

Massey University - 158.333 Applied Machine Learning & Big Data Processing

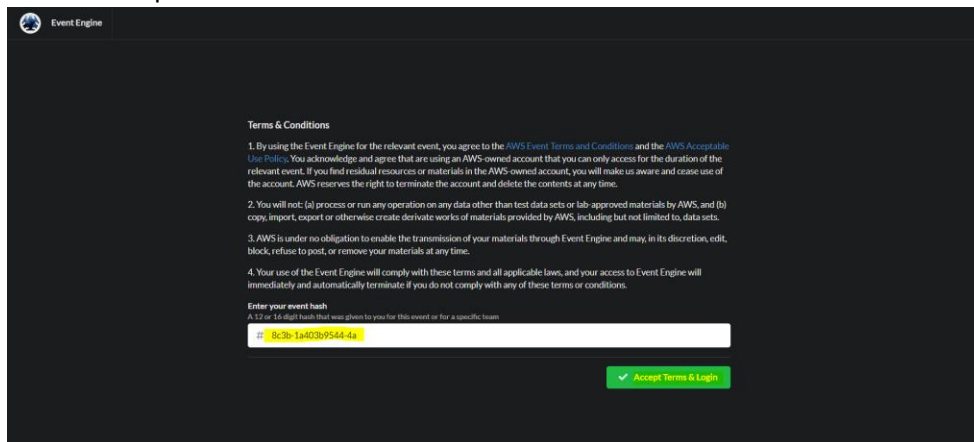
We've created some additional information to assist in navigating your way around the hands-on lab today. Our goal is to reduce complexity in the steps required and save time from menial tasks in setting up infrastructure tasks.

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>

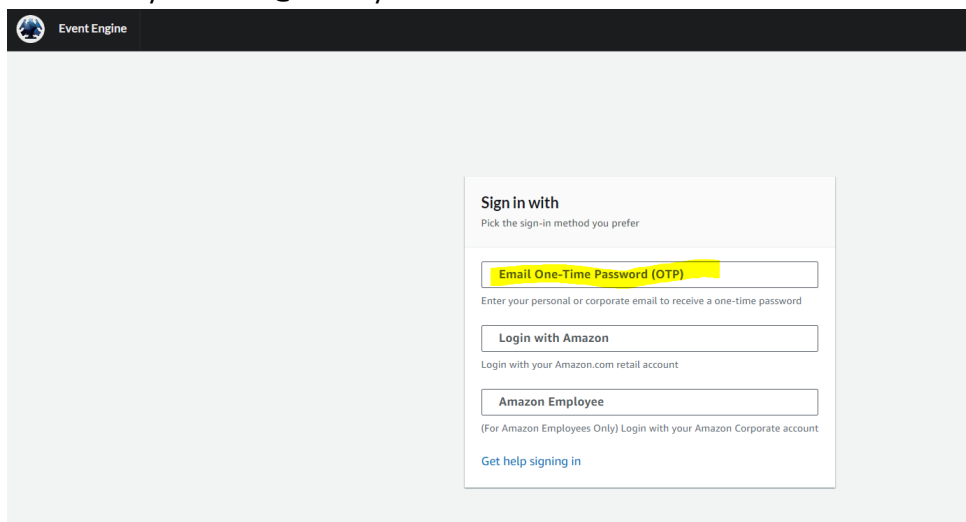
Logging into the Hands-on lab

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



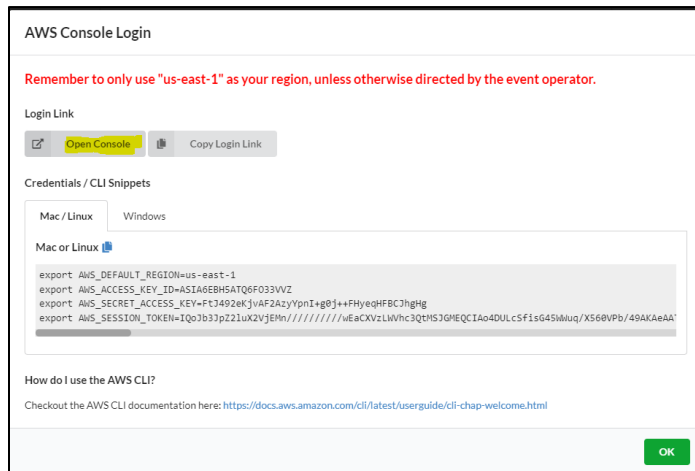
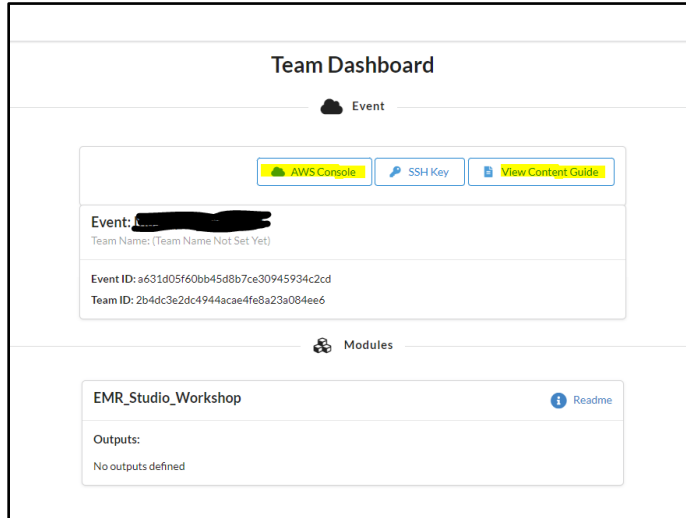
The screenshot shows the Event Engine login page with the "Terms & Conditions" section highlighted. The terms state that by using the Event Engine, the user agrees to the AWS Event Terms and Conditions and the AWS Acceptable Use Policy. It also outlines restrictions on data processing and AWS's right to terminate the account if terms are violated. A text input field contains the event hash "8c3b-1a403b9544-4a". A green button labeled "Accept Terms & Login" is at the bottom right.

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



The screenshot shows the Event Engine "Sign in with" page. It prompts the user to pick a sign-in method. The "Email One-Time Password (OTP)" option is highlighted. Below it, there is a text input field for the email address. Other options include "Login with Amazon" and "Amazon Employee". A link for "Get help signing in" is at the bottom.

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/>

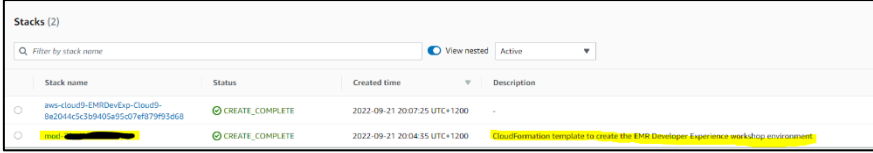
Notes and curated guidance for attendees

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:

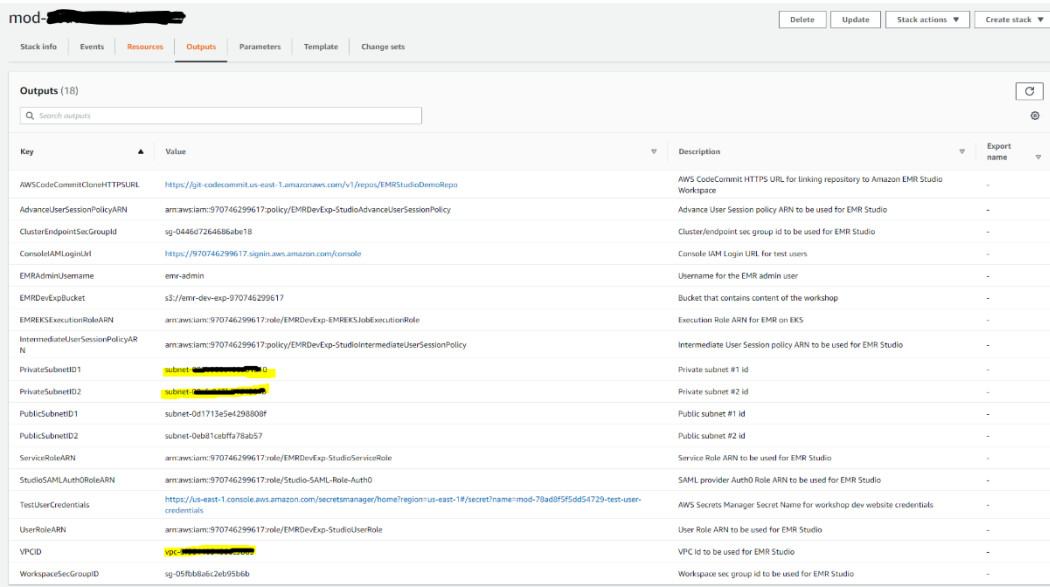
Access CloudFormation Stacks

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



Stack name	Status	Created time	Description
aws-cloud9-EMRDevExp-Cloud9-8e2044c5c3b9405a95d07ef879953d68	CREATE_COMPLETE	2022-09-21 20:07:25 UTC+1200	-
mod-#####	CREATE_COMPLETE	2022-09-21 20:04:35 UTC+1200	CloudFormation template to create the EMR Developer Experience workshop environment

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



Key	Value	Description	Export name
AWSCodeCommitCloneHTTPSURL	https://git-codecommit.us-east-1.amazonaws.com/v1/repos/EMRStudioDemoRepo	AWS CodeCommit HTTPS URL for linking repository to Amazon EMR Studio Workspace	-
AdvanceUserSessionPolicyARN	arn:aws:iam::970746299617:policy/EMRDevExp-StudioAdvanceUserSessionPolicy	Advance User Session policy ARN to be used for EMR Studio	-
ClusterEndpointSecGroupId	sg-0446d7264686abe18	Cluster/endpoint sec group id to be used for EMR Studio	-
ConsoleIAMLoginUrl	https://970746299617.signin.aws.amazon.com/console	Console IAM Login URL for test users	-
EMRAdminUsername	emr-admin	Username for the EMR admin user	-
EMRDevExpBucket	s3://emr-dev-exp-970746299617	Bucket that contains content of the workshop	-
EMREKSExecutionRoleARN	arn:aws:iam::970746299617:role/EMRDevExp-EMREKSJobExecutionRole	Execution Role ARN for EMR on EKS	-
IntermediateUserSessionPolicyARN	arn:aws:iam::970746299617:policy/EMRDevExp-StudioIntermediateUserSessionPolicy	Intermediate User Session policy ARN to be used for EMR Studio	-
PrivateSubnetID1	subnet-#####	Private subnet #1 id	-
PrivateSubnetID2	subnet-#####	Private subnet #2 id	-
PublicSubnetID1	subnet-0d1713e4e129800f	Public subnet #1 id	-
PublicSubnetID2	subnet-0eb11c0bfa78ab57	Public subnet #2 id	-
ServiceRoleARN	arn:aws:iam::970746299617:role/EMRDevExp-StudioServiceRole	Service Role ARN to be used for EMR Studio	-
StudioSAMLAuthRoleARN	arn:aws:iam::970746299617:role/Studio-SAML-Auth0	SAML provider Auth0 Role ARN to be used for EMR Studio	-
TestUserCredentials	https://us-east-1.console.aws.amazon.com/secretsmanager/home?region=us-east-1#/secret/name=mod-7ba6095f6d54729-test-user-credentials	AWS Secrets Manager Secret Name for workshop dev website credentials	-
UserRoleARN	arn:aws:iam::970746299617:role/EMRDevExp-StudioUserRole	User Role ARN to be used for EMR Studio	-
VPCID	vpc-#####	VPC id to be used for EMR Studio	-
WorkspaceSecGroupId	sg-05fbb8a6c2eb09366b	Workspace sec group id to be used for EMR Studio	-

- 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.

Authentication and IAM roles
Info

Authentication
Info

Choose an authentication method for your Studio.

☒ AWS Identity and Access Management (IAM)
Authenticate with single sign-on using AWS IAM identity Federation or AWS IAM credentials.

☐ AWS Single Sign-On (SSO)
Authenticate with single sign-on using AWS SSO federation.

Select identity provider
- optional

Select your identity provider, which auto-populates your identity provider RelayState parameter name below. Provide only if you have enabled IAM federation and you want users to login via EMR Studio-generated URL.

Select your identity provider

Identity provider login URL
- optional

Enter your identity provider login URL. Provide only if you have enabled IAM federation and you want users to login via EMR Studio-generated URL.

Enter the identity provider login URL

512 characters max.

Identity provider RelayState parameter name
- optional

This is the name of the RelayState parameter used by identity provider and differs based on the identity provider which you use. Provide only if you have enabled IAM federation and you want users to login via EMR Studio-generated URL.

Enter the RelayState parameter name

256 characters max.

Service role

The service role defines the allowable actions for EMR Studio when provisioning resources. Examples of such actions include attaching a workspace to a cluster or accessing the S3 backup location for workspaces. [AWS IAM](#)

EMRDevExp-StudioServiceRole

- 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

N. Virginia

Account ID: 5dd54729

Federated user: TeamRole/MasterKey

Account

Organization

Service Quotas

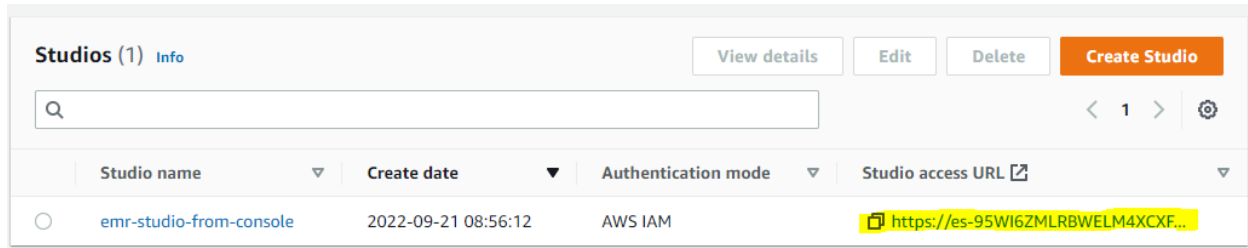
Billing Dashboard

Settings

Switch role

Sign out

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



The screenshot shows the AWS EMR Studio console. At the top, there's a header with 'Studios (1) Info', a search bar, and buttons for 'View details', 'Edit', 'Delete', and 'Create Studio'. Below the header is a table with the following columns: 'Studio name', 'Create date', 'Authentication mode', and 'Studio access URL'. There is one entry in the table: 'emr-studio-from-console', created on '2022-09-21 08:56:12', using 'AWS IAM' authentication, with an access URL 'https://es-95WI6ZMLRBWELM4XCXF...'. The URL is highlighted in yellow.

	Studio name	Create date	Authentication mode	Studio access URL
<input type="radio"/>	emr-studio-from-console	2022-09-21 08:56:12	AWS IAM	https://es-95WI6ZMLRBWELM4XCXF...

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