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/

Pre-requisite

1. **RDS Database Server Name**: If you are on an AWS hosted event, your instructor can provide you the information, or you can find it on the event's web console. Otherwise, check the *Outputs* tab on your CloudFormation Console, note down the RDS Server value.
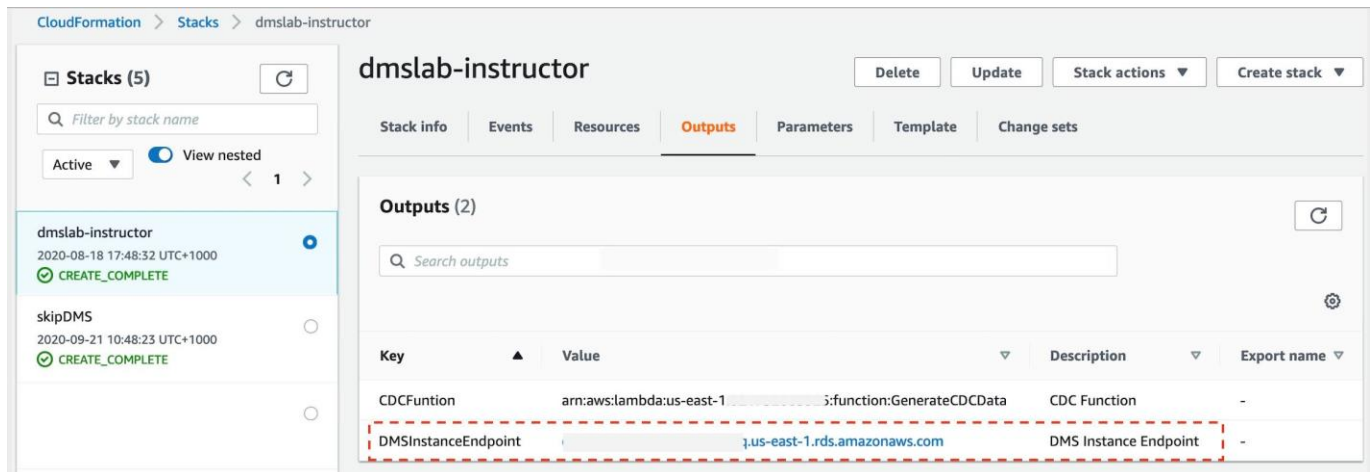
2. **dms-cloudwatch-logs-role** & **dms-vpc-role**: Check if the Identity and Access Managment (IAM) roles exist in your lab AWS account. Go to the IAM console, copy & paste the names in the search box.

   Note down whether these roles are present or not. In this example (shown as below), the **dms-cloudwatch-logs-role** role is absent.
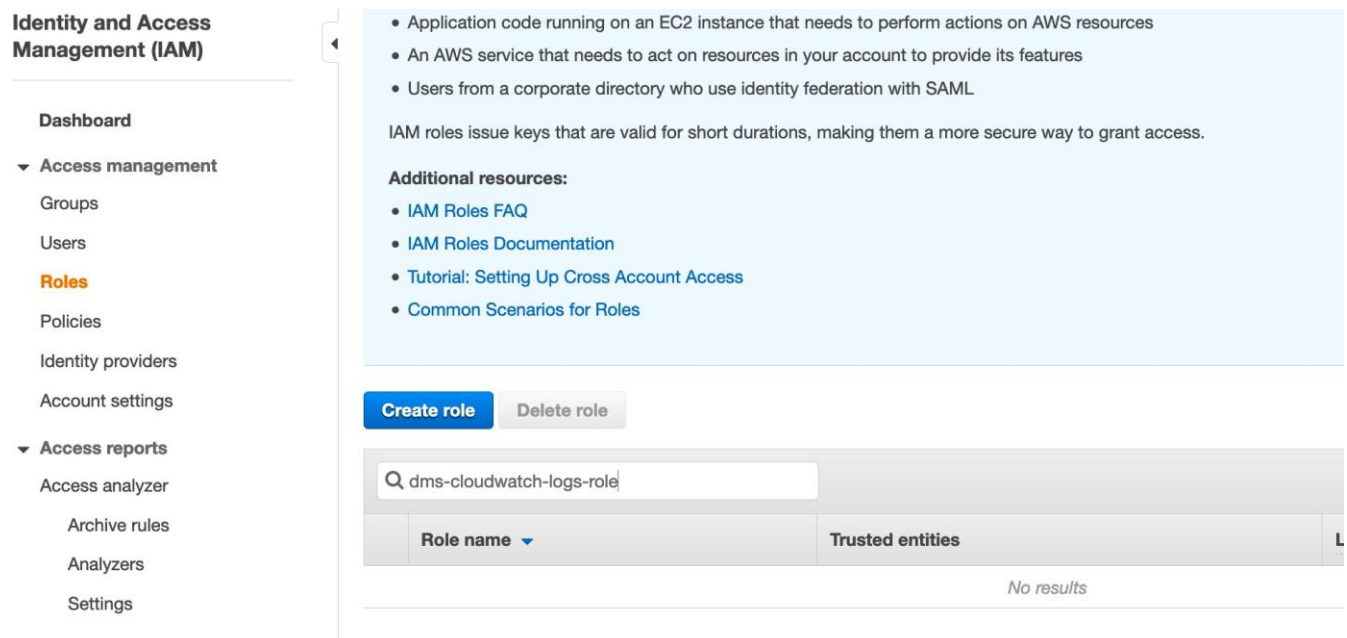
AutoComplete DMS

** 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. It will look similar to this: **dmslabinstance.xxxx<region>.rds.amazonaws.com**

CloudFormation > Stacks > Create stack

Quick create stack

Template

Template URL
https://s3.amazonaws.com/aws-dataengineering-day.workshop.aws/SkipDMSlab_student_CFN.yaml

Stack description
-

Stack name

Stack name
auto-dmslab

Stack name can include letters (A-Z and a-z), numbers (0-9), and dashes (-).

Parameters
Parameters are defined in your template and allow you to input custom values when you create or update a stack.

DMSCWRoleCreated
dms-cloudwatch-logs-role already created in account?

| true | ▼ |

DMSVPCRoleCreated
dms-vpc-role already created in account?

| false | ▼ |

ServerName

| dmslabinstance.~~xxxxxxxx~~.us-east-1.rds.amazonaws.com |

Capabilities

ⓘ **The following resource(s) require capabilities: [AWS::IAM::Role]**

This template contains Identity and Access Management (IAM) resources. Check that you want to create each of these resources and that they have the minimum required permissions. In addition, they have custom names. Check that the custom names are unique within your AWS account.Learn more ☑

☑ I acknowledge that AWS CloudFormation might create IAM resources with custom names.

Cancel | Create change set | **Create stack**

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.

6. 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 the name **prefix "auto-dmslab-".**