AWS Setup Guide

Link Tree:

* Presentations from Robert Wang
* Cross Account Guide

Notes:

* I’ve set our default region to us-west-2, but within IAM you’ll always default to Global.

# Set up Your Environment

## Join the Organization

This allows us to share credits and share services/permissions across AWS accounts.

Should be as simple as accepting the email invite. Let me know if it’s expired.

Once you do this, send me your AWS ID so that I can give you access to the S3 bucket by editing the policies.

## Setup Multi-Factor Authentication

This is a suggestion from Robert Wang to ensure security. This won’t affect any of the following setup, but it’s good practice.

* Go to the [IAM Dashboard](https://us-east-1.console.aws.amazon.com/iam/home?region=us-west-2#/home)
* The first recommendation in the **Security recommendations** section recommends MFA
* Click that, follow the instructions to link to an MFA app (I use Authenticator)

## Create User Groups

Now we’ll set up User Groups, which specify which services you can use within AWS & ensures that we don’t spend credits on random services by accident.

### Admin

This User Group gives you access to everything in case you need to make high-level changes.

* Go to [User Groups](https://us-east-1.console.aws.amazon.com/iam/home?region=us-west-2#/groups)
* Click “Create Group” and fill out details:
  + User group name: **Admin**
  + *Don’t add a user - we haven’t made one yet!*
  + Attach permissions policies: **AdministratorAccess**
* Click “Create Group” to confirm

### Dev

This User Group limits your access to the services we need, and should be used for development purposes.

* Go to [IAM User Groups](https://us-east-1.console.aws.amazon.com/iam/home?region=us-west-2#/groups)
* Click “Create Group” and fill out details:
  + User group name: **Dev**
  + *Don’t add a user - we haven’t made one yet!*
  + Attach permissions policies:
    - **AmazonS3FullAccess**
    - **AmazonEC2FullAccess**
    - **AmazonSagemakerFullAccess**
* Click “Create Group” to confirm

## Create Users

Now we’ll create users use the permissions set by the User Groups above.

### admin

* Go to [IAM Users](https://us-east-1.console.aws.amazon.com/iam/home?region=us-west-2#/users)
* Click “Create user” and fill out details:
  + Step 1 - Specify user details
    - User name: **admin**
    - Select “Provide user access to the AWS Management Console”
      * Specify that you’d like to create an IAM User (second option)
    - Console password
      * Leave “Autogenerated password” selected
      * Deselect “Users must create a new password at next sign-in”
  + Step 2 - Set permissions
    - Leave “Add user to group” selected
    - Select **Admin** User Group (created above)
  + Step 3 - Review and create
    - Click “Create user” to confirm
  + Step 4 - Retrieve password
    - In the bottom right corner, there’s a button to “download .csv”
    - Click this, and keep this file somewhere safe

### dev01

* Go to [IAM Users](https://us-east-1.console.aws.amazon.com/iam/home?region=us-west-2#/users)
* Click “Create user” and fill out details:
  + Step 1 - Specify user details
    - User name: **dev01**
    - Select “Provide user access to the AWS Management Console”
      * Specify that you’d like to create an IAM User (second option)
    - Console password
      * Leave “Autogenerated password” selected
      * Deselect “Users must create a new password at next sign-in”
  + Step 2 - Set permissions
    - Leave “Add user to group” selected
    - Select **Dev** User Group (created above)
  + Step 3 - Review and create
    - Click “Create user” to confirm
  + Step 4 - Retrieve password
    - In the bottom right corner, there’s a button to “download .csv”
    - Click this, and keep this file somewhere safe

# Set up Sagemaker

## Create a DataScientist Role

* Go to [Sagemaker role manager](https://us-west-2.console.aws.amazon.com/sagemaker/home?region=us-west-2#/role-manager-landing) and check that you’re in us-west-2
* Click “Create a role” and fill out details:
  + Step 1 - Enter role information
    - Role name suffix: **DataScientist**
    - Select a persona: **Custom role settings**
  + Step 2 - Configure ML Activities
    - Select the top 10 permissions, and deselect the rest
      * Your last permission should be **AmazonS3FullAccess**
    - S3Buckets: **lyricgen**
    - ECRRepoArns: **arn:aws:ecr:us-west-2:725212641896:repository/lyricgen**
  + Step 3 - Add additional policies & tags
    - Keep as-is
  + Step 4 - Review role:
    - Click Submit!

## 

## Create a Domain

* Go to [Sagemaker Domains](https://us-west-2.console.aws.amazon.com/sagemaker/home?region=us-west-2#/studio)
* Click “Create domain” and fill out details:
  + Select “Set up for Organizations”
  + Step 1 - Domain details
    - Domain Name: **sagemaker-lyricgen** (doesn’t matter)
  + Step 2 - Users and ML Activities
    - Keep “Login through IAM” selected
    - Add a user, and name whatever you like (I did **default**)
    - Select “Use an existing role” and select the role you made:
      * **Sagemaker-DataScientist**
  + Step 3 - Applications
    - Keep as-is
  + Step 4 - Network
    - Select “Public Internet Access” so you can access pip, for example
    - For VPC:
      * Choose **Default**
      * Select All 4 Subnets
      * *Don’t* Select a Security Group
  + Step 5 - Storage
    - Change “Maximum Space Size” to **300GB**
      * probably overkill, but better to do this than recreate a domain later
  + Step 6 - Review and create
    - Click “Submit”

## 

## 

## 

## 

## Add User to Domain

Once your domain gets created (takes a few minutes) you need to create a user

* Go to your [Domains](https://us-west-2.console.aws.amazon.com/sagemaker/home?region=us-west-2#/studio) and click on the domain you just created.
* Click “add user” in the top right and fill out the details:
  + Step 1 - General settings
    - Name: **your-name** (doesn’t matter)
    - Execution Role: **Sagemaker-DataScientist** (the role you just made)
  + Step 2 - Studio settings
    - Leave as-is
  + Step 3 - RStudio settings
    - Leave as-is
  + Step 4 - Canvas settings
    - Leave Canvas base permissions enabled
    - Canvas Storage Configuration:
      * Select Custom S3 and input **s3://lyricgen**
    - Turn off Canvas ready-to-use models
    - Turn off Amazon Kendra document query
    - Turn off Time Series Forecasting
    - Leave MLOps enabled
  + Submit!

## Launch Sagemaker Studio

Now you’re all set up and ready to use Sagemaker Studio!

* Go to your [Domains](https://us-west-2.console.aws.amazon.com/sagemaker/home?region=us-west-2#/studio)
* Click “Launch” then “Studio” next to the user you created
* Once you’re redirected, click on “JupyterLab”
* In the top right, click “Create JupyterLab Space”
  + Name it however you like. I’m planning to separate these spaces by function, since you can link Git repos to your space.
    - For example, I have one space for Lyric-to-plan and another for preprocessing
* Wait for your space to get created, then modify as needed:
  + If you know you’ll need more memory, change the **Storage** as you see fit
  + If you are training with a GPU, use **ml.g4dn.xlarge**
    - This is the same GPU used on Colab & will be most cost effective.
* Then click “Run Space” and launch your notebook instance!

## Accessing S3 from Sagemaker Studio

The sagemaker\_accessing\_s3.ipynb notebook within the aws\_setup folder shows how you can access files from S3 and write back to S3 in different formats.

* Download the notebook
* Then within JupyterLab, use the file system on the left to upload the notebook
* Then run, and confirm that you can access the S3 bucket.
  + I’ve tested this with a non-berkeley aws account to write this guide. If you run into any issues, let me know.