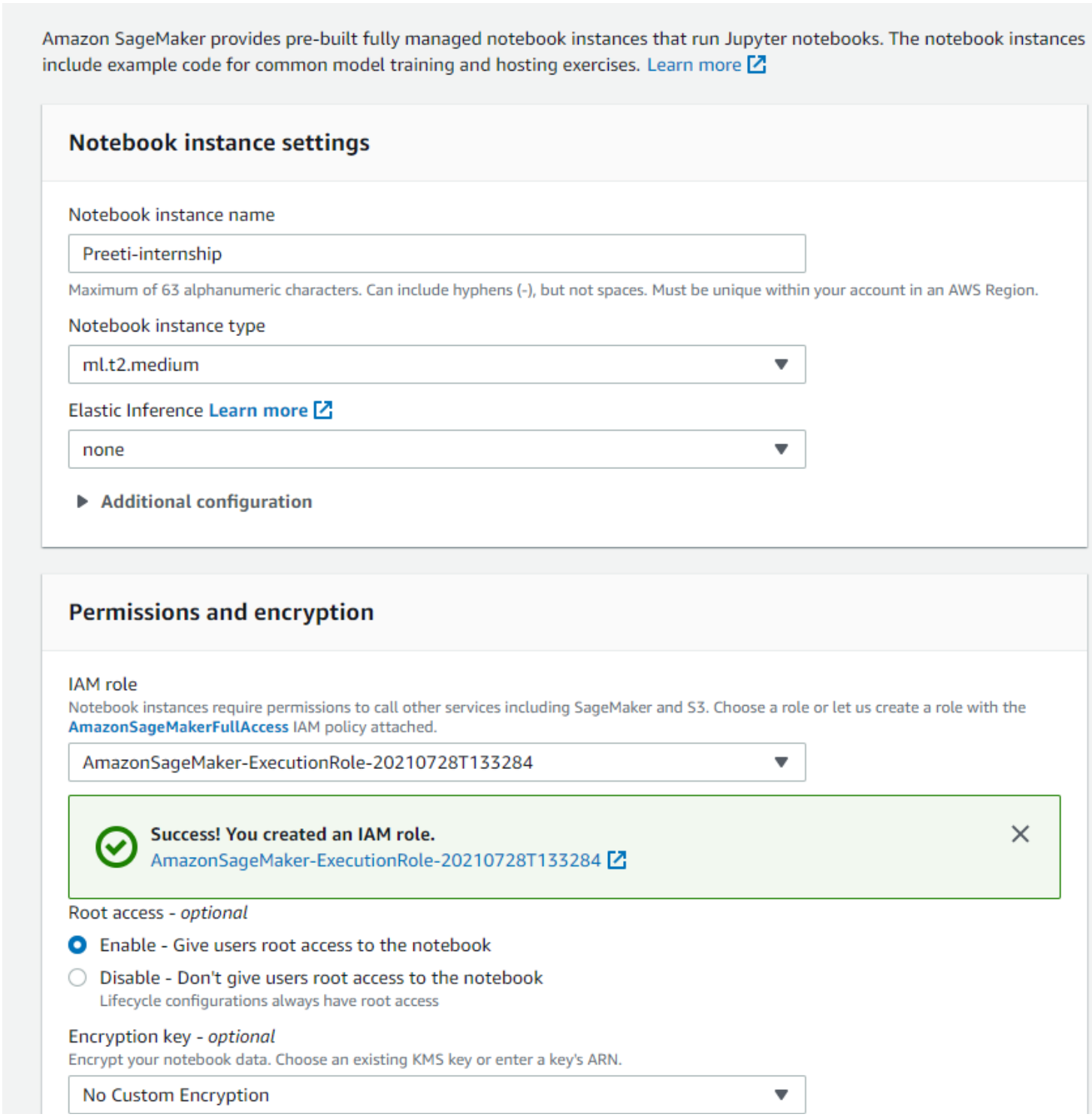
Notebook Instances:

To create a notebook instance in sagemaker we go to “Create notebook instance” in sagemaker.

After choosing notebook instance name, instance type as ml.t2.medium we create an IAM role.



After doing the above we would have successfully created an IAM role.
Note: We can also use an existing role if we already had one.

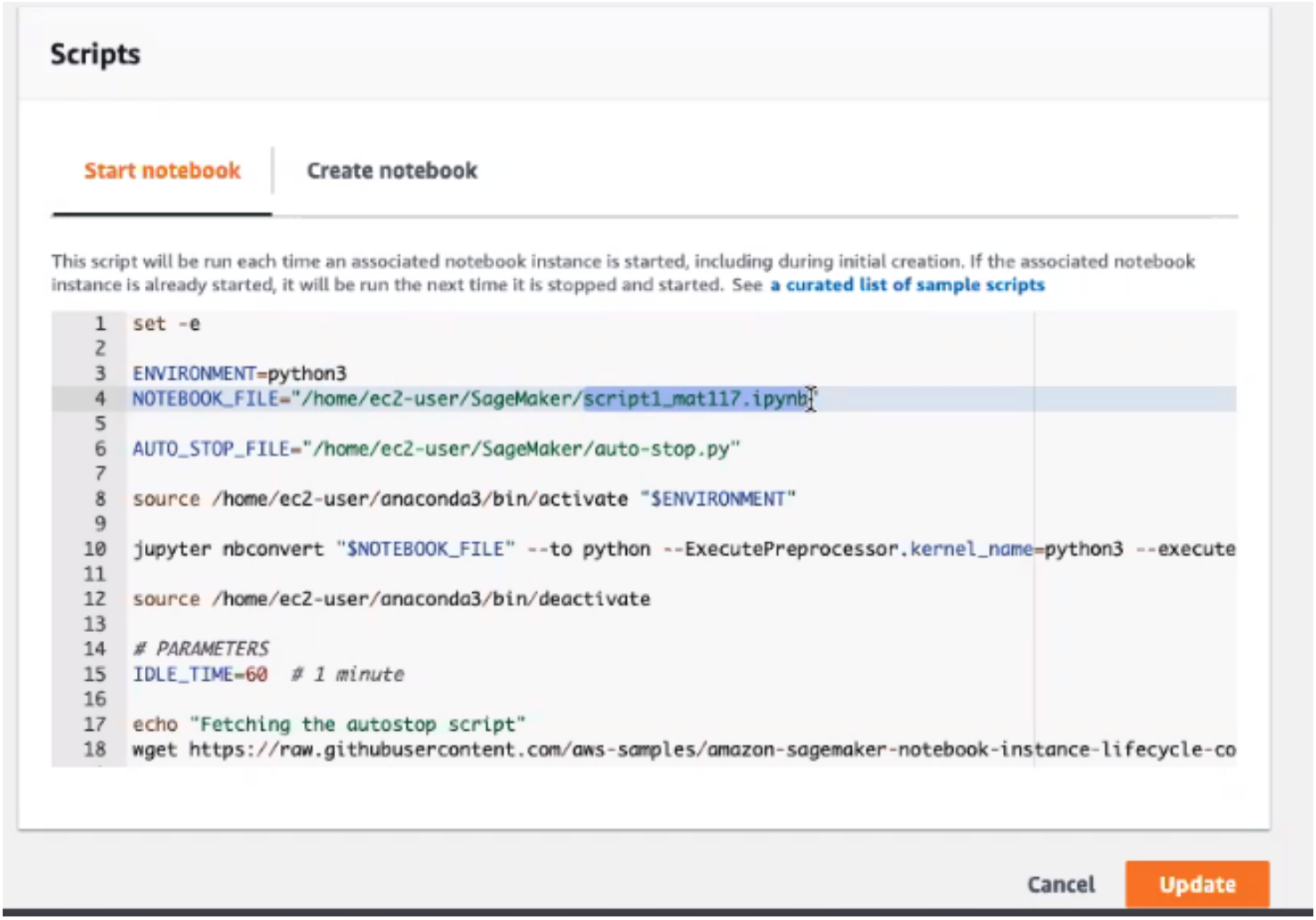


Then, we create a notebook instance and when the status is "InService" and open jupyter following that. **Our Notebook Instance is ready to be used now**.

Lifecycle Configurations:

Let us consider a scenario where we have scheduled our sagemaker notebook instance to run at a scheduled time. Now, since the notebook

instance started we don't want it to sit idle. So, we need some script to run as soon as the sagemaker instance starts.



We need to specify a few terms & concerns about the script here:

- ☐ NOTEBOOK_FILE - It runs the notebook file as soon as the sagemaker instance starts
- ☐ AUTO_STOP_FILE - This script stops our notebook instance without any manual intervention after IDLE_TIME
- ☐ IDLE_TIME - Parameter which decides how much time (in seconds) the sagemaker instance sits idle before it stops
- ☐ The more time our sagemaker instance runs, the more cost we incur

Studio Notebooks | On-Demand Notebook Instances | Processing | Data Wrangler | Feature Store | Training | Real-Time Inference | Batch Transform

Amazon SageMaker Studio Notebooks

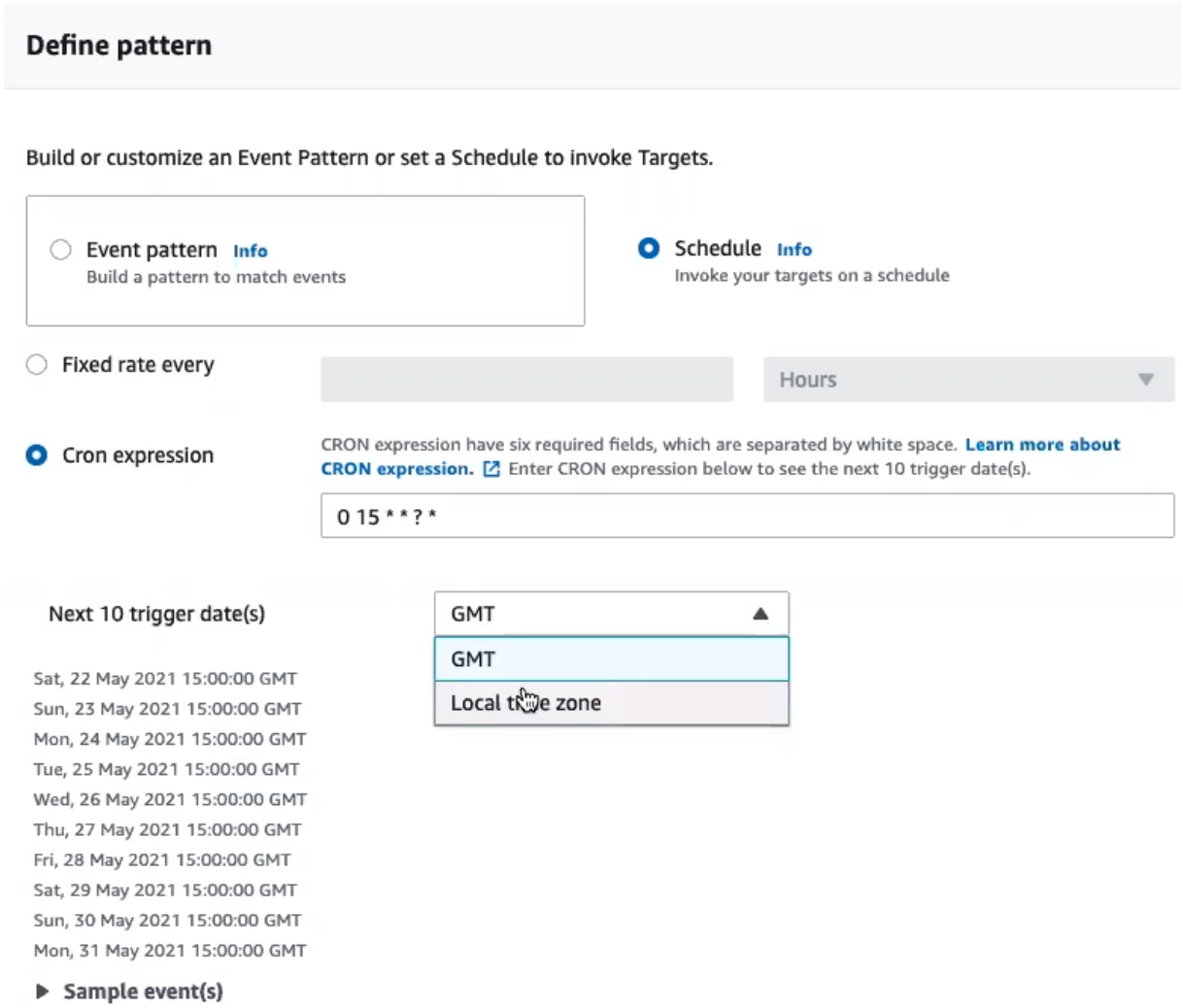
Amazon SageMaker Studio Notebooks are one-click Jupyter notebooks that can be spun up quickly. The underlying compute resources are fully elastic and the notebooks can be easily shared with others enabling seamless collaboration. You are charged for the instance type you choose, based on the duration of use.

Region: US West (Oregon)

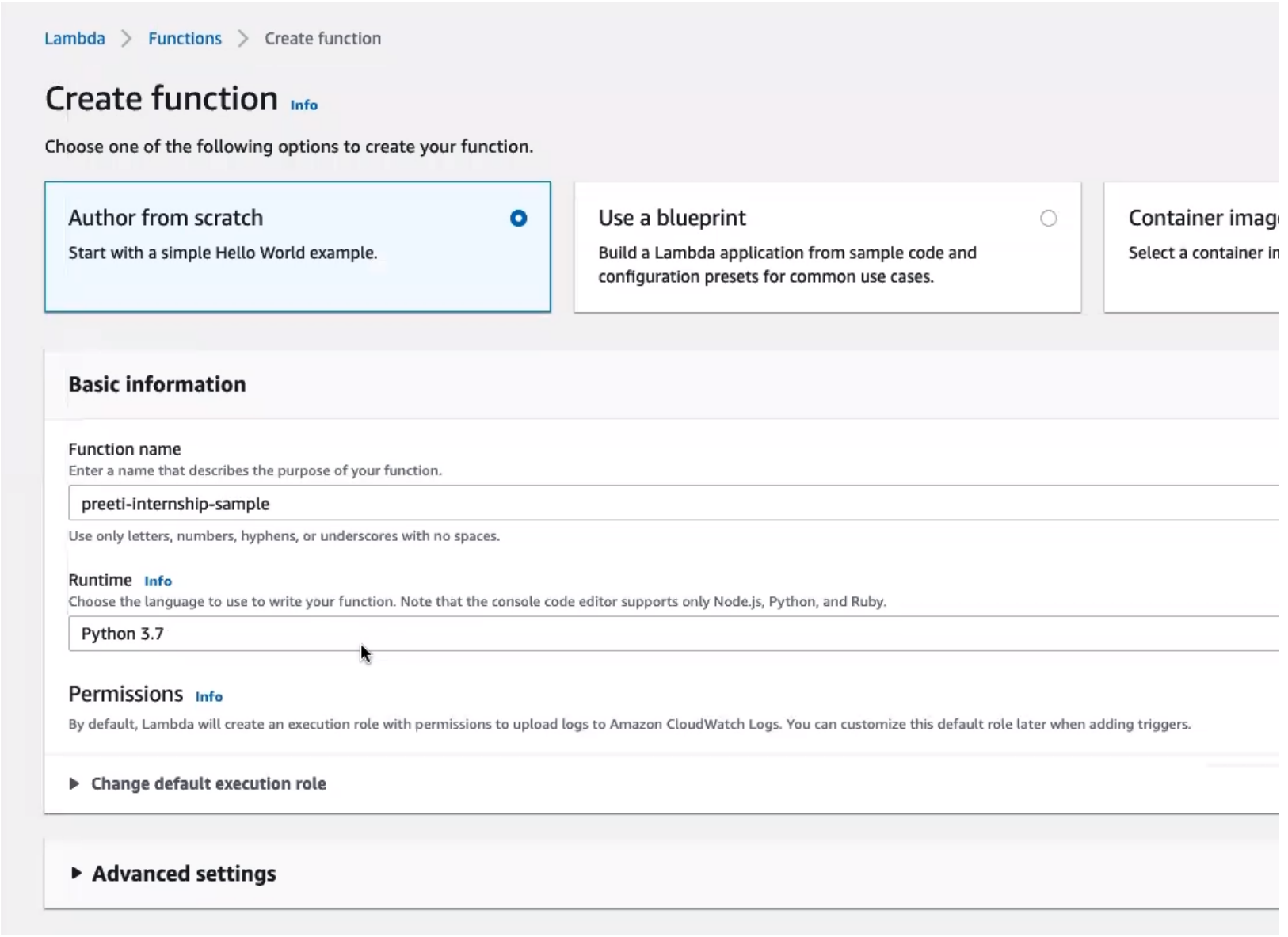
Standard Instances	vCPU	Memory	Price per Hour
ml.t3.medium	2	4 GiB	$0.05
ml.t3.large	2	8 GiB	$0.10
ml.t3.xlarge	4	16 GiB	$0.20
ml.t3.2xlarge	8	32 GiB	$0.399
ml.m5.large	2	8 GiB	$0.115
ml.m5.xlarge	4	16 GiB	$0.23
ml.m5.2xlarge	8	32 GiB	$0.461
ml.m5.4xlarge	16	64 GiB	$0.922
ml.m5.8xlarge	32	128 GiB	$1.843
ml.m5.12xlarge	48	192 GiB	$2.765
ml.m5.16xlarge	64	256 GiB	$3.686
ml.m5.24xlarge	96	384 GiB	$5.53

- **Amazon Eventbridge:** This helps us schedule any recurring tasks. To operate this, we have to create “Rules” under “Events” and then we *create a*

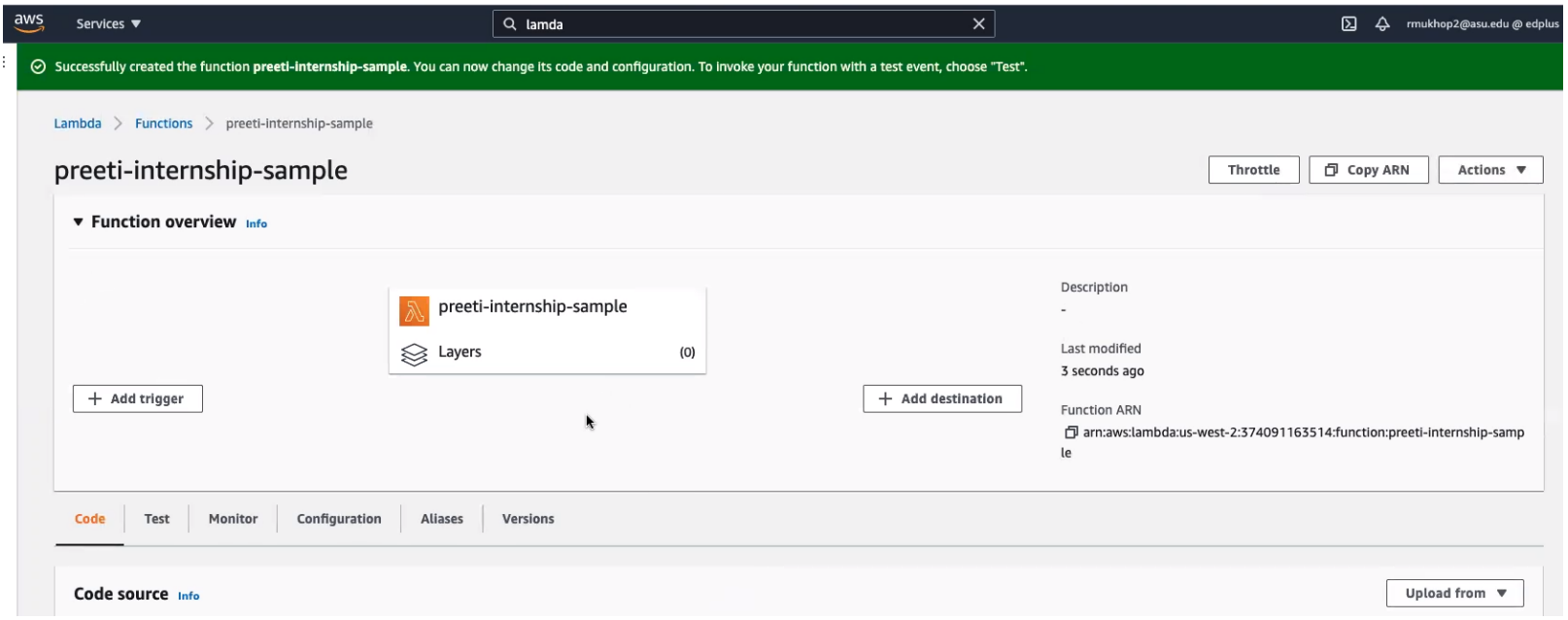
rule. While creating rules we have two options - either we can choose an **event pattern** or set a **schedule** to invoke targets. Event pattern implies that when something occurs then only the task gets triggered, for example we consider data deposits in s3 buckets as the trigger event. On the other hand, if we need our task to run at certain intervals of time using "Fixed rate every" or at a fixed time everyday using "Cron expression" [Image given below for reference], that would imply scheduling. **So, in conclusion, Amazon Eventbridge helps us schedule a particular task**.



- **AWS Lambda:** AWS Lambda is an event-driven, serverless computing platform provided by Amazon as a part of Amazon Web Services. It is a computing service that runs code in response to events and automatically manages the computing resources required by that code.
  To create a function, we go to Dashboard under AWS Lambda and click on "Create function". Then we choose the following and decide the IAM role from the admin.

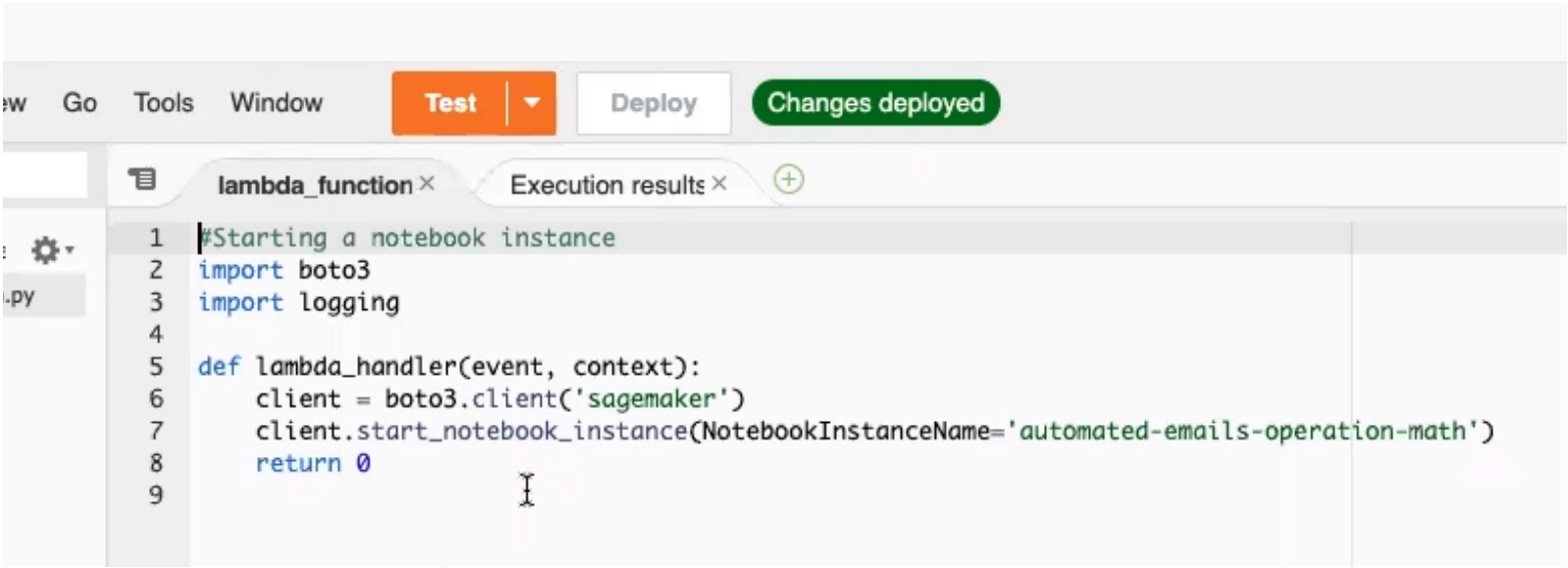


This creates our desired lambda function:



Lambda function named “preeti-internship-sample” created

Now, given below is an example code of how to start a notebook instance inside a lambda function:



A lambda function which starts a sagemaker notebook instance named “automated-emails-operation-math”

Bringing Everything Together

- Create a notebook instance
- Configure Lifecycle configurations and link it to notebook instance
- Creating relevant task using AWS Lambda to start the relevant notebook
- Setting a trigger by clicking on “Add trigger” option and choosing Amazon Eventbridge
- Create a new rule or use existing rules to schedule a trigger at a specific time or intervals of time