



Amazon Web Services Data Engineering Immersion Day

Lab 1. AutoComplete DMS
September 2020

Table of Contents

<i>Introduction</i>	<i>2</i>
<i>Get Started Using the Lab Environment.....</i>	<i>3</i>
<i>Pre-requisite.....</i>	<i>6</i>
<i>AutoComplete DMS</i>	<i>7</i>

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/>

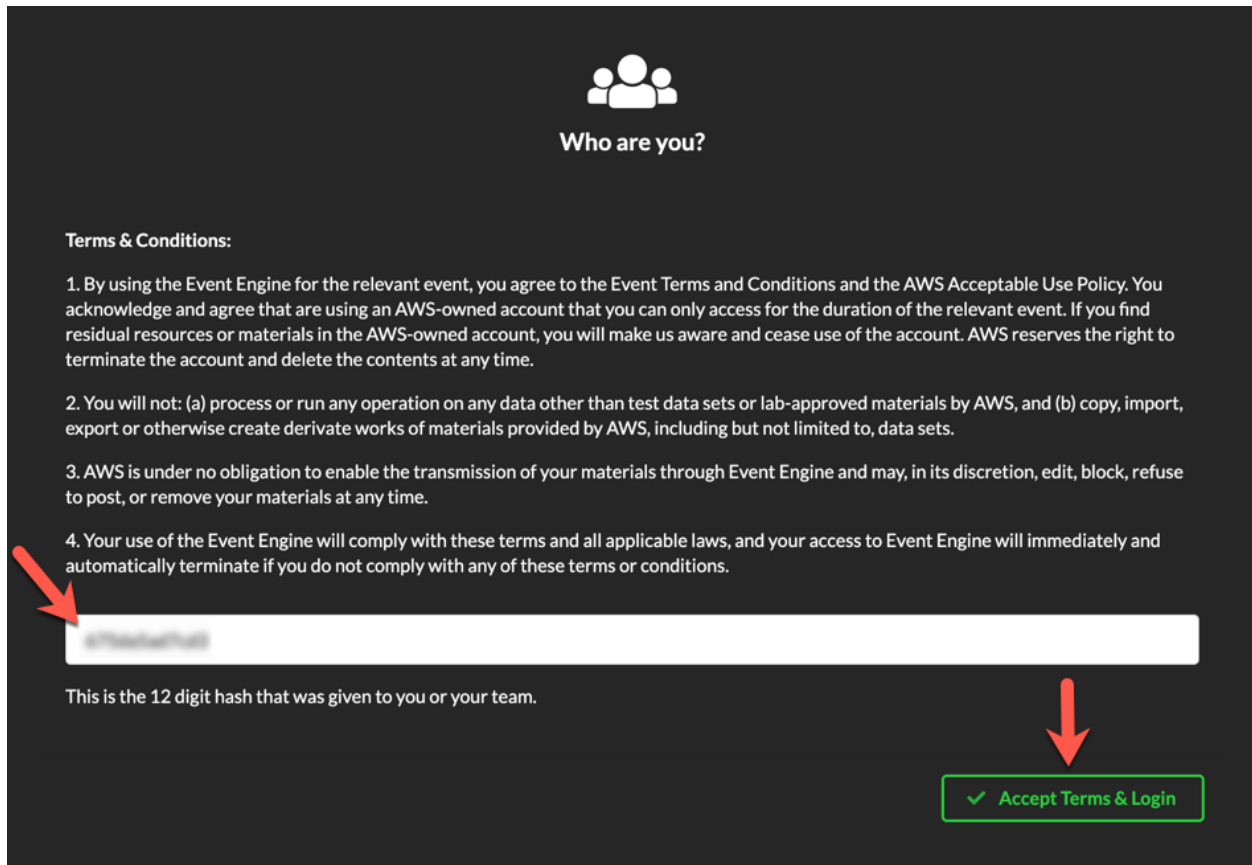
Get Started Using the Lab Environment

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:



Who are you?

Terms & Conditions:

1. By using the Event Engine for the relevant event, you agree to the Event Terms and Conditions and the AWS Acceptable Use Policy. You acknowledge and agree that are using an AWS-owned account that you can only access for the duration of the relevant event. If you find residual resources or materials in the AWS-owned account, you will make us aware and cease use of the account. AWS reserves the right to terminate the account and delete the contents at any time.
2. You will not: (a) process or run any operation on any data other than test data sets or lab-approved materials by AWS, and (b) copy, import, export or otherwise create derivate works of materials provided by AWS, including but not limited to, data sets.
3. AWS is under no obligation to enable the transmission of your materials through Event Engine and may, in its discretion, edit, block, refuse to post, or remove your materials at any time.
4. Your use of the Event Engine will comply with these terms and all applicable laws, and your access to Event Engine will immediately and automatically terminate if you do not comply with any of these terms or conditions.

This is the 12 digit hash that was given to you or your team.


✓ Accept Terms & Login

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:

Lab 1. AutoComplete DMS


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








RDS DB Info  [Readme](#)

Outputs:
No outputs defined

- S3 Bucket, IAM role for the DMS lab etc


 **Modules**



Environment Setup  [Readme](#)

Outputs:
S3 Bucket name
mod-3fccddd609114925-dmslabs3bucket-1ngcgzzcnd15u 
BusinessAnalystUser
mod-3fccddd609114925-BusinessAnalystUser-MBOXFZLQLOXX 
DMSLabRoleS3 ARN
arn:aws:iam::377243295828:role/mod-3fccddd609114925-DMSLabRoleS3-O2VT1RSN43SG 
Glue Lab Role
mod-3fccddd609114925-GlueLabRole-YLTJA13WW6WT 
S3BucketWorkgroupA
mod-3fccddd609114925-s3bucketworkgroupa-tbon3m1mkunh 
S3BucketWorkgroupB
mod-3fccddd609114925-s3bucketworkgroupb-18ygl8nfp8ead 
WorkgroupManagerUser
mod-3fccddd609114925-WorkgroupManagerUser-5IVE0UQNIBG4 

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

Team Dashboard

 **Event**

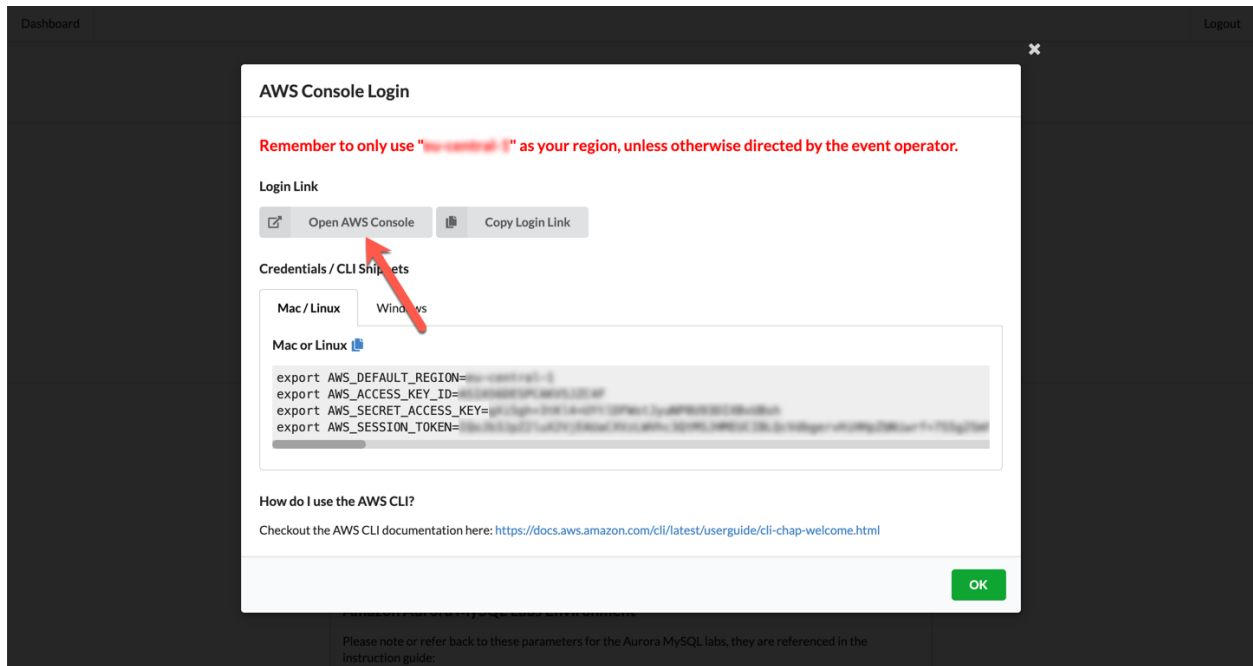
Event **Data Engineering Immersion Day - Test**

Team Name:

Event ID: d2302d4ae9ff4ea2857846b74f7de7e2
Team ID: 1c2f7ad7ec044b0b8276f917c5983133

Lab 1. AutoComplete DMS

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

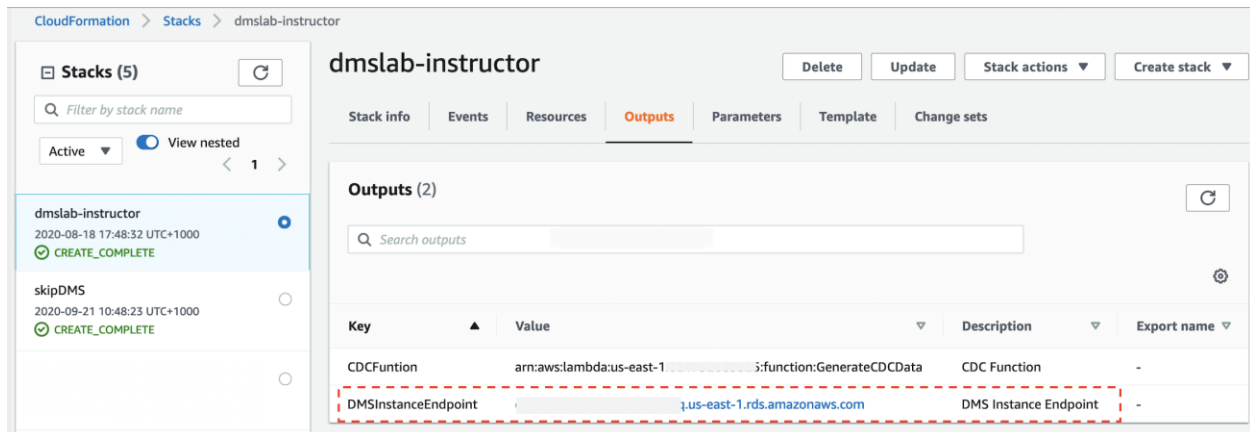


Once you have completed these steps, you can continue with the rest of this lab.

Pre-requisite

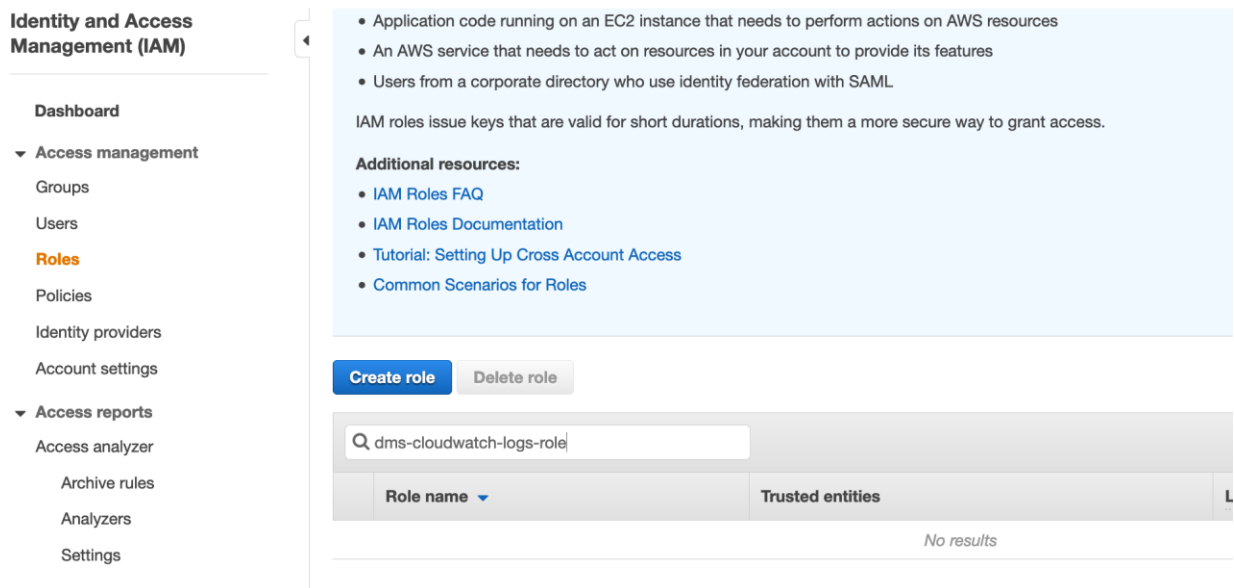
1. **RDS Database Server Name:** If you are on an AWS hosted event, your instructor should provide the detail.

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 Management (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.

Launch Template	Region
	N.Virginia (us-east-1)

2. Proceed through the wizard to launch the stack. At the Specify stack details page:

a) **Stack Name:** auto-dmslab

b) **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, eg.
dmslabinstance.xxxx<region>.rds.amazonaws.com

CloudFormation > Stacks > Create stack

Step 1
Specify template

Step 2
Specify stack details

Step 3
Configure stack options

Step 4
Review

Specify stack details

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?

no

DMSVPCRoleCreated
dms-vpc-role already created in account?

no

ServerName

dmslabinstance.crrprbscd9rq.us-east-1.rds.amazonaws.com

Cancel Previous **Next**

Lab 1. AutoComplete DMS

3. In Review page, review all the details, scroll down and check the box to acknowledge the policy and then click on Create Stack.

► Quick-create link

Capabilities

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

This template contains Identity and Access Management (IAM) resources that might provide entities access to make changes to your AWS account. Check that you want to create each of these resources and that they have the minimum required permissions. [Learn more](#)

☒ I acknowledge that AWS CloudFormation might create IAM resources.

Cancel Previous Create change set **Create stack**

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

CloudFormation > Stacks > auto-dmslab

Stacks (4)

Filter by stack name

Active View nested

- auto-dmslab
2020-09-26 23:11:06 UTC+1000
CREATE_COMPLETE
- kinesis-prelab
2020-09-25 09:53:21 UTC+1000
CREATE_COMPLETE
- serverlessrepo-AthenaJdbcConnector
2020-08-27 17:14:50 UTC+1000
CREATE_COMPLETE
- dmslab-instructor
2020-08-18 17:48:32 UTC+1000

auto-dmslab

Delete Update Stack actions

Stack info **Events** Resources Outputs Parameters Template Change sets

Events (100+)

Search events

Timestamp	Logical ID	Status	Status reason
2020-09-26 23:16:51 UTC+1000	auto-dmslab	CREATE_COMPLETE	-
2020-09-26 23:16:47 UTC+1000	CDCReplicationTaskParameter	CREATE_COMPLETE	-
2020-09-26 23:16:46 UTC+1000	CDCReplicationTaskParameter	CREATE_IN_PROGRESS	Resource creation Initiated

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-".

AWS DMS

Dashboard

Migration

Database migration tasks

Resource management

Replication instances

Endpoints

Certificates

DMS > Database migration tasks

Database migration tasks (2)

Find database migration tasks

Identifier	Status	Progress	Type	Source	Target
cdctask	Replication ongoing	100%	Ongoing replication	rds-source-endpoint	rds-cdc-endpoint
dms-task-full-dump	Load complete	100%	Full load	rds-source-endpoint	s3-target-endpoint