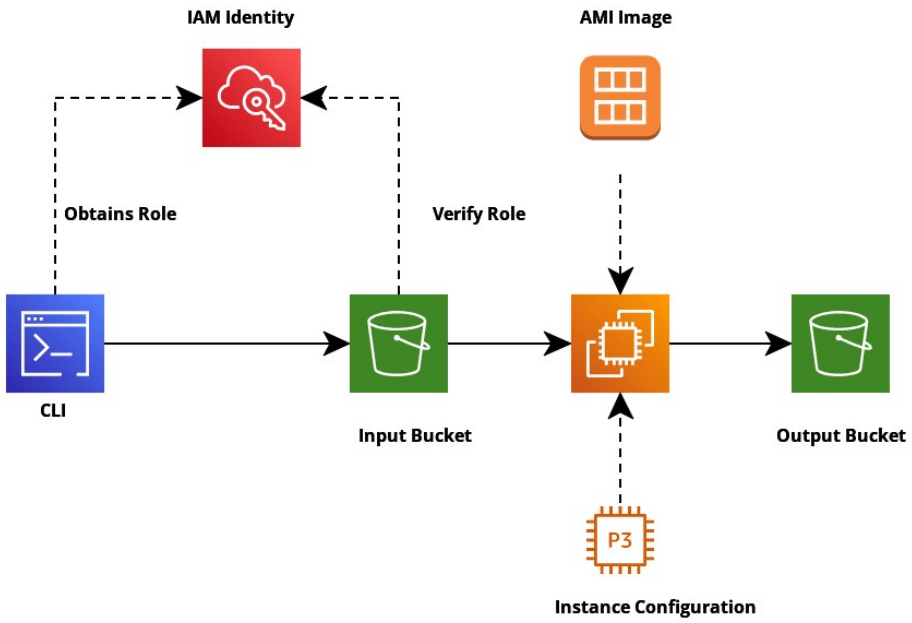
AWS Pipeline for Modeling

Pipeline



Summary

This pipeline enables one to quickly upload video and data to S3 buckets, which the EC2 instances grab data from to process data.

Part I - Setup

Credentials

First, acquire credentials. Ask your administrator. Details on what credentials to acquire can be found here Github AWS Docs

Ensure that your administrator has assigned you roles that enable you to connect to EC2 instances, alongside accessing S3 buckets.

CLI

Installation of CLI (Command Line Interface) can be found via this page

> Installing or updating the latest version of the AWS CLI - AWS Command Line Interface
> The AWS CLI is an open source tool built using the AWS SDK for Python (Boto) that provides commands for interacting with AWS services. With minimal configuration, you can start using all of the functionality provided by the AWS Management Console from your favorite terminal program.
> https://docs.aws.amazon.com/cli/latest/userguide/getting-started-install.html

CLI enables one to interface with AWS clients.

Part II - Operation

Upload files through CLI to S3

You can upload your files to S3 via the CLI and remove them faster than the console access. Instructions on how to do that can be found here -

> Using high-level (s3) commands with the AWS CLI - AWS Command Line Interface
> The AWS CLI is an open source tool built using the AWS SDK for Python (Boto) that provides commands for interacting with AWS services. With minimal configuration, you can start using all of the functionality provided by the AWS Management Console from your favorite terminal
> https://docs.aws.amazon.com/cli/latest/userguide/cli-services-s3-commands.html

OR

Utilize the AWS plugin for Visual Studio. Complete the setup there, and directly upload and download the files.

The bucket to utilize is `cis-input-bucket`

EC2 Provisioning

If the instance has been created and is assigned to you, go to the next step.

If the instance has not been created…

1. First, go to the EC2 dashboard. Then go click `launch instance`
2. Give the instance a name
3. Choose the image (OS) of the instance. Do note that there are specific instances for machine learning under ubuntu.

> 💡 You can also create your images through AMI - more on how to do that later.
>
> The benefit of AMI's is that one doesn't have to reconfigure an entire machine and can quickly load the configuration desired.

4. Select the keypair for login. This may either be one that your administrator has assigned to you, or you may need to create a new one.

> 💡 If generating a new keypair, remember to save it!

5. Network settings may be configured depending on your sensitivity use case.
6. Configure storage - based on how much storage your application needs, provision the proper amount. You'll see the acronym EBS

> 💡 EBS's are storage disks of an instance. You can disconnect an EBS from an instance and save it to be used later or for another instance's use.

Running EC2

Checkmark which instance you wanted to launch, and click the dropdown for `instance state` , and click `start instance` .

To stop an instance, checkmark the instance you want to stop, and click the drop down for `instance state` , and click `stop instance`

`reboot instance` - restart

`terminate instance` - delete instance

> 💡 Make sure to turn off your instance when its not in use.

For this project - ML 1 and ML 2 are instances for training

To Connect

Checkmark the instance you would like to connect to, and click `actions` followed by `connect`

Options to Connect

There are two methods to connect:

- through the SSH console - this is where the .pem file comes in handy, follow the instructions

- through the EC2 viewer, follow the instructions

Retrieving files inside EC2

Use the following to download the files you need from the S3 buckets

```
import boto3

s3 = boto3.client('s3')
s3.download_file('BUCKET_NAME', 'OBJECT_NAME', 'FILE_NAME')
```

Write to S3

Use this script:

https://github.com/CurveAssure/CISII/blob/main/AWS/Pipeline_Files/s3Upload.py