Massey University - 158.333 Applied Machine Learning & Big Data Processing

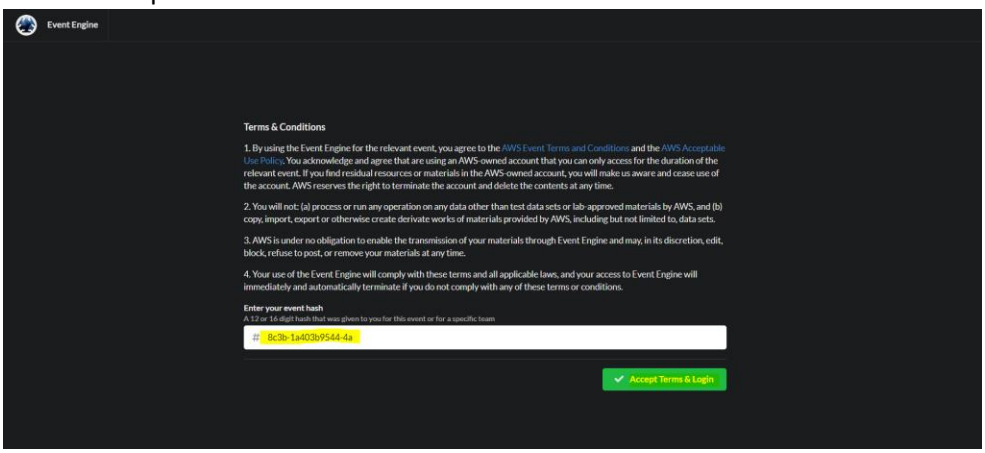
We've created some additional information to assist in navigating your way around the hands-on lab today. Our goal is to reduce complexity in the steps required and save time from menial tasks in setting up infrastructure tasks.

A couple of resources up front to have handy

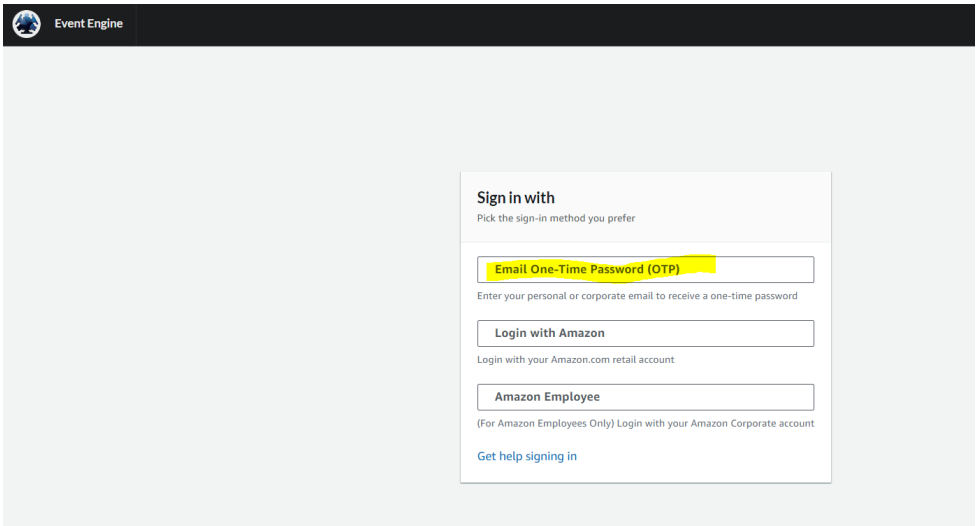
- Sample notebook repo - https://github.com/chindaws/EMRWorkshopsamples
- Link to workshop guide - https://catalog.us-east-1.prod.workshops.aws/workshops/3c29bc13-0f30-42f7-9f97-4ce8e2ef9b17/en-US

Logging into the Hands-on lab

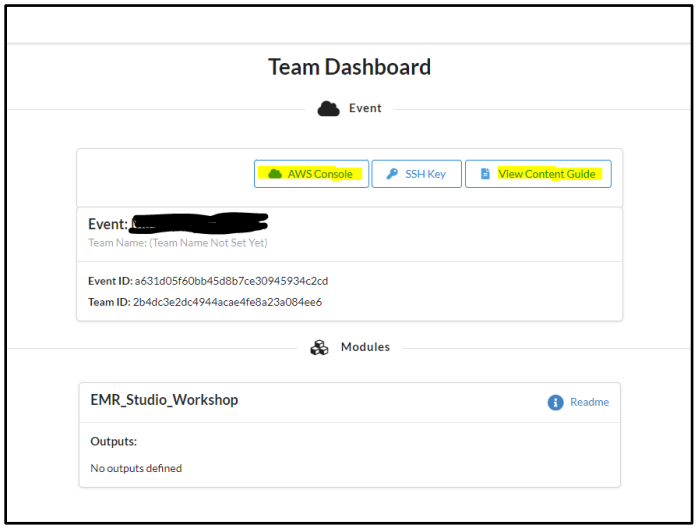
1. Event engine link https://dashboard.eventengine.run/login?hash=8c3b-1a403b9544-4a
2. Read & accept "Terms and Conditions"

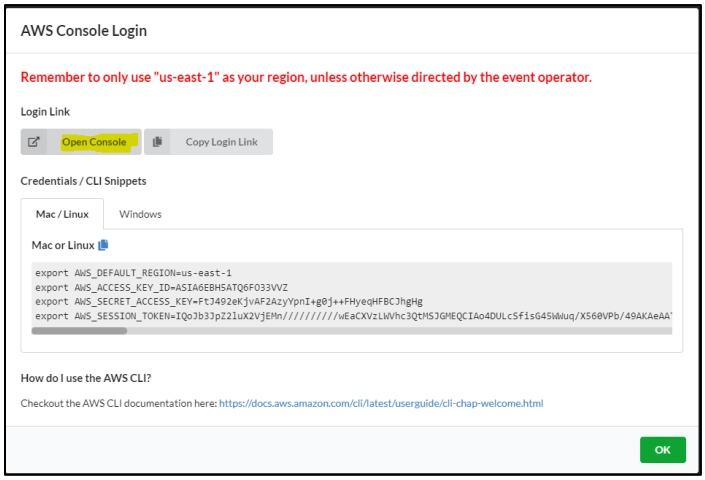


3. Select "Sign in with", Email One Time Password (OTP)
   Note: Use <youemail>@massey.ac.nz email address



4. Log into AWS Workshop dashboard using your One Time Password after entering code sent to your *@massey.ac.nz email.
   - Select “AWS Console”
   - Select “Open Console” login link





Developer Experience EMR lab guide

Access the guide:
https://catalog.us-east-1.prod.workshops.aws/workshops/3c29bc13-0f30-42f7-9f97-4ce8e2ef9b17/

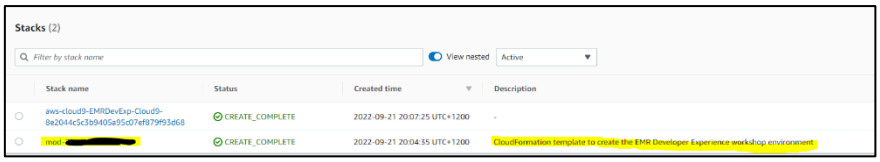
Notes and curated guidance for attendees

1. Introduction – Begin here to learn about the workshop.
   Then move to the “Setting up the EMR Studio"

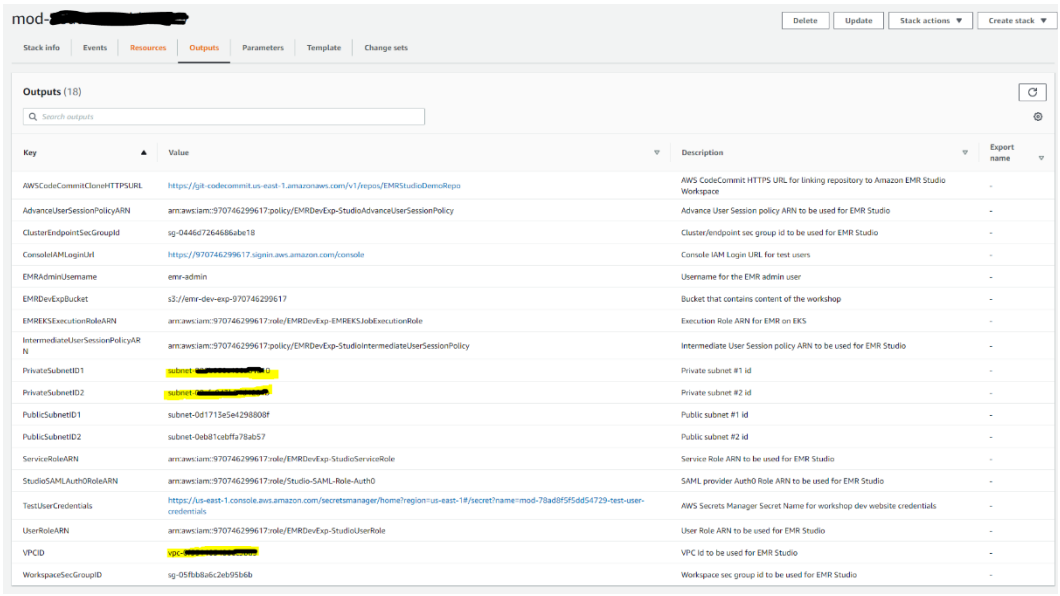
2. Setting up the EMR Studio – Using AWS Console
   - Skip bullet point (1) “Log out from the existing user.”
   - Bullet point (5) “Under the **Networking and security** section…”

Select the corresponding matching **VPC id** & **two private subnet ids** that you can reference from CloudFormation output tab below:

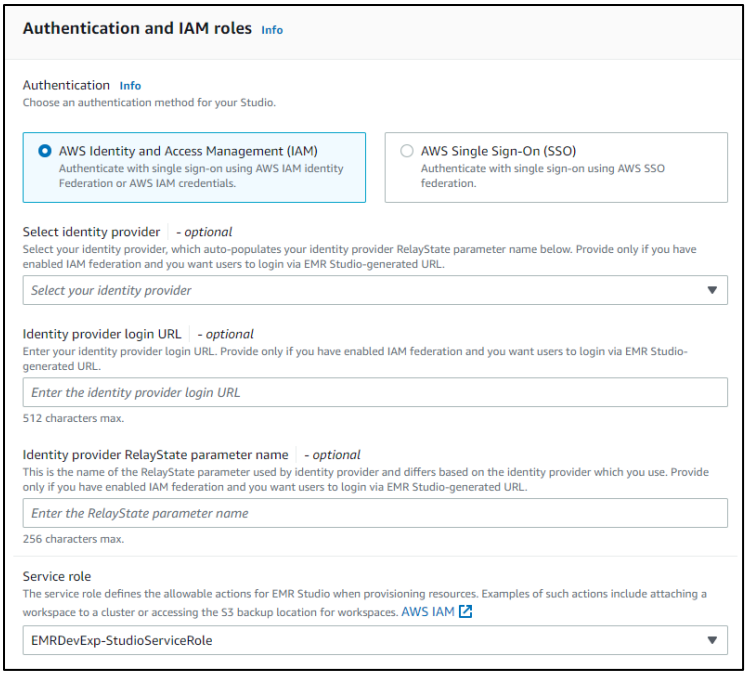
Access CloudFormation Stacks
https://us-east-1.console.aws.amazon.com/cloudformation/home?region=us-east-1#/stacks?filteringStatus=active&filteringText=&viewNested=true&hideStacks=false



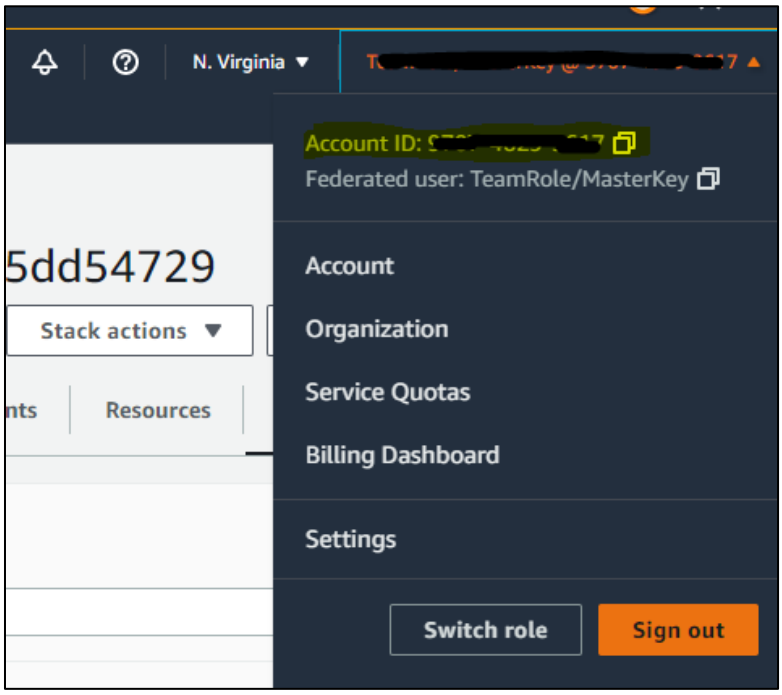
Click on the stack named: mod-############



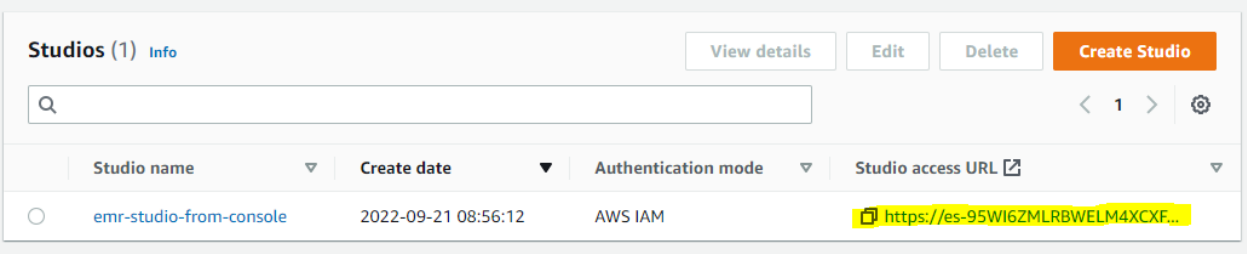
- At bullet point (6) “Now move on to **Authentication and IAM roles** section…” Disregard instructions here. Leave Authentication and IAM roles form as default to reflect the following image below.



- At bullet point (7) “And finally under the **Workspace storage…”**
  For s3://emr-dev-exp-<account-id>
  Locate <account-id> in the top right logged in account number



Once Studio has been created you can click onto the Studio access URL



Proceed to

3. Create EMR Studio Workspace

   Note: Workspace creation time should take <1min
   Once workspace has been created proceed to Getting familiarity with Workspace

4. Cluster Attachment
   - Carry out steps in Option 1 : Attach an existing Amazon EMR cluster
   - Proceed to Running Spark Jobs

5. Play around with additional Jupyter Notebook examples
   Download notebooks available at https://bit.ly/aws-massey

   machine-learning-with-pyspark-linear-regression.ipynb
   plot-graph-using-bokeh.ipynb
   visualize-data-with-pandas-matplotlib.ipynb

   Note: Refer to Readme for guidance on KERNEL RESTART when installing prerequisites on Amazon EMR cluster