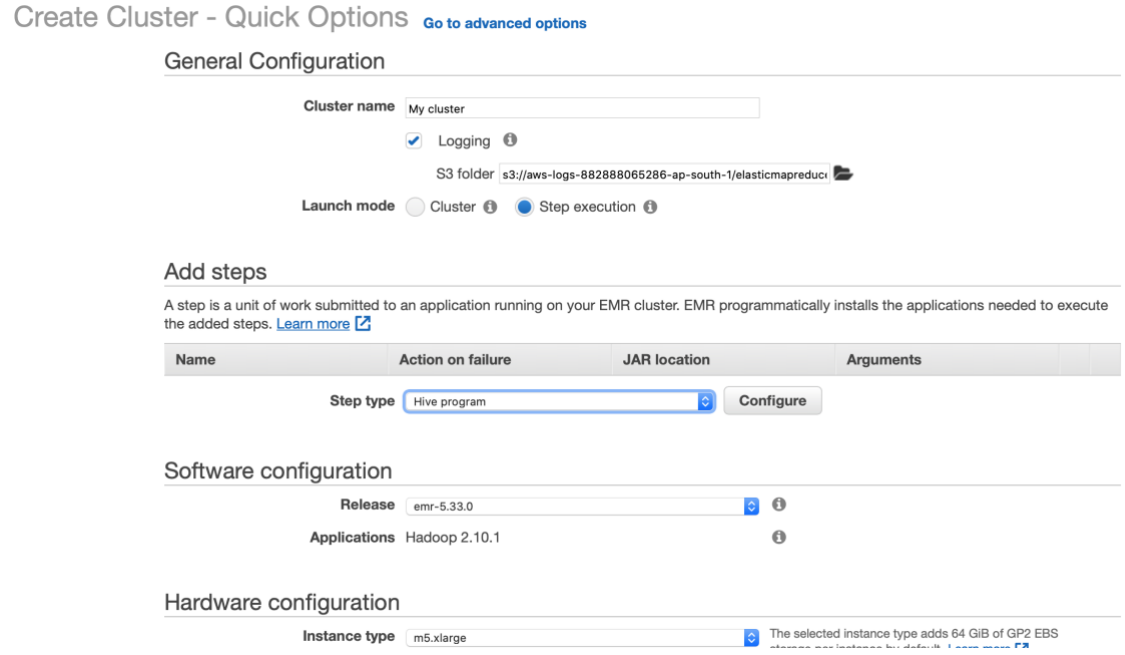
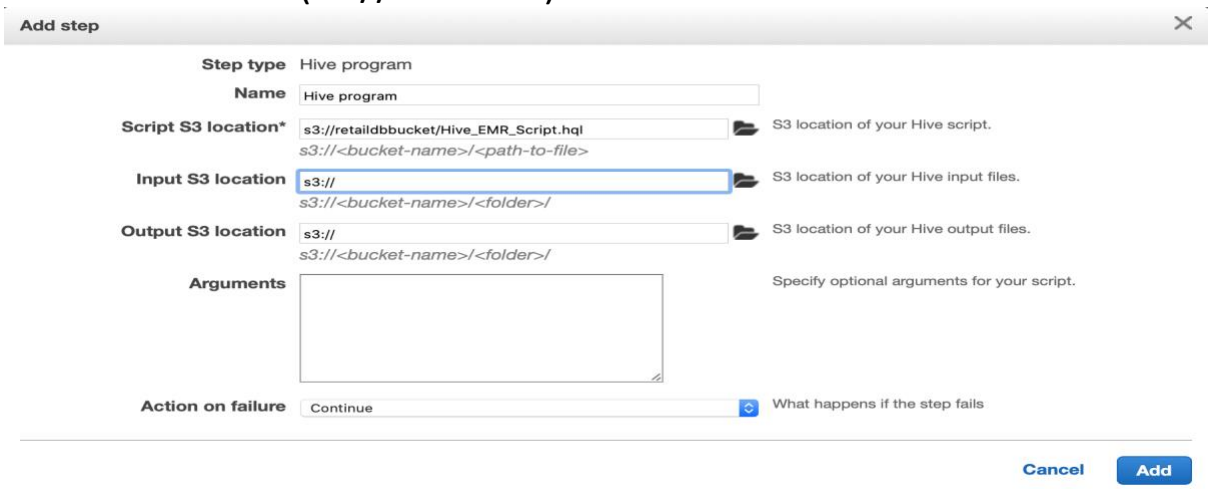
Lab – Data Processing using EMR – Step Execution – Hive Script

Step 1: Create a S3 bucket name “retail-db” and upload the files/directories from the link https://github.com/aws-training/retail_db.git or download from the following link “ https://github.com/aws-training/retail_db/archive/refs/heads/main.zip “ and upload all the contents.

Step 2: In AWS Console, select EMR and Click on create a cluster and give a name or use default. Select step execution and step type as Hive program. Select latest release and select m4.large machines(1).



Step 3: Click Configure beside the step type and select the script location from S3 (s3://retail-db)



Step 4: Click create cluster and wait till the status changes to **Waiting**
Step 5: Keep checking the status of your cluster and when it starts terminating go to S3 bucket and check if a directory is created with name “daily_product_revenue” .

Step 6: Click the file under the “daily_product_revenue” folder and query it through S3 select to check the contents