Amazon Web Services
Data Engineering Immersion Day

Lab 1. AutoComplete DMS

July 2021

Table of Contents

Introduction

Labs in the Data Engineering workshop are to be completed in sequence. This lab is designed to automate the Data Lake hydration with AWS Database Migration Service (AWS DMS), so we can fast forward to the next Glue lab.

If you prefer to get hands-on with AWS DMS service, please skip this lab and go to Lab 1 - Hydrating Data Lake (DMS).

In this lab, the automated process will complete the following tasks on your behalf:

- Set up lab environment, including S3 bucket and IAM roles
- Create a DMS subnet group within the VPC
- Create a DMS replication instance
- Create a source endpoint for RDS source database
- Create a target endpoint for full data load
- Create a target endpoint for CDC
- Create a task to perform the initial full data migration
- Create a task to support the ongoing replication of data changes (CDC)

If you'd like to run the workshop on your own after the AWS hosted event, please follow the lab instruction here: https://aws-dataengineering-day.workshop.aws/

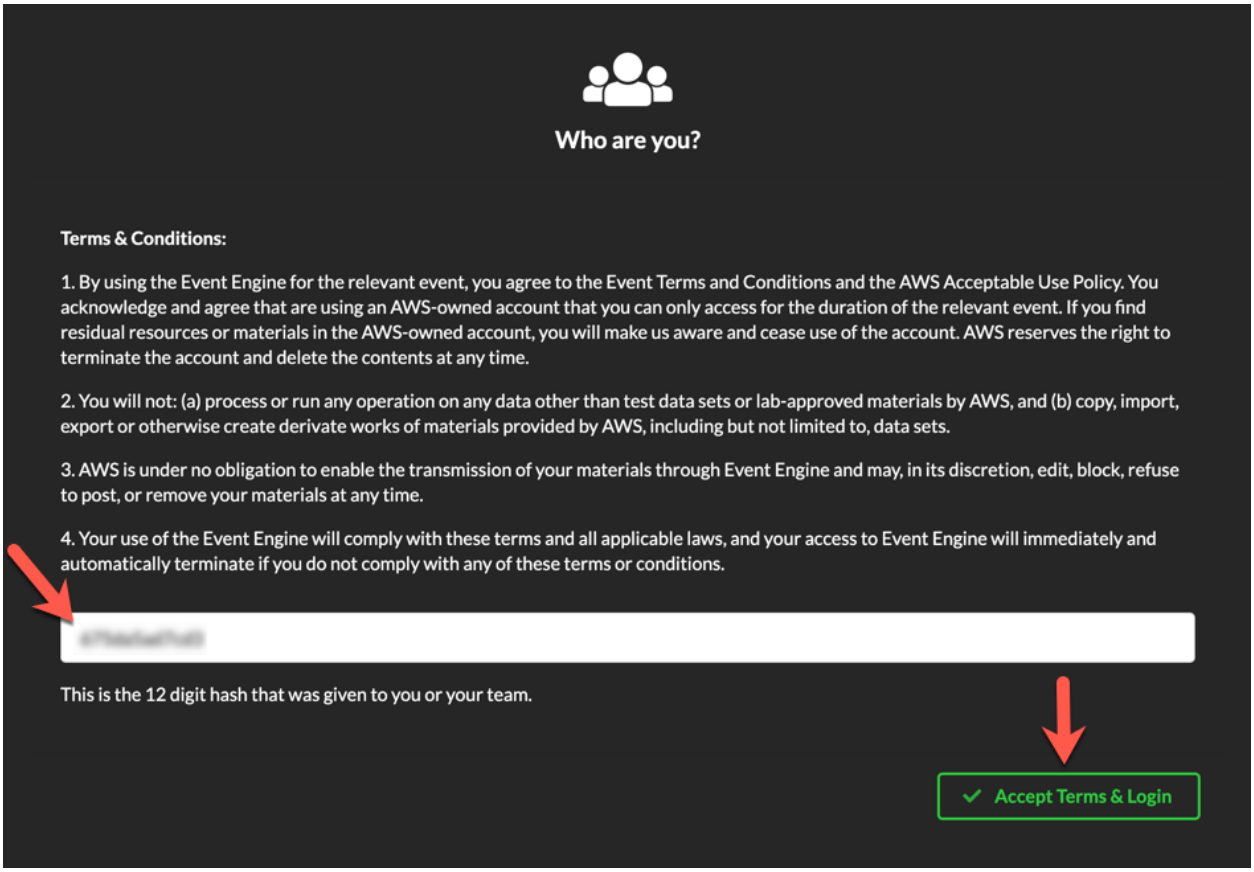
Get Started Using the Lab Environment

Please skip this section if you are running the lab on your own AWS account.

Today, you are attending a formal event and you will have been sent your access details beforehand. If in the future you might want to perform these labs in your own AWS environment by yourself, you can follow instructions on GitHub - https://github.com/aws-samples/data-engineering-for-aws-immersion-day.

A 12-character access code (or 'hash') is the access code that grants you permission to use a dedicated AWS account for the purposes of this workshop.

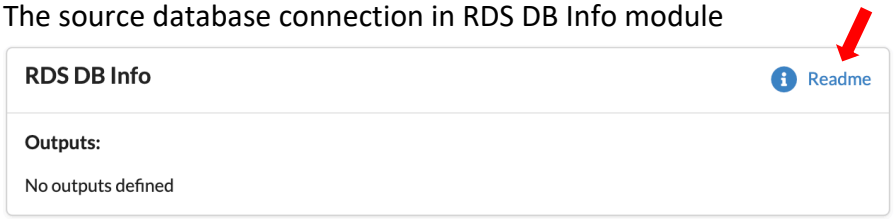
1. Go to https://dashboard.eventengine.run/, enter the access code and click Proceed:



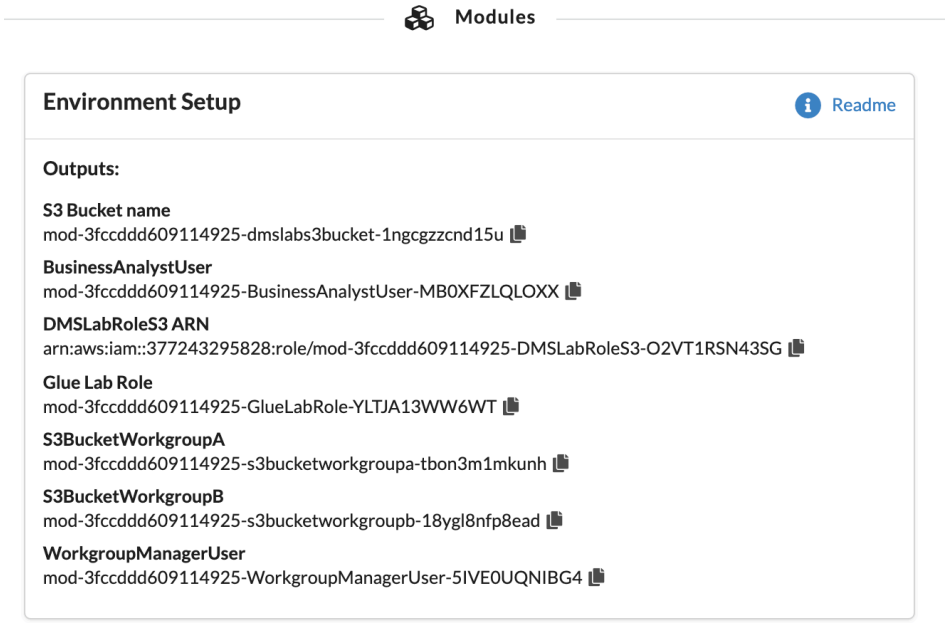
2. On the Team Dashboard web page you will see a set of parameters that you will need during the labs. Best to save them to a text file locally, alternatively you can always go to this page to review them. Replace the parameters with the corresponding values from here where indicated in subsequent labs:

Because you're at a formal event, some AWS resources have been pre-deployed for your convenience, for example:

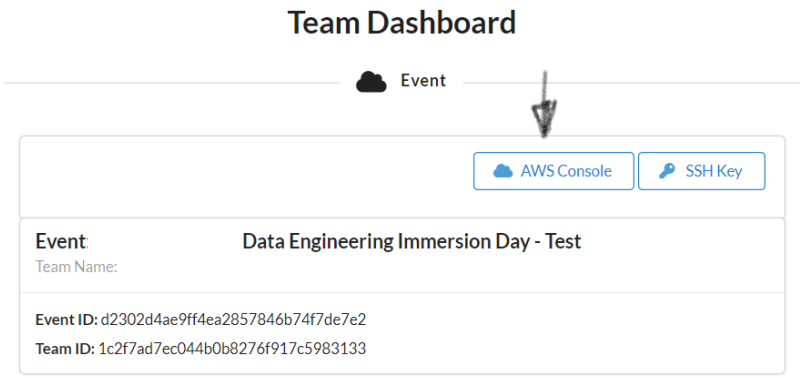
- The source database connection in RDS DB Info module



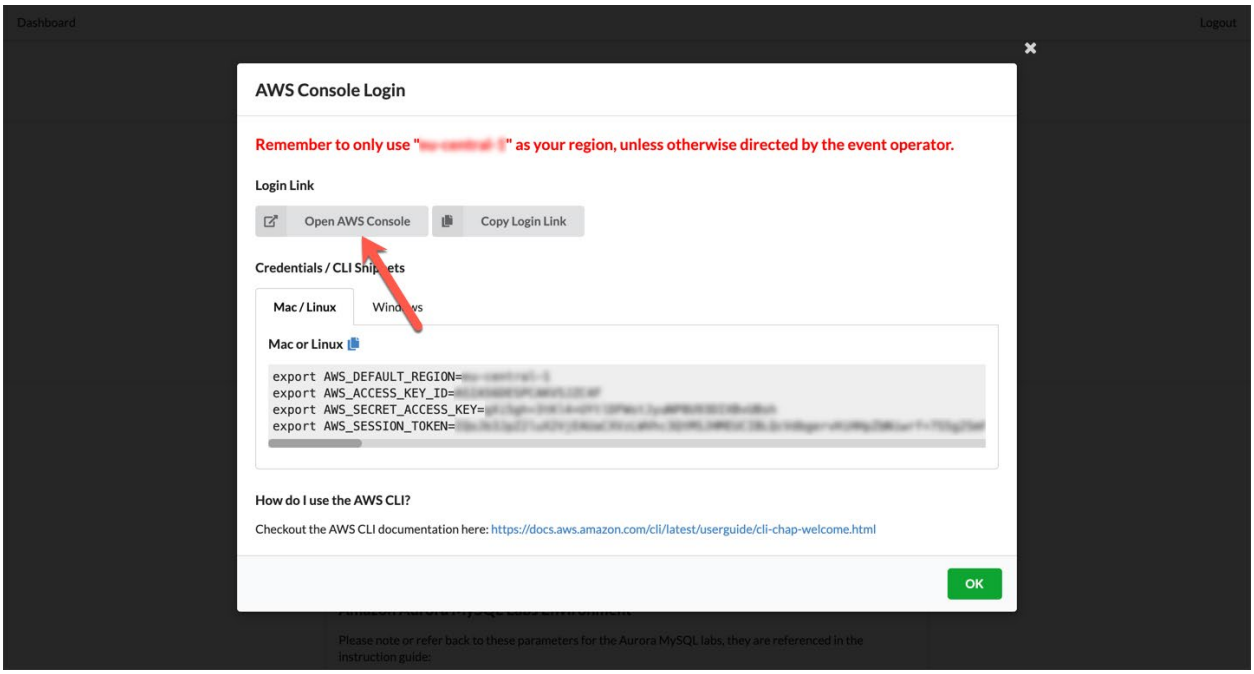
- S3 Bucket, IAM role for the DMS lab etc



3. On the Team Dashboard, please click AWS Console to log into the AWS Management Console:



4. Click Open AWS Console. For the purposes of this workshop, you will not need to use command line and API access credentials:

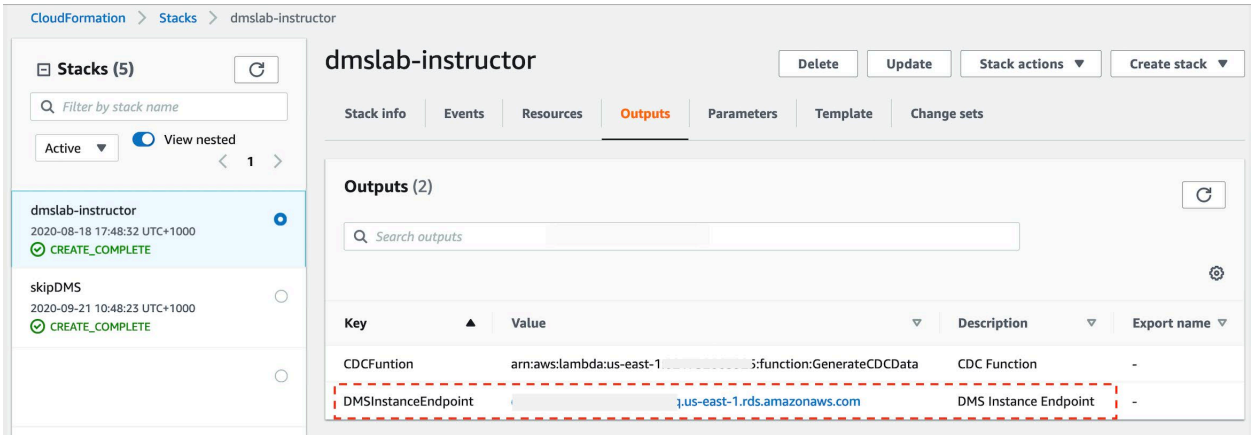


Once you have completed these steps, you can continue with the rest of this lab.

Pre-requisite

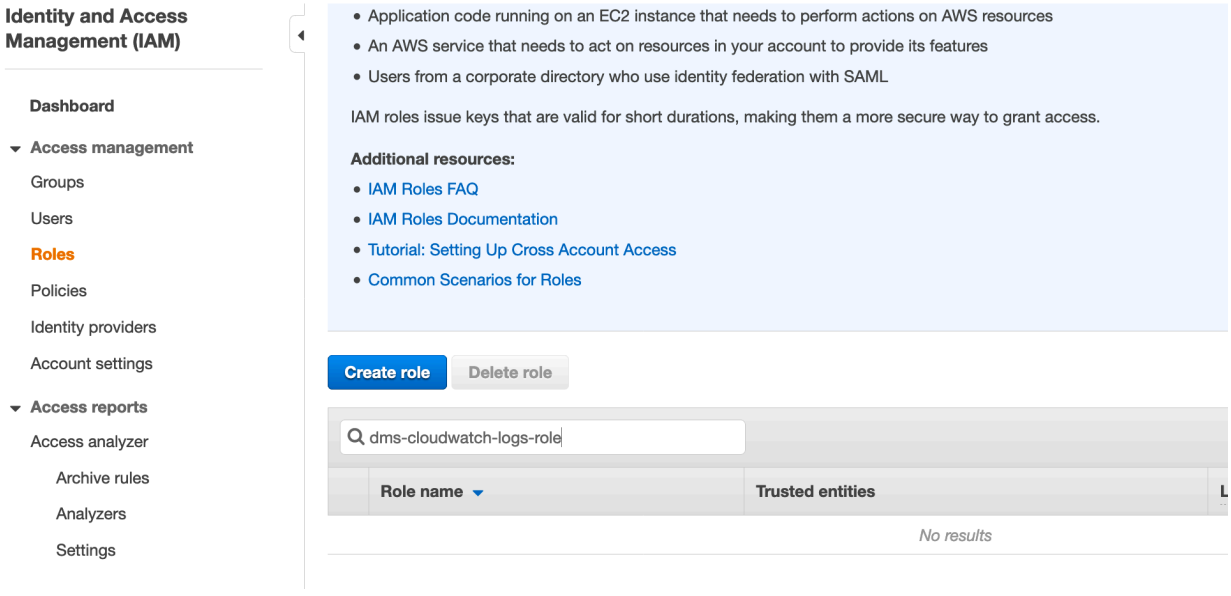
1. **RDS Database Server Name**: If you are on an AWS hosted event, you shall find this detail on Event Engine Team Dashboard under **Environment Setup** module.

Otherwise, check the *Outputs* tab on your CloudFormation Console, note down the RDS Server value.



2. **dms-cloudwatch-logs-role** & **dms-vpc-role**: Check if the Identity and Access Managment (IAM) roles exist in your lab account. Go to the IAM console, copy & paste the names in the search box.

   Note whether these roles are present or not. In this example screenshot the **dms-cloudwatch-logs-role** role is absent.

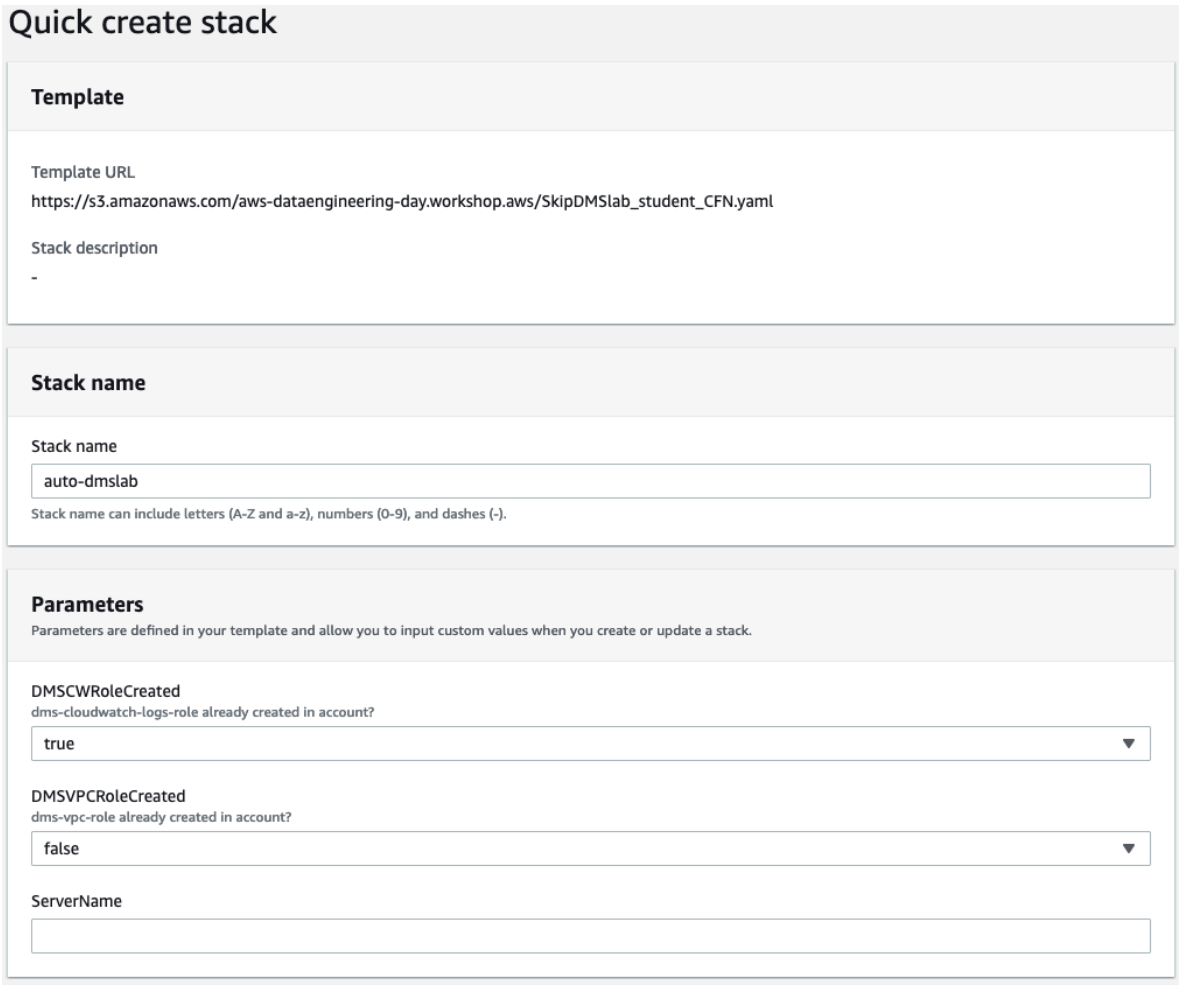


AutoComplete DMS

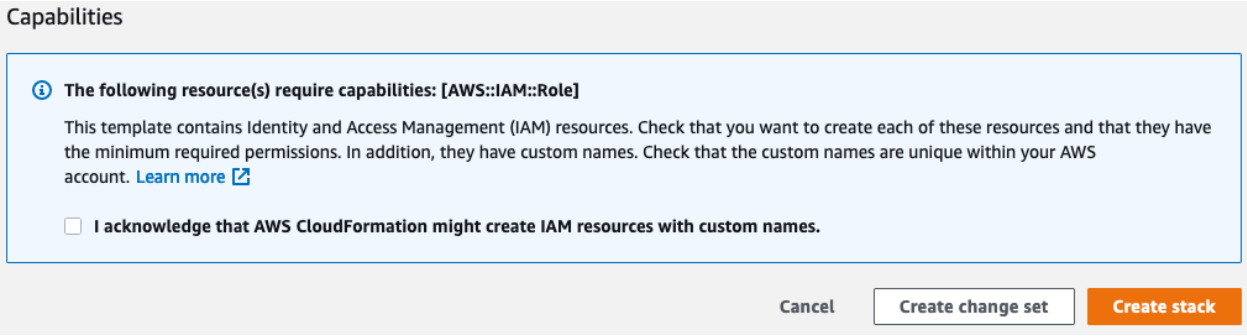
** Warning: You may find duplicate S3 buckets and IAM roles in your lab environment, once deploy the following template. If that happens, please use resources with name prefix "auto-dmslab-". NOT “dmslab-student-” **

1. Click the "Deploy to AWS" icon and open the link in a new web browser tab. It will load the CloudFormation dashboard to start the DMS automation process.

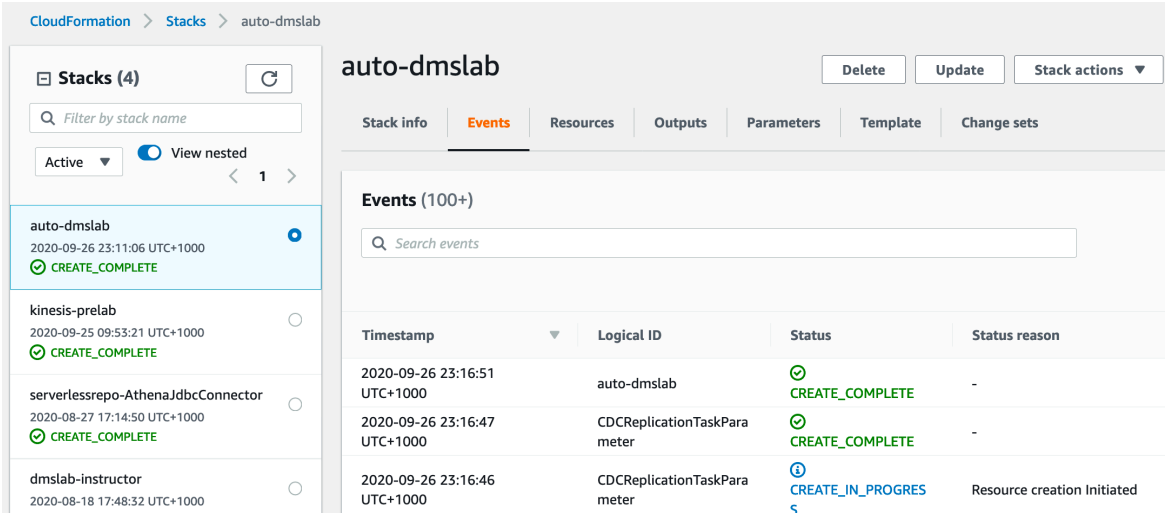
2. **Under Parameters:**
    - *DMSCWRoleCreated*: - If the role exists, keep to **yes**. If doesn’t exist, change to **no**.
    - *DMSVPCRoleCreated*: - **no** if the role doesn't exist. Otherwise, **yes**
    - *ServerName*: - Enter the RDS Database Server Name obtained from Environment Setup module in Event Engine Team Dashboard, e.g. ***dmslabinstance.xxxx<region>.rds.amazonaws.com***



3. Under **Capabilities** select the checkbox for "**I acknowledge that AWS CloudFormation might create IAM resources**" and select "**Create stack**".



4. The stack launch may take 5-6 minutes. Wait until your stack status advances to "CREATE_COMPLETE".



5. At this point, the source data has been fully loaded from RDS database to your S3 bucket via DMS. Go to AWS DMS console, you should see two **Database migration tasks** are 100% completed. If not, please wait until they are finished, then proceed to the Glue lab. **Going forward, ensure to use S3 bucket and IAM roles with name prefix "auto-dmslab-".**