

# **Amazon Web Services Data Engineering Immersion Day**

Lab 1. AutoComplete DMS July 2021

#### **Table of Contents**

Introduction	2
Get Started Using the Lab Environment	3
Pre-requisite	6
AutoComplete DMS	7

## 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: <u>https://aws-dataengineering-day.workshop.aws/</u>

# 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 - <u>https://github.com/aws-samples/data-engineering-for-aws-immersion-day</u>.

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 <a href="https://dashboard.eventengine.run/">https://dashboard.eventengine.run/</a>, 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 an automatically terminate if you do not comply with any of these terms or conditions.									
	statutut								
	This is the 12 digit hash that was given to you or your team.								
	<ul> <li>Accept Terms &amp; Login</li> </ul>								

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



🔏 Modules

• S3 Bucket, IAM role for the DMS lab etc

Environment Setup	i Readme
Outputs:	
S3 Bucket name mod-3fccddd609114925-dmslabs3bucket-1ngcgzzcnd15u 🌗	
BusinessAnalystUser mod-3fccddd609114925-BusinessAnalystUser-MB0XFZLQLOXX	
DMSLabRoleS3 ARN arn:aws:iam::377243295828:role/mod-3fccddd609114925-DMSLabRoleS3-O2V	T1RSN43SG 🍺
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							
	AWS Console SSH Key							
Event: Team Name:	Data Engineering Immersion Day - Test							
Event ID: d2302d4ae Team ID: 1c2f7ad7ec	ff4ea2857846b74f7de7e2 44b0b8276f917c5983133							

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

Dashboard	
AWS Console Login	×
Remember to only use " as your region, unless otherwise directed by the event operator. Login Link C Open AWS Console C Copy Login Link Credentials / CLI Ship ets	
Mac/Linux Wind vs  Mac or Linux  export AWS_DEFAULT_REGION= export AWS_SECRET_ACCESS_KEY_ID= export AWS_SECRET_ACCESS_KEY= export AWS_SECRET_ACCESS_KEY= export AWS_SESSION_TOKEN=	
How do I use the AWS CLI? Checkout the AWS CLI documentation here: https://docs.aws.amazon.com/cli/latest/userguide/cli-chap-welcome.html	
Please note or refer back to these parameters for the Aurora MySQL labs, they are referenced in the instruction guide:	

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 <u>CloudFormation</u> Console, note down the RDS Server value.

CloudFormation > Stacks >	dmslab-instr	ıctor		
Stacks (5)	C	dmslab-instructor	Delete Updat	e Stack actions V Create stack V
Q Filter by stack name	sted	Stack info Events Resources Outp	Parameters Template Ch	nange sets
	< 1 >	Outputs (2)		C
dmslab-instructor 2020-08-18 17:48:32 UTC+1000 CREATE_COMPLETE	0	Q Search outputs		
skipDMS	0			٥
2020-09-21 10:48:23 UTC+1000 CREATE_COMPLETE		Key 🔺 Value	$\nabla$	Description $\nabla$ Export name $\nabla$
	0	CDCFuntion arn:aws:lambda:us-east	-1	CDC Function -
		DMSInstanceEndpoint	ı.us-east-1.rds.amazonaws.com	DMS Instance Endpoint -

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 <u>IAM console</u>, copy & paste the names in the search box.

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

Identity and Access	Application code running on an EC2 instance that needs to perform actions on AWS resources						
Management (IAM)	<ul> <li>An AWS service that needs to act on resources in y</li> </ul>	our account to provide its features					
Dashboard → Access management Groups Users Roles	<ul> <li>Users from a corporate directory who use identity from a corporate directory who use identity for IAM roles issue keys that are valid for short durations, Additional resources:</li> <li>IAM Roles FAQ</li> <li>IAM Roles Documentation</li> <li>Tutorial: Setting Up Cross Account Access</li> <li>Common Scenarios for Roles</li> </ul>	ederation with SAML , making them a more secure way to grant access.					
Policies							
Identity providers							
Account settings	Create role Delete role						
<ul> <li>Access reports</li> </ul>							
Access analyzer	Q dms-cloudwatch-logs-role						
Archive rules	Role name 👻	Trusted entities					
Analyzers	<ul> <li>An AWS service that needs to act on resources in your account to provide its features</li> <li>Users from a corporate directory who use identity federation with SAML</li> <li>IAM roles issue keys that are valid for short durations, making them a more secure way to grant access.</li> <li>Additional resources:         <ul> <li>IAM Roles FAQ</li> <li>IAM Roles Documentation</li> <li>Tutorial: Setting Up Cross Account Access</li> <li>Common Scenarios for Roles</li> </ul> </li> <li>Create role Delete role</li> <li>Q dms-cloudwatch-logs-role</li> <li>Role name          <ul> <li>Trusted entities</li> <li>No results</li> </ul> </li> </ul>						
Settings		No results					

### 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

Template URL https://s3.amazona Stack description	ws.com/aws-dataengineering-day.workshop.aws/SkipDMSlab_student_CFN.yaml
-	
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 DMSCWRoleCreated ms-cloudwatch-logs-	in your template and allow you to input custom values when you create or update a stack.
Parameters Parameters are defined DMSCWRoleCreated dms-cloudwatch-logs-1 true	in your template and allow you to input custom values when you create or update a stack. ole already created in account?
Parameters Parameters are defined DMSCWRoleCreated dms-cloudwatch-logs true DMSVPCRoleCreated dms-vpc-role already c	in your template and allow you to input custom values when you create or update a stack.  ole already created in account?
Parameters Parameters are defined DMSCWRoleCreated dms-cloudwatch-logs-t true DMSVPCRoleCreate dms-vpc-role already c false	in your template and allow you to input custom values when you create or update a stack.   ole already created in account?    t ated in account?

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

Capabilities	
The following resource(s) require capabilities: [AWS::IAM::Role]	
This template contains Identity and Access Management (IAM) resources. Check that you want to c the minimum required permissions. In addition, they have custom names. Check that the custom n account. Learn more 🔀	reate each of these resources and that they have ames are unique within your AWS
I acknowledge that AWS CloudFormation might create IAM resources with custom names.	
Cancel	Create change set Create stack

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

#### Lab 1. AutoComplete DMS

CloudFormation > Stacks > auto-dmsla	b			
🗆 Stacks (4)	auto-dmslab		Delete	odate Stack actions 🔻
<b>Q</b> Filter by stack name	Stack info Events R	esources Outputs Par	ameters Template	Change sets
Active View nested				
	Events (100+)			
auto-dmslab 2020-09-26 23:11:06 UTC+1000	<b>Q</b> Search events			
kinesis-prelab				
2020-09-25 09:53:21 UTC+1000 <b>⊘ CREATE_COMPLETE</b>	Timestamp	Logical ID	Status	Status reason
serverlessrepo-AthenaJdbcConnector	2020-09-26 23:16:51 UTC+1000	auto-dmslab	⊘ CREATE_COMPLETE	-
2020-08-27 17:14:50 UTC+1000	2020-09-26 23:16:47 UTC+1000	CDCReplicationTaskPara meter	⊘ CREATE_COMPLETE	-
dmslab-instructor 02020-08-18 17:48:32 UTC+1000	2020-09-26 23:16:46 UTC+1000	CDCReplicationTaskPara meter	CREATE_IN_PROGRES	Resource creation Initiated

At this point, the source data has been fully loaded from RDS database to your S3 bucket via DMS. Go to <u>AWS DMS console</u>, 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	X DMS	> Database migra	ation tasks									
Dashboard	Dat	tabase migrati	ion task	<b>s</b> (2)			C Acti	ons 🔻	Quick view	/ and comp	are	Create task
<ul> <li>Migration</li> </ul>	Q	Find database mi	gration tas	ks							<	1 > @
Database migration tasks												
Resource management		Identifier	$\nabla$	Status	$\bigtriangledown$	Progress 🛡	Туре	$\nabla$	Source	$\bigtriangledown$	Target	$\bigtriangledown$
Replication instances		cdctask		Replication ongo	oing	<b>—</b> 100%	Ongoing rep	olication	rds-source-e	endpoint	rds-cdc-e	ndpoint
Endpoints		dms-task-full-	dump	O Load complete		<b>—</b> 100%	Full load		rds-source-e	ndpoint	s3-target-	-endpoint
Certificates				C temptere					Jource e		target	