module_1.md

Module 1: Jupyter-based integrated solution

Resources

- Blog Post: De-identify medical images with the help of Amazon Comprehend Medical and Amazon Rekognition
- Source code: GitHub: aws-samples/amazon-comprehend-medical-image-deidentification

Process

Step 1: Launch the CloudFormation Stack

- This CloudFormation stack deploys the infrastructure we will need in Module 2, and creates an S3 bucket we will use in this module.
- Reference: Implementation Guide, Step 1
- Sign in to the AWS console
- Click this CloudFormation console link to launch the CloudFormation stack:
  - The template location is pre-filled in the form
    - If the link does not work, copy this link and paste it into the 'Amazon S3 URL' field of the **Stacks** → **Create Stack** page in the CloudFormation console.
    - (Interest only: the template file is ai-powered-health-data-masking.template in Module 2's source code).
  - **Specify stack details** page: Enter a stack name such as **image-workshop**
  - **Configure stack options** page: Accept defaults
  - **Review** page: Select Acknowledge box
- From the **Outputs** of the stack, make note of the **ImageBucketName**

Step 2: Upload the sample image files

- Locate the downloaded course resources on your local machine
- Open the S3 service console, and click on the bucket created in Step 1
- Upload into this bucket the sample image files from the `images` directory of the course resources

Step 3: Create a SageMaker instance

- Open the **SageMaker** service console
- Select **Notebook** → **Notebook instances** from the left-hand sidebar
- Click **Create notebook instance** from the Notebook Instances page
  - Enter a suitable name such as **workshop-sm**
  - Select **ml.t3.medium** instance type (2 vCPU, 4 GiB memory, $0.0582/hr)
  - In **Permissions and encryption** → **IAM role**, select **Create a new role**
  - In the **Create an IAM role** popup, select **Any S3 bucket**
  - Note the name of the newly-created role (it will start with **AmazonSageMaker-ExecutionRole-**), and click on it to open the role in a new tab, for Step 4.
  - Click *Create notebook instance* to launch the instance (2 minutes)

Step 4: Add permission to use AWS services

- Open the **IAM** service console
- Locate the IAM role created in Step 2, and click on its name
- It contains these policies:
  - **AmazonSageMakerFullAccess** to run SageMaker
  - **AmazonSageMaker-ExecutionPolicy** allows access to S3
- Click **Attach Policies** and search for and select these policies:
  - **AmazonRekognitionFullAccess**
  - **ComprehendMedicalFullAccess**
- Click **Attach policy** and confirm that you now have 4 policies attached to this role

Step 5: Install the Jupyter notebook on the SageMaker instance

- In the **SageMaker** → **Notebook Instances** tab, once the notebook instance is running, click on its name and click on **Open Jupyter** to open Jupyter in a new tab.
- In the Jupyter tab, upload the files in the `python` directory of the course resources:
  - `medical_image_de_id.ipynb` is the Jupyter notebook file for this module
  - `api_imagemask.py` and `api_textmask.py` will be used in Module 2
- Click on the notebook file to run the notebook in a new tab

Step 6: Edit and run the Jupyter notebook

- In the Jupyter notebook tab, in the second code cell, edit the values for the **bucket** and **object** (name) of the uploaded image file in your bucket from Step 2
- Run the notebook by selecting **Cell** → **Run All** in the top menu bar
- The first time you run this, some software will be installed
- Experiment by adjusting the **phi_detection_threshold** to values between 0.00 (more redaction) and 1.00 (less redaction)