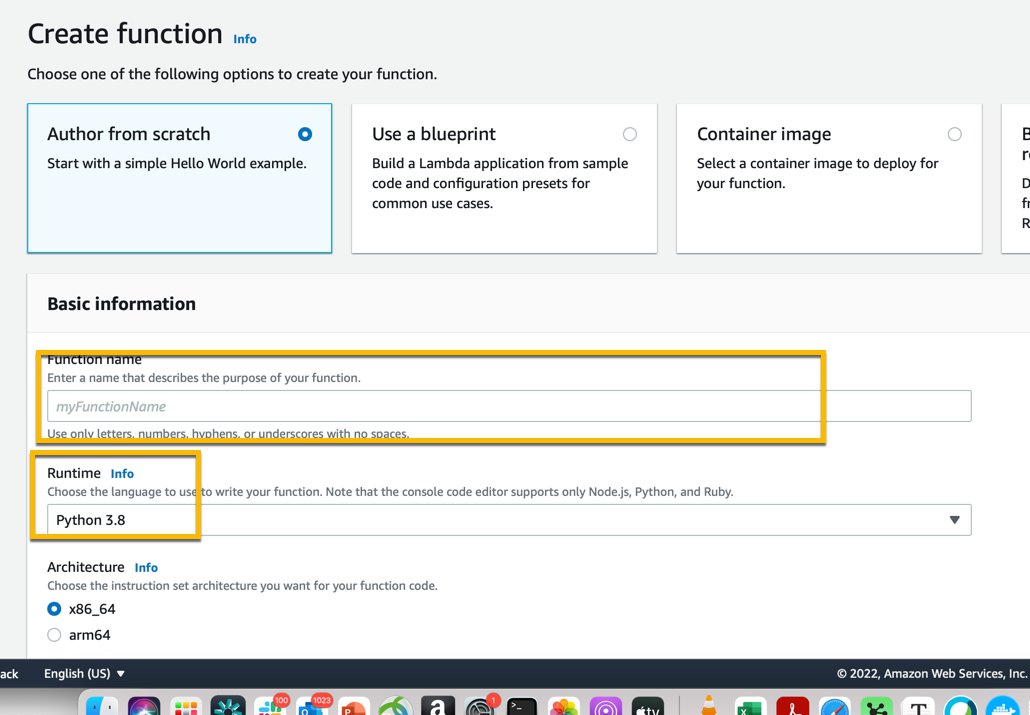
Inference With SageMaker Batch.

We will create 2 lambda functions, a SNS topic and provision IAM permissions to lambda functions.

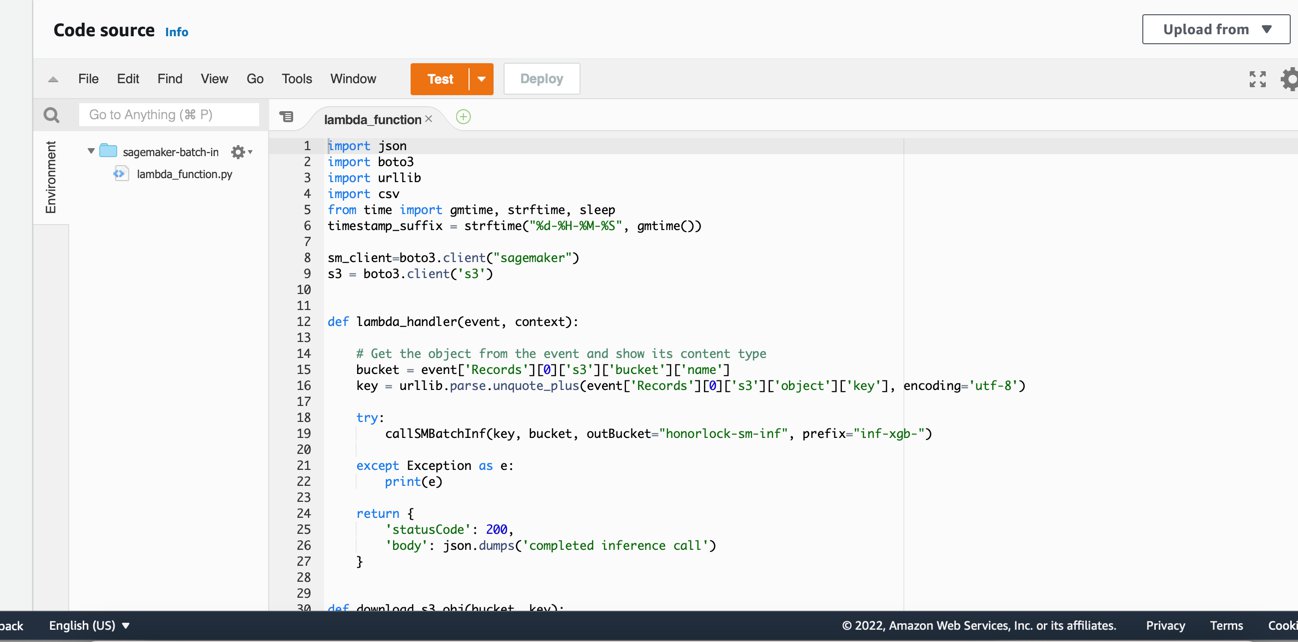
## Lambda Functions

### Lambda Function for starting SageMaker Batch Inference

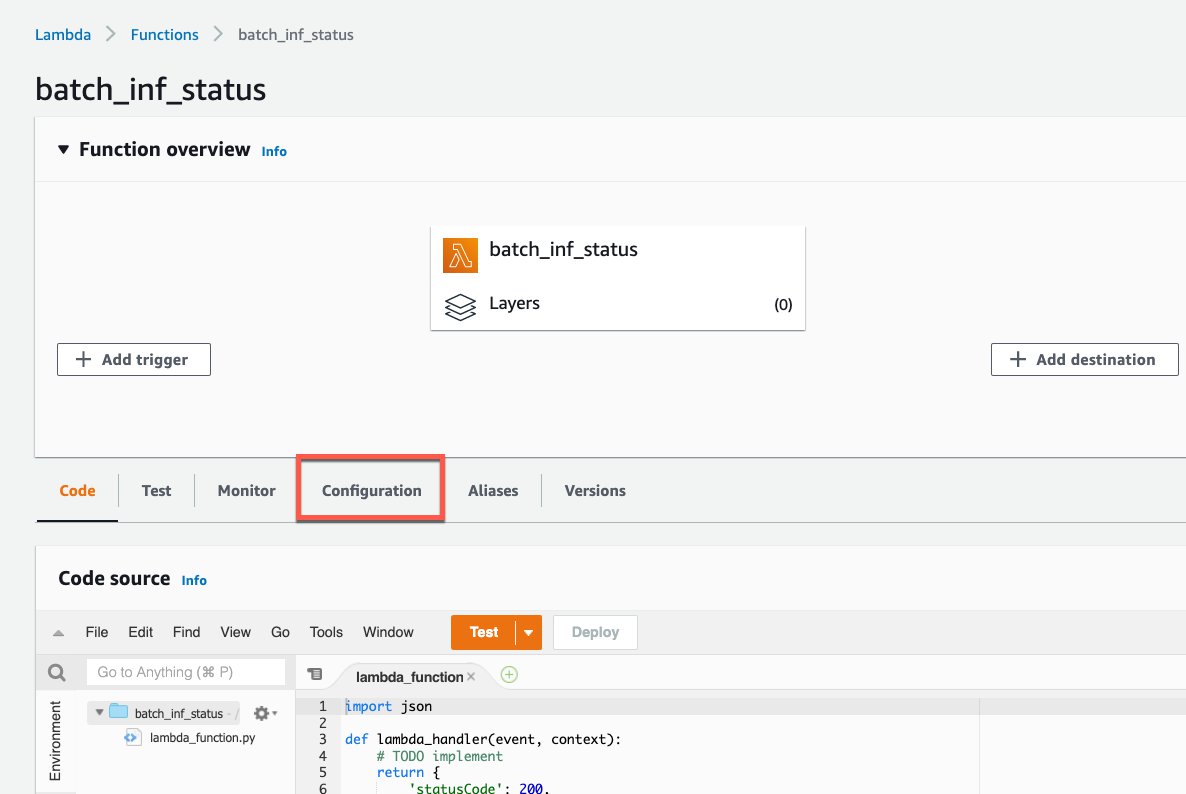
1. go to lambda console on aws console and create a new lambda function, select author your own and for language select python3.8

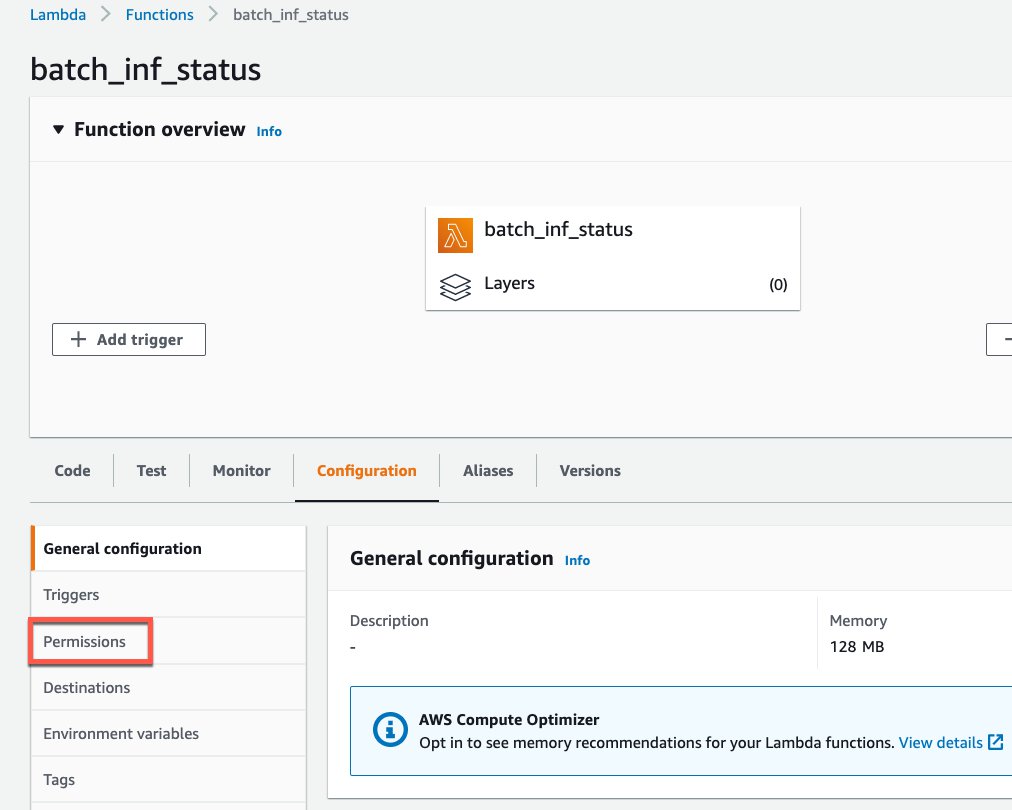


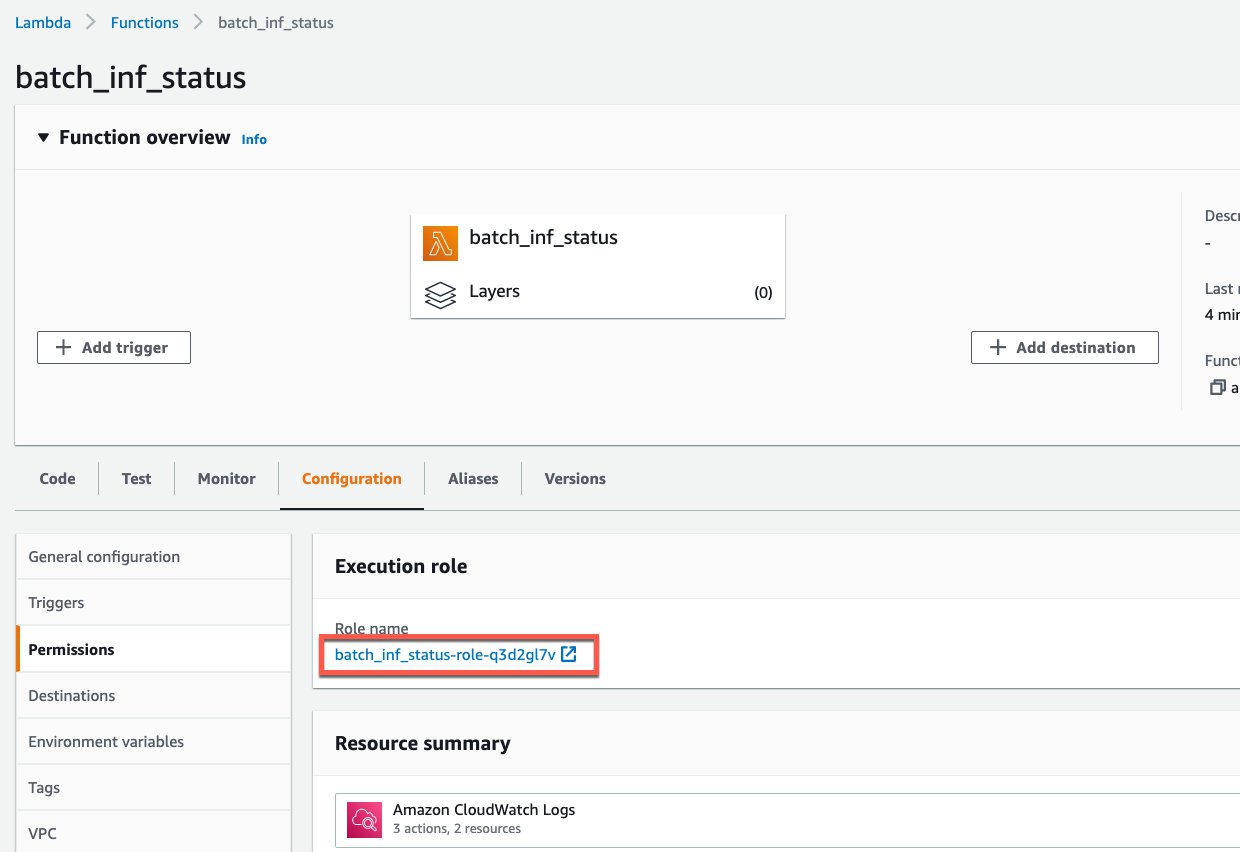
1. Once the function is created, go to code section of the lambda function and paste in the code sm-batch-inf.py
2. click File-->save.
3. Look at the code and change any setting , like naming conventions, logic to use the latest model as you see fit, One way to decouple that from code would be to have a config.json file stored in S3 that can be updated without touching the code. The config file can contain output file naming, model to use for inference or any other variables.

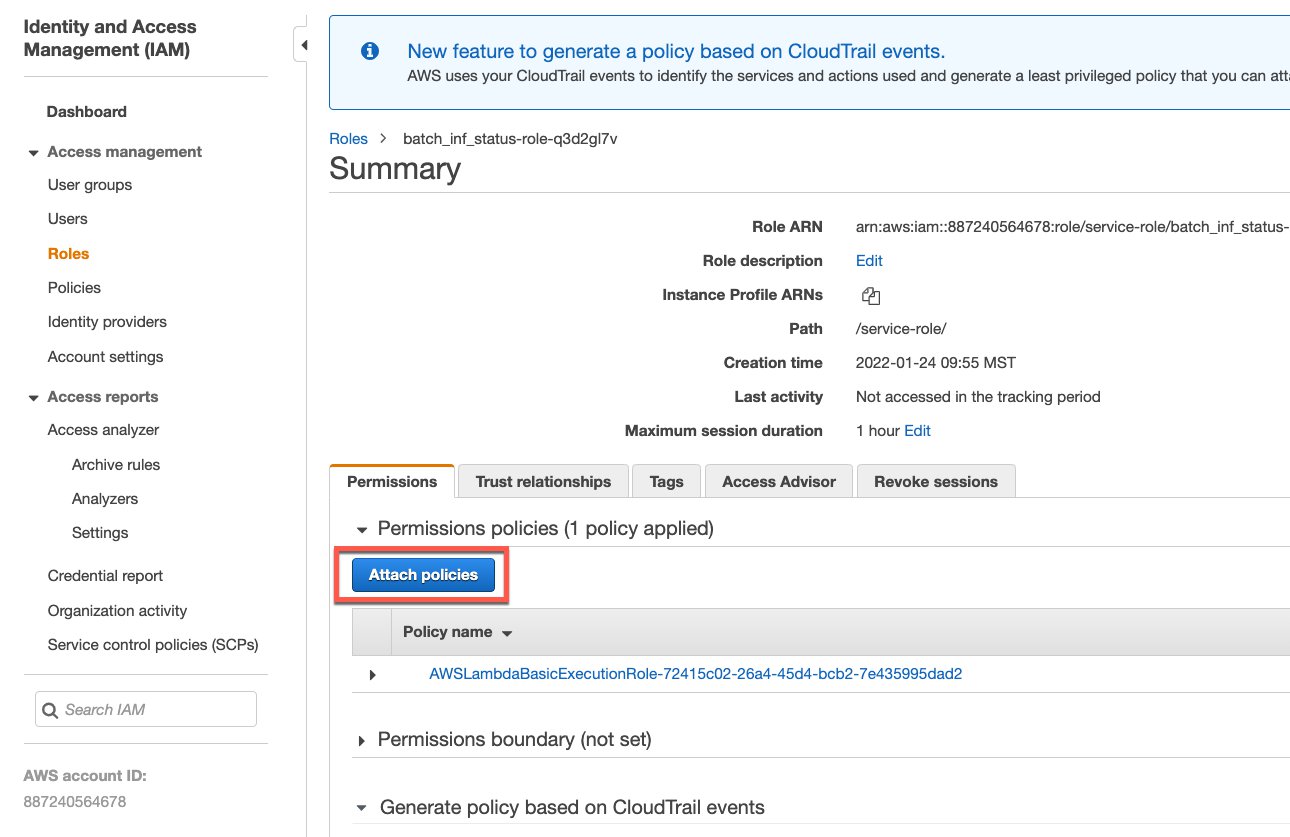


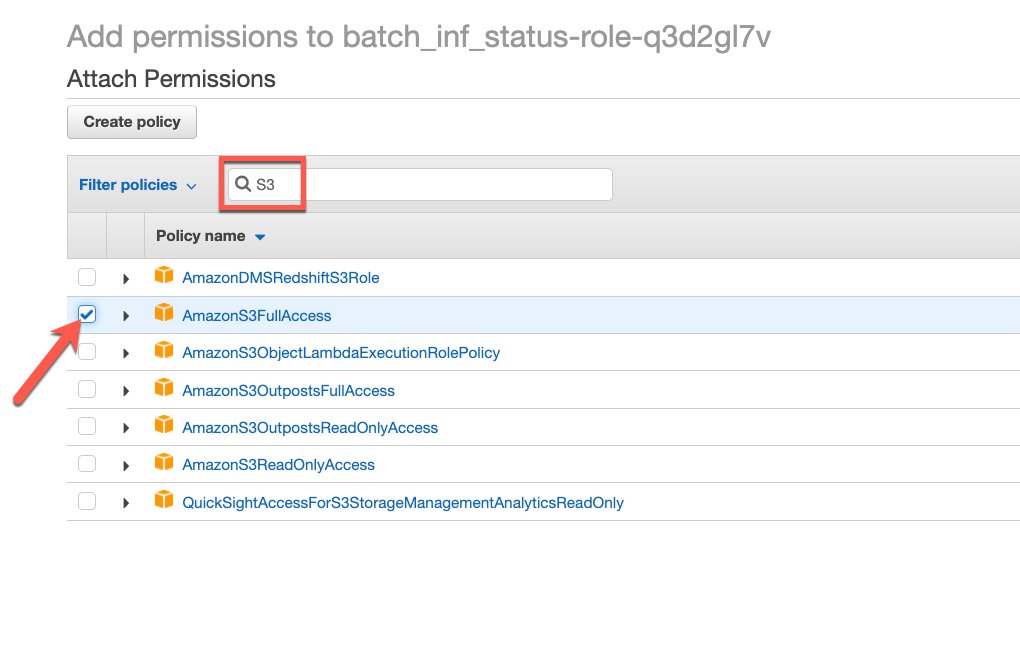
1. Next add a trigger to Lambda function, Choose one of the S3 buckets assigned as landing bucket for inference input.
2. Next we will update permissions so lambda can access Sagemaker and S3.

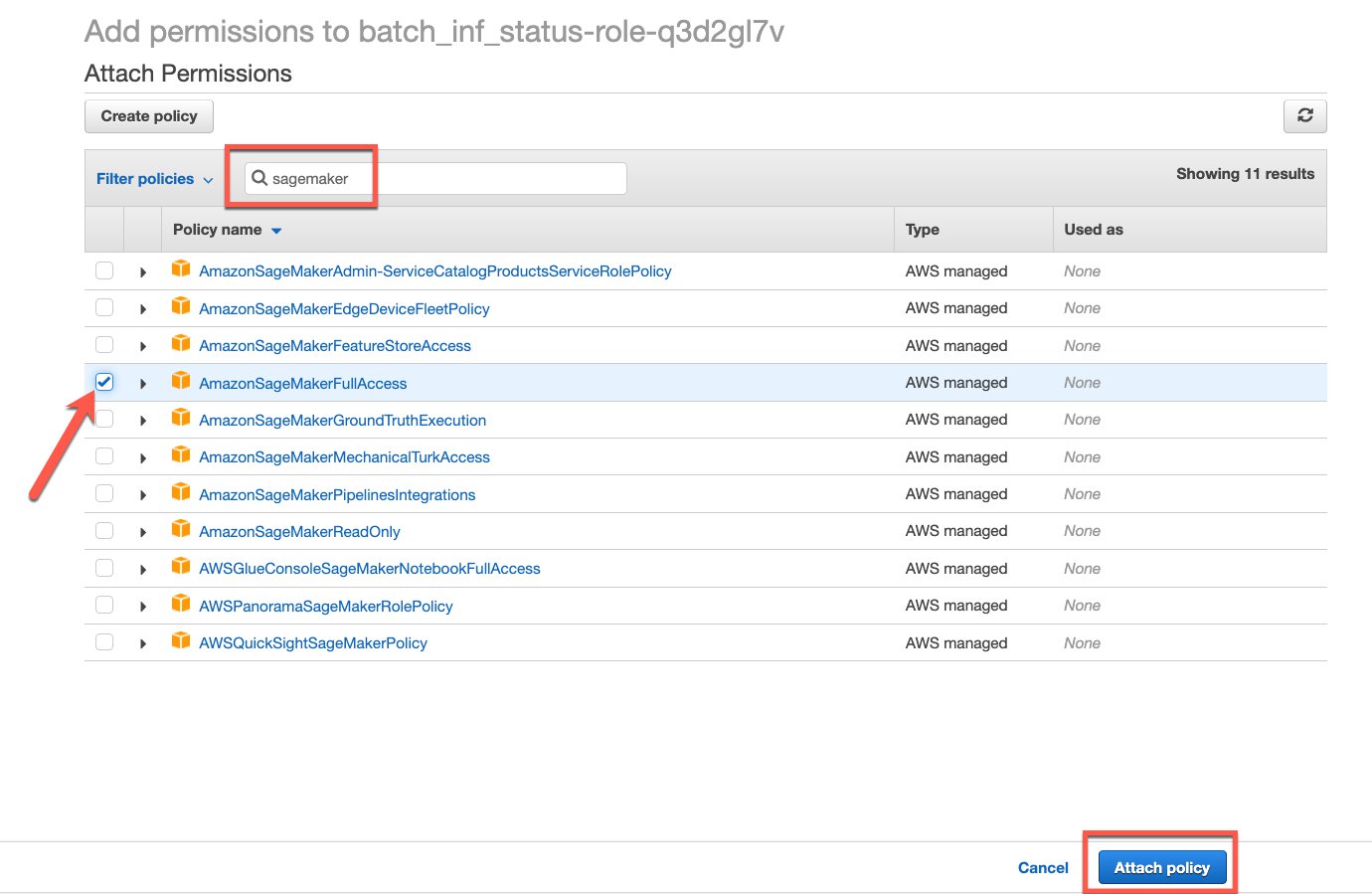












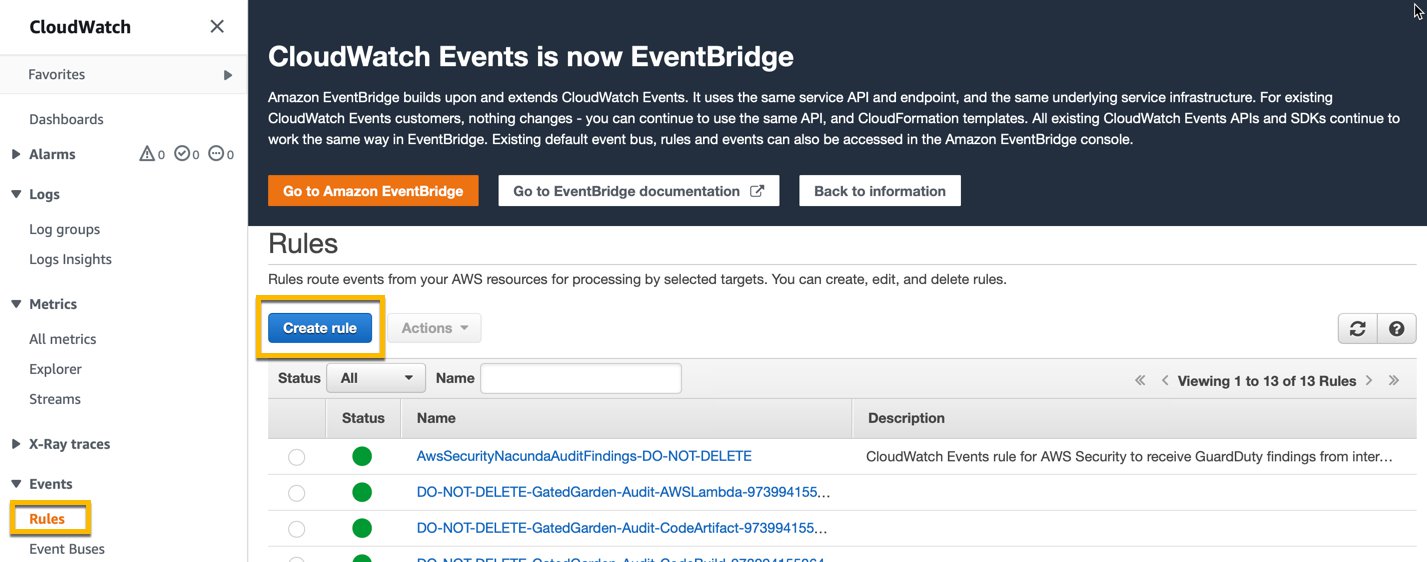
The Lambda function is now ready to react to a file drop in S, look at the file for headers, remove if there is any and call SageMaker Batch Inference with the file. We can do a few more things to the Lambda function to make it efficient.

In configurations, we can remove multiple retries in asynchronous section, we can also increase the memory to 1GB and change the timeout to 15min in General configuration sections.

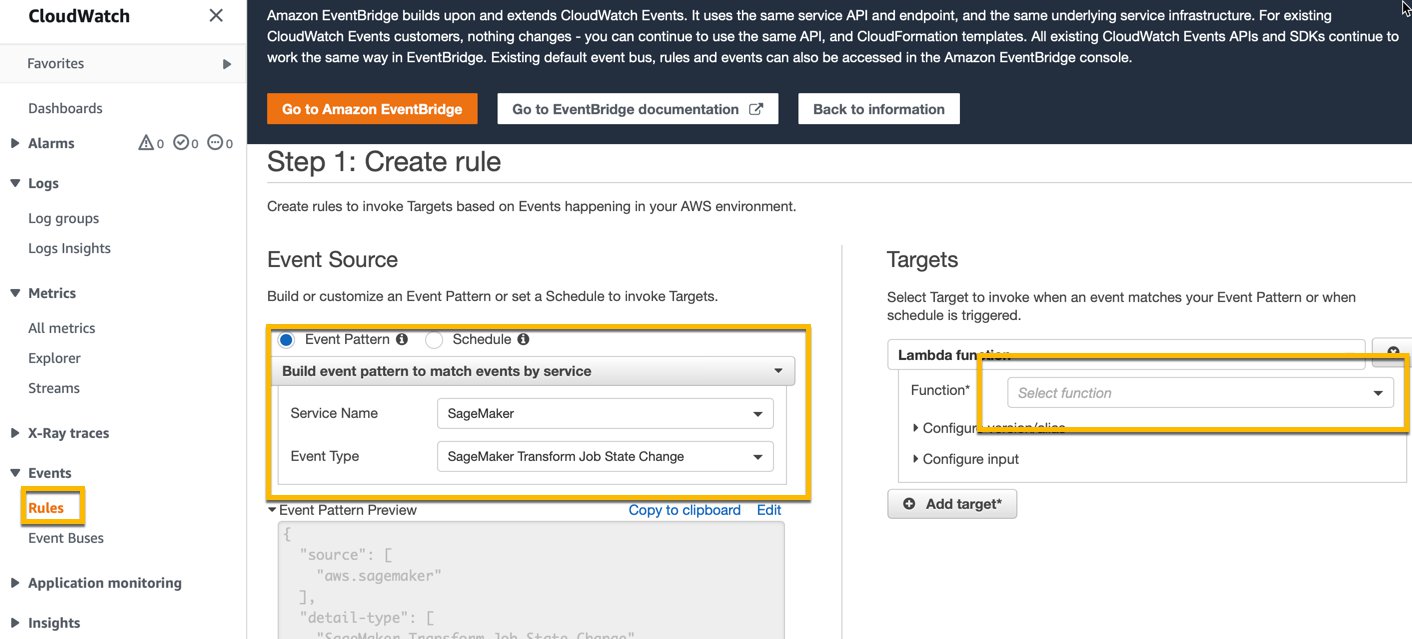
### **Lambda Function for Checking completion status and re-assemble the output csv**

This lambda function is created using the code from batch-ing-status.py code. The trigger for this is cloudwatch events.

1. Let’s start with lambda function as above and add an addition permission for SNS to this lambda function.
2. For the Trigger, Let’s create a cloudwatch event and associate with this lambda function.



1. Create a rule and select SageMaker as service and Transform Job State change as event type, on the right hand side select the second lambda function created above as your target and follow the rest of the prompts.



This lambda function will kick off after the first lambda function that initiated the inference. As a part of this, we have also provided code to create a SNS topic and subscribe to updates on Inference results. Whoever is on subscription list will get an email asking to confirm subscription and once confirmed they will get an email when the inference is successfully completed. In your case, your operations people will get an email saying new predictions are available.