

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

Database Migration Services Lab September 2021

### **Table of Contents**

Introduction	2
Create the Subnet Group	3
Create the Replication Instance	4
Create the Source Endpoint	6
IAM Policy for DMS->S3 Access	8
Create the Target Endpoint	11
Create a task to perform the initial full copy	12
(Optional) Create the CDC endpoint to replicate ongoing changes	18
(Optional) Create a task to perform the ongoing replication	20

# Introduction

This lab will give you an understanding of the AWS Database Migration Service (AWS DMS). You will migrate data from an existing Amazon Relational Database Service (Amazon RDS) Postgres database to an Amazon Simple Storage Service (Amazon S3) bucket that you create.



In this lab you will complete the following tasks:

- 1. Create a subnet group within the DMS Lab VPC
- 2. Create a DMS replication instance
- 3. Create a source endpoint
- 4. Create a target endpoint
- 5. Create a task to perform the initial migration of the data.

Optionally, you can add ongoing replication of data changes on the source: (Only one of the DMS replication instances will enable this feature.)

- 6. Create target endpoint for CDC files to place these files in a separate location than the initial load files
- 7. Create a task to perform the ongoing replication of data changes

Your instructor has already created and populated the RDS Postgres database that you will use as your source endpoint in this lab.

# Create the Subnet Group

1. Search **DMS service** in the AWS console

aws	Services 🔺	Resource Groups 🗸	🙆 EKS 🛛 🚺
History		DMS	
Console Home	•	Database Migration S	Service
Database Migr	ation	Managed Database Migrati	ion Service
BDO		EC2	
RD5		Lightsail 🗗	
CloudFormatic	on	ECR	
CloudWatch		ECS	
EC2		EKS	
		Lambda	

2. On the DMS console, select **Subnet Groups**.

AWS DMS ×	DMS > Subnet group			
Dashboard	Subnet groups		C Actions V	Create subnet group
<ul> <li>Conversion &amp; migration</li> <li>Database migration tasks</li> </ul>	Q Find subnet group			< 1 > @
<ul> <li>Resource management</li> </ul>	Name	▼ Status	VPC ID	
Replication instances		Empty subnet gro	oup table	
Endpoints		You don't have any su	ubnet groups.	
Certificates				
Subnet groups				
Events				
Event subscriptions				

- 3. Click Create subnet group.
- 4. In the Identifier box, type a descriptive name that you will easily recognize (e.g., dms-lab-subnet-grp).
- 5. In the Description box, type an easily recognizable description (e.g., Replication instance for production data system).
- 6. For VPC, select the name of the VPC that you created earlier (e.g., dmslstudv1). The subnet list populates in the Available Subnets pane.
- 7. Select as many subnets as you want and click Add. The selected subnets move to the Subnet Group pane. Note: DMS requires at least two separate availability zones to be selected.

AWS DMS $\times$	DMS $>$ Create subnet group
Dashboard	Create replication subnet group
<ul> <li>Conversion &amp; migration</li> <li>Database migration tasks</li> </ul>	Subnet group configuration
<ul> <li>Resource management</li> <li>Replication instances</li> <li>Endpoints</li> <li>Certificates</li> <li>Subnet groups</li> <li>Events</li> <li>Event subscriptions</li> </ul>	Name A regionally scoped unique identifier you will use to identify your Replication Subnet Group dms-lab-subnet-grp Description Free form text to describe your Replication Subnet Group Replication instance for production data system VPC vpc-0314e829ba12d9481  V
	Add subnets Add Subnet(s) to this Subnet Group. You may add subnets one at a time or add all the subnets related to this VPC. You may make additions/edits after this group is created.  subnet-006522a5bbf92c0c5 - dmslstudv1 × us-east-1a 10.0.0.0/26 Private subnet-097495655d22662d2 - dmslstudv1 × us-east-1b 10.0.0.64/26 Private
	subnet-05f628a564cf25622 - dmslstudv1 X us-east-1c 10.0.0.128/26 Private Cancel Create subnet group

- 8. Click Create subnet group
- 9. On the DMS console, the subnet group status displays Complete.

DMS > Subnet group			
Subnet groups (1)			C Actions V Create subnet group
Q Find subnet group			< 1 > ©
Name	▼ Status	VPC ID	v
dms-lab-subnet-grp	⊘ Complete	vpc-0314e829ba12d9481	

## Create the Replication Instance

- 1. On the DMS console, select **Replication instances**.
- 2. Click Create replication instance.

AWS DMS $\times$	DMS > Replication instance
Dashboard	Replication instances C Actions V Create replication instance
Conversion & migration	Q Find replication instance < 1 > (2)
Database migration tasks	
Resource management	Name V Class
Replication instances	Empty replication instance table
Endpoints	You don't have any replication instances.
Certificates	

- 3. For Name, type a name for the replication instance that you will easily recognize.
- 4. For Description, type a description you will easily recognize. (e.g., DMS-Replication-Instance).
- 5. For Instance class, choose **dms.t2.micro**.
- 6. For VPC, choose the **dmslsstudv1** that you created earlier in pre-lab.

NOTE: Keep the existing default settings. (You may see a newer engine version than what is shown in the example image.)

AWS DMS	X DMS > Create replication instance
Dashboard	Create replication instance
<ul> <li>Conversion &amp; migration</li> <li>Database migration tasks</li> </ul>	Replication instance configuration
Database migration tasks    Replication instances  Endpoints Certificates Subnet groups Events Event subscriptions	Replication instance configuration         Name         The name must be unique among all of your replication instances in the current AWS region.         DMS-Replication-instance         Replication instance name must not start with a numeric value         Description         DMS replication instance for prod         The description must only have uniced letters, digits, whitespace, or one of these symbols:i/++.@. 1000         DMS replication instance for prod         The description must only have uniced letters, digits, whitespace, or one of these symbols:i/++.@. 1000         Italiance class         Onsets an appropriate instance class for your replication needs. Each instance class provides differing levels of compute, network and memory capacity.         If miss is based on DMS pricing [2].         Engine version         Choose an AWS DMS version to run on your replication instance.         3.1.3       Image: Compute symbols.         Jocose an aux 50 DMS version to run on your replication instance.         J.1.3       Image: Compute symbols.         Descret be amount of storage space you want for your replication instance. AWS DMS uses this storage for log files and cached transactions while replication tasks are in progress.         So       Image: Compute symbols.         Marcel       Image: Compute symbols.         So       Image: Compute symbols.         So
	Choose an Amazon Virtual Private Cloud (VPC) where your replication instance should run.  vpc-0314e829ba12d9481 - dmslstudv1
	Multi AZ If you choose this option, AWS DMS will perform a multi-AZ deployment, with a primary instance in one availability zone (AZ) and a standby instance in another AZ. This configuration provides a highly available, fault-tolerant replication environment.

- 7. Click **Advanced** to expand the section.
- 8. Select the security group with **sgdefault** in the name.

AWS DMS	×	Publicly accessible If you choose this option, AWS DMS will assign a public IP address to your replication instance, and you'll be able to connect to databases outside of your Amazon VPC.		
Dashboard				
<ul> <li>Conversion &amp; migration</li> <li>Database migration tasks</li> </ul>		<ul> <li>Advanced security and network configuration</li> </ul>		
<ul> <li>Resource management</li> <li>Replication instances</li> <li>Endpoints</li> </ul>		Replication subnet group Choose a subnet group for your replication instance. The subnet group defines the IP ranges and subnets that your replication instance can use within the Amazon VPC you've chosen. dms-lab-subnet-grp		
Certificates Subnet groups Events		Availability zone Choose an availability zone (AZ) where you want your replication instance to run. The default is "No preference", meaning that AWS DMS will determine which AZ to use.		
Event subscriptions		VPC security group(s) Choose one or more security groups for your replication instances. The security group(s) specify inbound and outbound rules to control network access to your replication instance. Use default dmslab-student-sgdefault-G2VY06TNTMNZ		
		(Default) aws/dms  Account		
		Description Key ARN		
		► Maintenance		
			Cancel	Create

- 9. Click **Create**.
- 10. The DMS console displays **creating** for the instance status. When the replication instance is ready, the status changes to **available**.

AWS DMS $\times$	DMS > Replication Instance
Dashboard	Replication instances (1) Create replication instance
Conversion & migration	Q. Find replication instance < 1 > Ø
Database migration tasks	
Resource management	Name     VPC     VPC     VPC     VP     VPL     V
Replication instances	dms-replication-Instance dms.t2.micro 🙆 vailable 3.1.3 us-east-1a vpc-0314e829ba12d9481 Yes 3.213.132.171 10.0.0.7
Endpoints	

# Create the Source Endpoint

1. On the DMS console, select Endpoints.

AWS DMS	×	DMS > Endpoint	
Dashboard		Endpoints	C Actions  Create endpoint
▼ Conversion & migration		Q Find endpoint	< 1 > @
Database migration tasks			
Resource management		Name V Type V Status V Engine V Server name V Port V Migration Hub Mapping	v ARN v Certificate ARN v
Replication instances		Empty endpoint table	
Endpoints		You don't have any endpoints.	
Certificates			

- 2. Click Create endpoint.
- 3. On the Create endpoint page, for Endpoint type, select **Source**.
- 4. For Endpoint identifier, select your easily recognized name.
- 5. For Source engine, select **postgres**.
- 6. Enter the **Server name** provided by your instructor, or if you ran instructor lab then take recorded endpoint from the same pre-lab.
- 7. For Port, enter **5432**.
- 8. For SSL mode, choose **none**.
- 9. For User name, type **adminuser**.
- 10. For Password, type **admin123**.
- 11. For Database name, type **sportstickets**.

AWS DMS $\times$	Create endpoint	
Dashboard	Endpoint type Info	
<ul> <li>Conversion &amp; migration</li> </ul>		
Database migration tasks	<ul> <li>Source endpoint</li> </ul>	Target endpoint
<ul> <li>Resource management</li> </ul>	A source endpoint allows AWS DMS to read data from a database (on-premises or in the cloud), or from other data source such as Amazon S3.	A target endpoint allows AWS DMS to write data to a database, or to other data source.
Replication instances		
Endpoints	Select RDS DB instance	
Certificates		
Subnet groups		
Events		
Event subscriptions	Endpoint configuration	
	Endpoint identifier Info A label for the endpoint to help you identify it.	
	prodendpoint-postgre	
	Source engine The type of database engine this endpoint is connected to.	
	postgres	▼
	Server name	
	dmslabinstance.c1ny3gywsvdz.us-east-1.rds.amazona	ws.com
	Port The port the database runs on for this endpoint.	Secure Socket Layer (SSL) mode The type of Secure Socket Layer enforcement
	5432	none 🔻
	User name Info	Password Info
	master	
	sportstickets	

12. Expand the **Test endpoint connection (optional) section**, and choose your DMS Lab VPC name on the VPC drop-down list.

13. Click **Run test**. This step tests connectivity to the source database system. If successful, the message "Connection tested successfully" appears.

AWS DMS	×	Database name sportstickets			
Dashboard					
<ul> <li>Conversion &amp; migration</li> <li>Database migration tasks</li> </ul>		Endpoint-specific sett	ings		
<ul> <li>Resource management</li> </ul>					
Replication instances		KMS master key			
Endpoints					
Certificates		<ul> <li>Test endpoint connect</li> </ul>	ion (optional)		
Subnet groups					
Event subscriptions		Test your endpoint connection b After clicking "Run test", an endp connect to the instance. If the co aren't saved will be deleted. VPC	y selecting a replication instance within yu soint will be created with the details provi nnection fails, you can edit and test it aga	our desired VPC. ded and attempt to ain. Endpoints that	
		vpc-0314e829ba12d9481 - dr	nslstudv1 💌		
		Replication instance A replication instance performs the di dms-replication-instance	stabase migration		
		Run test After tiicking "Run test", an endpoint faïks, you can edit and test it again. Er	will be created with the details provided and atte dpoints that aren't saved will be deleted.	impt to connect to the inst	ance. If the connection
		Endpoint identifier	Replication instance	Status	Message
		prodendpoint-postgre	dms-replication-instance	successful	
				Cancel	Create endpoint

- 14. Click **Create endpoint** to create the endpoint.
- 15. When available, the endpoint status changes to **active**.

AWS DMS $\times$	DMS > Endpoint	
Dashboard	Endpoints (1) C Action	ns V Create endpoint
Conversion & migration	Q Find endpoint	< 1 > 💿
Database migration tasks		
Resource management	□ Name ▼ Type ▼ Status ▼ Engine ▼ Server name ▼ Port ▼ Migration Hub Mapping ▼ ARN	
Replication instances	prodendpoint-postgre     Source     I     OActive     PostgreSQL     dmslabinstance.r1ny3gywsvdz.us-east-1.rds.amazonaws.com     5432     arn:aw	vs:dms:us-east-1:341259728059:endp
Endpoints		

### IAM Policy for DMS->S3 Access

Now that we have created the source endpoint from which we want to replicate and/or export data from, we now need a security policy and role that DMS can run under to store the results against our target.

**The policy and role have been created for you** in the student prelab by an AWS CloudFormation template, with a permission set that allows the DMS service to access the S3 bucket.

Below IAM policy for the IAM role granted to your S3 bucket endpoint, enabling DMS to write to the S3 bucket. This policy grants *GetObject, PutObject, DeleteObject and ListBucket* to a bucket with a name that starts with dmslab. See the following code for an example:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "s3:GetObject",
        "s3:PutObject",
        "s3:DeleteObject"
      ],
      "Resource": [
        "arn:aws:s3:::dmslabstudent-dmslabs3bucket-1eegnc2tj056l/*"
      ],
      "Effect": "Allow"
    },
    {
      "Action": [
        "s3:ListBucket"
      ],
      "Resource": [
        " arn:aws:s3:::dmslabstudent-dmslabs3bucket-1eegnc2tj056l"
      ],
      "Effect": "Allow"
    }
 ]
}
```

Explore the IAM Role by following below steps:

- 1. On the IAM console, select Roles.
- 2. On the Roles page, in the search box, type **dmslab** to filter the results.
- 3. Click the **DMSLabRoleS3** role name.

Search IAM	IAM Roles Documentation     Tutorial: Setting Up Cross	Account Access	
Dashboard	Common Scenarios for Ro	bles	
Groups			
Users	Create role Delete role		
Roles			
Policies	Q dmslab		
Identity providers	Role name 👻	Description	Trusted entities
Account settings	dmslab-student-DMSI a	bBo	AWS service: dms
Credential report			

```
Encryption keys
```

### 4. Look at the **ROLE ARN** value for this role.

Search IAM	Roles > dmslab-	-student-DMSLabRoleS3-	1SR7IR2GC5VD5			Delete role
Dashboard Groups Users Roles Policies Identity providers Account settings	Maxi	Role ARN Role description Instance Profile ARNs Path Creation time mum CLI/API session duration	arn:aws:iam::34125 Edit 2 / 2018-09-29 21:51 F 1 hour Edit	9728059:role/dmslab-sl DT	udent-DMSLabRoleS3-1SR7IR2GC5VD5	2
Credential report	Permissions	Trust relationships	Access Advisor	Revoke sessions		
Encryption keys	<ul> <li>Permissi</li> <li>Attach polic</li> </ul>	ons policies (1 polic	y applied)			Add inline policy
	Policy	name 👻			Policy type 👻	
	► DM	SLabS3Policy			Inline policy	×

5. Expand the *DMSLabS3Policy*, record the **s3 bucket name** from the Resource section. NOTE: don't copy the ARN prefix '*arn:aws:s3:::*'

DMSLabS3Policy Policy summary {} JSON Edit policy 11 "arn:aws:s3:::dmslabstudent-dmsla	Inline policy Simulate
Policy summary {} JSON Edit policy	Simulate
11 "arn:aws:s3:::dmslabstudent-dmsla	
	bs3bucket-1eegnc2tj056l/*"
12 ],	
13 "Effect": "ALLow"	
15 - 5	
16 "Action": E	
17 "s3:ListBucket"	
18 ],	
19- "Resource": [	
20 "arn:aws:s3:::dmslabstudent-dmsla	bs3bucket-1eegnc2tj056l"
21 ],	
22 "Effect": "Allow"	

## Create the Target Endpoint

Before start, make sure you have the following information on hand:

- **DMSLabRoleS3 ARN** arn:aws:iam::xxxx:role/xxxxx
- **s3 Bucket Name** xxxx-dmslabs3bucket-xxxxx
- 1. On the DMS console, select Endpoints.

AWS DMS	×	DMS > Endpoint	
Dashboard		Endpoints	C Actions V Create endpoint
Conversion & migration		Q. Find endpoint	< 1 > @
Database migration tasks			
<ul> <li>Resource management</li> </ul>		Name V Type V Status V Engine V Server name V Port V Migration Hub Mapping	▼ ARN ▼ Certificate ARN ▼
Replication instances		Empty endpoint table	
Endpoints		You don't have any endpoints.	
Certificates			

- 2. Click Create endpoint.
- 3. For Endpoint type, select Target.
- 4. For Endpoint identifier, type an easily recognized name.
- 5. For Target engine, choose **s3**.
- 6. For Service access role ARN, paste the DMSLabRoleS3 ARN number
- 7. For Bucket name, paste the **s3 bucket name**
- 8. For Bucket folder, type tickets.

AWS DMS	×	DMS > Create endpoint
Dashboard		Create endpoint
<ul> <li>Conversion &amp; migration</li> <li>Database migration tasks</li> </ul>		Endpoint type Info
<ul> <li>Resource management</li> <li>Replication instances</li> <li>Endpoints</li> <li>Certificates</li> <li>Subnet groups</li> <li>Events</li> </ul>		Source endpoint     A source endpoint allows AWS DMS to read data from a     database (on-premises or in the cloud), or from other data     source such as Amazon 53.     Select RDS DB instance
Event subscriptions		
		Endpoint configuration
		Endpoint identifier Info A label for the endpoint to help you identify it. Target engine Target engine Target engine this endpoint is connected to. \$3 Service access role ARN Role that can access target arr:aws:iam::341259728059:role/dmslab-student-DMSLabRole53-8DVW2RR7J7QZ Bucket name The name of an Amazon 53 bucket where DMS will read the files from
		dmslab-student-dmslabs3bucket-woti4bf73cw3
		Bucket folder The Amazon 53 bucket path where the CSV files can be found tickets

- 9. Click Endpoint-specific settings to expand the section.
- 10. In the **Extra connection attributes** box, type **addColumnName=true**. This attribute includes the column names in the files in the S3 bucket.

- 11. Expand the **Test endpoint connection (optional) section**, and choose your **dmslstudv1** name on the **VPC** drop-down list.
- 12. Click **Run test**. This step tests connectivity to the source database system. If successful, the message "Connection tested successfully" appears.
- 13. Click **Create Endpoint**. If the button is grey out, just click **Cancel** button.

AWS DMS	×				,
Dashboard		<ul> <li>Endpoint-specific set</li> </ul>	ttings		
<ul> <li>Conversion &amp; migration</li> </ul>			-		
Database migration tasks		Extra connection attributes	rameters here. See the documentation for more in	formation.	
<ul> <li>Resource management</li> </ul>		addColumnName=true	_		
Replication instances		• • • • • • • • • • • • • • • • • • •			
Endpoints					
Certificates					
Subnet groups		<ul> <li>Test endpoint conner</li> </ul>	ection (optional)		
Events					
Event subscriptions		Test your endpoint connection After clicking "Run test", an er connect to the instance. If the aren't saved will be deleted. VPC vpc-0314e829ba12d9481 - Replication instance A replication instance dms-replication-instance	n by selecting a replication instance within indpoint will be created with the details pro connection fails, you can edit and test it a dmslstudv1 e database migration	your desired VPC. vided and attempt to gain. Endpoints that	
		After clicking "Run test", an endpo fails, you can edit and test it again Endpoint identifier	Int will be created with the details provided and at Endpoints that aren't saved will be deleted.  Replication instance	tempt to connect to the ins	Message
		targetendpoint	dms-replication-instance	successful	
				Cancel	Create endpoint

When available, the endpoint status changes to **active**.

AWS DMS	×	DM	S > Endpo	int											
Dashboard			Endpoints	(2)									C	Actions v	Create endpoint
Conversion & migration			Q Find end	ipoint											< 1 > 💿
Database migration tasks															
<ul> <li>Resource management</li> </ul>			Nam	e		Туре 🔻	Status 🔻	Engine 🔻	Server name	Ŧ	Port v	Migration Hub	Mapping 🔻	ARN	
Replication instances			prode	endpoint-pos	stgre	Source	⊘ Active	PostgreSQL	dmslabinstance.c1ny3gywsvd	dz.us-east-1.rds.amazonaws.com	5432			arn:aws:dms:us	-east-1:341259728059:endp
Endpoints			targe	tendpoint		Target	⊘ Active	Amazon S3	-					arn:aws:dms:us	-east-1:341259728059:endp
Certificates		-	-			-	-								

## Create a task to perform the initial full copy

1. On the DMS console, select **Database Migration Tasks**.

AWS DMS	:	DMS > Database migration tasks	
Dashboard		Database migration tasks	ì
Conversion & migration		Q. Find task < 1 > @	
Database migration tasks			
<ul> <li>Resource management</li> </ul>		Rame v Status v Source v Target v Type v Progress v Elapsed time v Tables loaded v Tables loading v Tables queued v Tables errored v	
Replication instances		Empty replication task table	
Endpoints		You don't have any replication tasks.	
Certificates			_

2. Click Create Task.

- 3. Type an easily recognized **Task name**.
- 4. Select your **Replication instance**.
- 5. Select your **Source endpoint**.
- 6. Select your Target endpoint.
- 7. For Migration type, choose **Migrate existing data**.
- 8. Select the **Start task on create** check box.

AWS DMS $\times$	DMS > Create replication task
Dashboard	Create data migration task
<ul> <li>Conversion &amp; migration</li> <li>Database migration tasks</li> </ul>	Task configuration
Resource management	Task identifier
Replication instances	DMS-task
Endpoints Certificates	Replication instance
Subnet groups	dms-replication-instance - vpc-0314e829ba12d9481
Events	Source database endpoint
Event subscriptions	prodendpoint-postgre
	Target database endpoint
	targetendpoint
	Migration type Info
	Migrate existing data
	Start task on create

- 9. Expand Task Settings.
- 10. Select the Enable CloudWatch logs check box.

AWS DMS X	Task settings
Dashboard	
<ul> <li>Conversion &amp; migration</li> <li>Database migration tasks</li> </ul>	Target table preparation mode Info <ul> <li>Do nothing</li> <li>Drop tables on target</li> </ul>
Resource management	O Truncate
Replication instances Endpoints Certificates Subnet groups Events	Include LOB columns in replication Info <ul> <li>Don't include LOB columns</li> <li>Full LOB mode</li> <li>Limited LOB mode</li> </ul> Maximum LOB size (KB) Info
Event subscriptions	32         Enable validation         Choose this setting if you want AWS DMS to compare the data at the source and the target, immediately after it performs a full data load. Validation ensures that your data was migrated accurately, but it requires additional time to complete.         I Image: Enable CloudWatch logs       Info 1         Image: CloudWatch logs       Info 1         Image: CloudWatch logs       Info 1

- 11. Go to Table Mappings.
- 12. Click on Add new selection rule
- 13. For Schema name, select **dms\_sample** from drop down. Keep the settings for the remaining fields

<ul> <li>Guided UI</li> <li>Set up your table mapping rules using a step- interface.</li> </ul>	by-step guided	n more 🔀 ning rules directly, in JSON format.
ecify at least one selection rule with an inc	clude action. After you do this, you can add on	ne or more transformation rules.
Choose the schema and/or tables you wa migration task. Info	nt to include with, or exclude from, your	Add new selection rule
where schema name is like 'dms_same is like '	nple' and <b>table name</b> is like '%', include	o ×
Schema		
dms_sample		•
Table name Use the % character as a wildcard		
% Action		
Include	objects, or exclude to ignore them during the migra	▼
Source filters Info		Add column filter
Transformation rules		
Advanced task settings		

#### 14. Click Create task.

Your task is created and starts automatically. (Note: The complete creation and data extraction process takes 5 to 15 minutes.)

DMS > Data	base migration tasks									
Database	e migration tasks (1)								C Actions V	Create task
Nar	me 🔻 Status 🔻	Source 💌	Target 💌	Туре 🔻	Progress V	Elapsed time 🔻	Tables loaded 🔻	Tables loading 🔻	Tables queued 🔻	Tables errored 🔻
dms	s-task 💬 Running	prodendpoint-postgre	targetendpoint	Full load	66 %	4 m	15	1	0	0

Once complete, the console displays 100% complete. Select your task and explore the summary:

ms-task				C Actions			
Summary							
Status	Type	Source	Target				
Ø Load complete	Full load	prodendpoint-postgre	targetendpoint				
Overview details							
Basic configuration							
Task ARN	k:MUYIRRCLBYT45EZESVNFNGAUL4 🗗	Status	Status				
am:aws:dms:us-east-1:341259728059:tas		Status Complete	Ø Load complete				
Type		Replication instance	Replication instance				
Full load		dms-replication-instance	dms-replication-instance				
Source		Target	Target				
prodendpoint-postgre		targetendpoint	targetendpoint				
Last failure message		Created	Created				
-		5/29/2019, 10:55:15 AM GMT-0700	5/29/2019, 10:55:15 AM GMT-0700				
Started		Migration task logs Info	Migration task logs Info View logs 🔄				

Scroll down and you can observe all table information loaded in S3 from RDS by DMS

Table	e statistics (16)								C	Valldate again	Reload table data
Q /	Ind schema										< 1 > ©
	Schema name 🔻	Table 👻	Load state 🗢	Inserts 🔻	Deletes 🔻	Updates 🔻	DDLs V	Full load rows 🔻	Total 🔻	Validation state 🔻	Validation pending $=$
	dms_sample	seat_type	Table completed	0	0	0	0	6	6	Not enabled	0
	dms_sample	seat	Table completed	0	0	o	0	603,631	603,631	Not enabled	o
	dms_sample	mlb_data	Table completed	0	0	0	0	2,230	2,230	Not enabled	0
	dms_sample	player	Table completed	0	o	0	o	5,157	5,157	Not enabled	0
	dms_sample	ticket_purchase_hist	Table completed	0	0	0	0	6,038,756	6,038,756	Not enabled	0
	dms_sample	person	Table completed	0	o	0	o	7,025,584	7,025,584	Not enabled	0
	dms_sample	name_data	Table completed	0	0	0	0	5,360	5,360	Not enabled	0
	dms_sample	sport_team	Table completed	0	o	0	0	62	62	Not enabled	0
	dms_sample	sport_league	Table completed	0	0	0	0	2	2	Not enabled	0
	dms_sample	sporting_event	Table completed	0	o	0	o	1,158	1,158	Not enabled	0
	dms_sample	sporting_event_ticket	Table completed	0	0	0	0	15,212,460	15,212,460	Not enabled	0
	dms_sample	sport_division	Table completed	0	o	0	o	14	14	Not enabled	0
	dms_sample	sport_location	Table completed	0	0	0	0	62	62	Not enabled	0
	dms_sample	sport_type	Table completed	0	0	0	o	0	0	Not enabled	0
	dms_sample	nfl_stadium_data	Table completed	0	0	0	0	32	32	Not enabled	0
	dms_sample	nfl_data	Table completed	0	0	o	0	2,928	2,928	Not enabled	0

15. Open the S3 console and view the data that was copied by DMS.

Your S3 bucket name will look like below : BucketName/bucket\_folder\_name/schema\_name/table\_name/objects/

	Amazon S	S3 >	dmslab-stud	ent-dmslabs3buc	ket-woti4bf73cw3	>	tickets	>	dms_sample
		Overvie	ew						
ſ									
	<b>Q</b> Type	a prefix	and press Ent	er to search. Press	ESC to clear.				
	🛓 Uploa	d +	- Create folde	Download	Actions ~				
	Nan	ne 👻							
		mlb_da	ata						
		name_	data						
		nfl_dat	a						
		nfl_sta	dium_data						
		person	i i						
		player							
		seat							
		seat_ty	/pe						
		sport_c	division						
		sport_l	eague						
		sport_l	ocation						
		sport_t	team						
		sportin	ig_event						
		sportin	ig_event_ticke	t					
	□ ►	ticket_	purchase_hist						

- 16. Navigate to one of the files and review it using <u>S3 Select</u>:
  - a. Navigate in to the directory named \*\*player\*\* and select the check box next to the file name.
  - b. Click the Actions dropdown button and choose Query with S3 Select. Amazon 53 > mod-3fccddd609114925-dmslabs3bucket-1ut2vprjgnoe1 > tickets/ > dms\_sample/ > player/

player/		Copy S3 URI
Objects Properties		
Objects (1)         Objects are the fundamental entities stored in Amazon 53. You can use Amazon 53 inventory [2] to get a list of all objects in yo grant them permissions. Learn more [2]         C       C Copy S3 URI       C Copy URL       Download       Open [2]       Delete         Q. Find objects by prefix       Image: Copy Copy Copy Copy Copy Copy Copy Copy	ur bucket. For others to access your obje Actions ▲ Create fold Download as Calculate total size Copy Move Initiate restore Query with 53 Select Edit actions Rename object Edit storage class Edit server-side encryption Edit metadata Edit metadata	er 🕞 Upload < 1 > 💿 Storage class $\nabla$ Standard
	Make public	

c. In the Query with S3 Select page, leave the default value for *Input Settings* and SQL Query and click **Run SQL query**.

d. It will execute the specified SQL query and return the first 5 lines from the CSV file.



You will notice that the file contains the column headers in the first row as requested by the **"addColumnName=true"** connection attribute we included when we created the s3 target endpoint. Note that column names are included in the file in the first row.

### Explore the objects in the S3 directory further.

(Optional) Create the CDC endpoint to replicate ongoing changes

As of now we are enabling only one schema replication for CDC

1. On the DMS console, select **Endpoints**.

AWS DMS $\times$	DMS > Endpoint	
Dashboard	Endpoints	C Actions  Create endpoint
Conversion & migration	Q Find endpoint	< 1 > @
Database migration tasks		
<ul> <li>Resource management</li> </ul>	Name V Type V Status V Engine V Server name V Port V Migration Hub Mapping	V ARN V Certificate ARN V
Replication instances	Empty endpoint table	
Endpoints	You don't have any endpoints.	
Certificates		

- 2. Click **Create endpoint**.
- 3. For Endpoint type, select **Target**.
- 4. For Endpoint identifier, type an easily recognized name that includes "cdc".
- 5. For Target engine, choose **Amazon S3**.
- 6. For Service Access Role ARN, paste the ARN value that you copied in the IAM role console group.

NOTE: The value is similar to the following string, where the account number is specific to your account number: "arn:aws:iam::119911911299:role/data-eng-dms-role"

- 7. For Bucket name, type the name of the s3 bucket you noted down from pre-lab.
- 8. For Bucket folder, type **cdc** and For CDC path , leave blank

<ul> <li>Source endpoin</li> </ul>	It Target endpoint
database (on-pren	allows Aws DMs to read data from a A target endpoint allows Aws DMs to write data to mises or in the cloud), or from other data data and database, or to other data source.
source such as Am	lazon \$3.
Select RDS DB ins	stance
Endpoint config	guration
Endpoint identifier	Info
A label for the endpoint	to help you identify it.
cdcendpoint	
cdcendpoint	
cdcendpoint Target engine The type of database env	gine this endpoint is connected to.
cdcendpoint Target engine The type of database eng	gine this endpoint is connected to.
cdcendpoint Target engine The type of database eng \$3	gine this endpoint is connected to.
cdcendpoint Target engine The type of database engine \$3 Service access role Af	gine this endpoint is connected to.
cdcendpoint Target engine The type of database en- s3 Service access role AR Role that can access targ	gine this endpoint is connected to.
cdcendpoint Target engine The type of database en \$3 Service access role Af Role that can access targ am:aws:lam::34125	gine this endpoint is connected to. RN jet i9728059:role/dmslab-student-DMSLabRole53-6DVW2RR7J7QZ
cdcendpoint Target engine The type of database en \$3 Service access role Af Role that can access targ arn:aws:iam::34125 Bucket bases	gine this endpoint is connected to. RN pet 19728059:role/dmslab-student-DMSLabRoleS3-8DVW2RR7J7QZ
cdcendpoint Target engine The type of database eng \$3 Service access role Af Role that can access targ arn:aws:lam::34125 Bucket name The name of an Amazon	gine this endpoint is connected to.  RN pet 19728059role/dmslab-student-DMSLabRoleS3-8DVW2RR7J7QZ SS bucket where DMS will read the files from
cdcendpoint Target engine The type of database en \$3 Service access role Af Role that can access target arr:aws:iam:341255 Bucket name The name of an Amazon dmslab-student-dm	gine this endpoint is connected to. RN jet S3 Ducket where DMS will read the files from rstabs5bucket-worl40/f73cvt3
dcendpoint Target engine The type of database em \$3 Service access role Ai Role that can access targ arn:aws:lam::34125 Bucket name The name of an Amazon dmslab-student-dm	gine this endpoint is connected to.  RN get 19728059:role/dmslab-student-DMSLabRoleS3-8DVW2RR7J7QZ 153 bucket where DMS will read the files from nslabs3bucket-woti4bf73cw3
dcendpoint Target engine The type of database en \$3 Service access role Ai No Role that can access arg arrcaws:lam::34125 Bucket name The name of an Amazon dmsiab-student-dm Bucket folder	igine this endpoint is connected to.  RN  ret i9728059:role/dmslab-student-DMSLabRoleS3-8DVW2RR7J7QZ  S3 bucket where DMS will read the files from nslabs3bucket-woli4bf73cw3
deendpoint Target engine The type of database en § 3 Service access role Al Role that can access targ arr:aws:iam::34125 Bucket name The name of an Amazon dmslab-student-drn Bucket folder The Amazon S3 bucket p	igine this endpoint is connected to.  RN pet 19728059:role/dmslab-student-DMSLabRoleS3-8DVW2RR7J7QZ 155 bucket where DMS will read the files from rslabs3bucket-wol4bf73cw3 aith where the CSV files can be found
dcendpoint Target engine The type of database engine § 3 Role that can access targ arri:aws:lam::34125 Bucket name The name of an Amazon dmslab-student- Bucket folder The Amazon 35 bucket pi	ighe this endpoint is connected to.

- 9. Click **Endpoint-specific settings** to expand the section.
- 10. In the **Extra connection attributes** box, type **addColumnName=true**. This attribute includes the column names in the files in the S3 bucket.
- 11. Expand the **Test endpoint connection (optional)** section, and choose your **dmslstudv1** name on the VPC drop-down list.
- 12. Click Run test. This step tests connectivity to the source database system. If successful, the message "Connection tested successfully" appears.

<ul> <li>Endpoint-specific setting</li> </ul>	s						
Extra connection attributes Type any additional connection parameter addColumnName=true	s here. See the documentation for more	information.					
▼ Test endnoint connection	(optional)						
Test your endpoint connection by se After clicking "Run test", an endpoin connect to the instance. If the conne aren't saved will be deleted. VPC	lecting a replication instance within t will be created with the details pr cction fails, you can edit and test it	n your desired VPC. ovided and attempt to again. Endpoints that					
vpc-0314e829ba12d9481 - dmsistudv1							
Replication instance A replication instance performs the databa	ise migration						
dms-replication-instance	▼						
Run test After clicking "Run test", an endpoint will i fails, you can edit and test it again. Endpo	pe created with the details provided and nts that aren't saved will be deleted.	attempt to connect to the insta	nce. If the connection				
Endpoint identifier	Replication instance	Status	Message				
cdcendpoint	dms-replication-instance	successful					
		Cancel	Create endpoint				

- 13. Click **Create endpoint**.
- 14. When available, the endpoint status changes to active.

MS >	Endpoint							
Endp	points (3)						C	Actions  Create endpoint
Q	Find endpoint							< 1 > 🐵
	Name 🔻	Туре 🔻	Status 🔻	Engine 🔻	Server name v	Port V	Migration Hub Mapping 🔻	ARN
<	cdcendpoint	Target	⊘ Active	Amazon S3	-	-		arn:aws:dms:us-east-1:341259728059:end
	prodendpoint-postgre	Source	⊘ Active	PostgreSQL	dmslabinstance.c1ny3gywsvdz.us-east-1.rds.amazonaws.com	5432		arn:aws:dms:us-east-1:341259728059:end
	targetendpoint	Target	⊘ Active	Amazon S3	-	-		arn:aws:dms:us-east-1:341259728059:end

(Optional) Create a task to perform the ongoing replication

1. On the DMS console, select **Database Migration Tasks**.

AWS DMS	DMS 🗦 Database migration tasks	
Dashboard	Database migration tasks	C Actions V Create task
<ul> <li>Conversion &amp; migration</li> </ul>	Q. Find task	< 1 > @
Database migration tasks		
<ul> <li>Resource management</li> </ul>	Name V Status V Source V Target V Type V Progress V Elapsed time V Tables loaded V Tables loading V	Tables queued v Tables errored v
Replication instances	Empty replication task table	
Endpoints	You don't have any replication tasks.	
Certificates		

- 2. Click Create Task.
- 3. Type an easily recognized **Task Identifier**. For example "cdctask".
- 4. Select your **Replication instance**.
- 5. Select your **Source endpoint**.
- 6. Select your **Target endpoint** as cdc endpoint created in previous section.
- 7. For **Migration type**, choose **Replicate data changes only**.
- 8. Select the Start task on create check box.

AWS DMS	×	DMS > Create replication task
Dashboard		Create data migration task
<ul> <li>Conversion &amp; migration</li> <li>Database migration tasks</li> </ul>		Task configuration
<ul> <li>Resource management</li> <li>Replication instances</li> <li>Endpoints</li> <li>Certificates</li> <li>Subnet groups</li> <li>Events</li> <li>Event subscriptions</li> </ul>		Task identifier         cdctask         Replication instance         dms-replication-instance - vpc-0314e829ba12d9481         Source database endpoint         prodendpoint-postgre         Target database endpoint         cdcendpoint         Migration type Info         Replicate data changes only         ✓         Start task on create

9. In Task Settings, Select the Enable CloudWatch logs check box.

AWS DMS $\times$	Task settings
Dashboard	
Conversion & migration	Target table preparation mode Info
Database migration tasks	Do nothing     Dran tables on target
Resource management	
Replication instances	Include LOB columns in replication Info
Endpoints	<ul> <li>Don't include LOB columns</li> </ul>
Certificates	Full LOB mode
Subnet groups	Limited LOB mode
Events	Maximum LOB size (KB) Info
Event subscriptions	32
	<ul> <li>Enable validation</li> <li>Choose this setting if you want AWS DMS to compare the data at the source and the target, immediately after it performs a full data load. Validation ensures that your data was migrated accurately, but it requires additional time to complete.</li> <li>Enable CloudWatch logs Info</li> <li>CloudWatch logs usage will be charged at standard rates. See here for more details.</li> </ul>

- 10. Go to **Table Mappings**.
- 11. Click on Add new selection rule
- 12. For **Schema name**, select **dms\_sample** from drop down. Keep the settings for the remaining fields

int	t up your table mapping rules using a step-by-step guided terface.	nore 🖸 rules directly, in JSON format.
oecify Sel	at least one selection rule with an include action. After you do this, you can add one o ection rules	r more transformation rules.
Cho mig	ose the schema and/or tables you want to include with, or exclude from, your ration task. Info	Add new selection rule
•	where schema name is like 'dms_sample' and table name is like '%', include	۵ ×
	Schema	
	dms_sample	•
	Use the % character as a wildcard %	
	Action Choose "Include" to migrate your selected objects, or "Exclude" to ignore them during the migration	n.
	Include	•
	Source filters Info	Add column filter
Tra	nsformation rules	
Adv	vanced task settings	

- 13. Click Create task.
- 14. Your task is created and starts automatically. You can see status as ongoing replication, after couple of minutes.

oms >	Database migr	ration tasks								
Data	base migrat	tion tasks (2)							C Actions	Create task
Q	Find task									< 1 > ©
	Name 🔻	Status 🔻	Source v	Target 🔻	Туре 🔻	Progress v	Elapsed time 🔻	Tables loaded ▼	Tables loading ▼	Tables queued v
	dms-task	O Load complete	prodendpoint-postgre	targetendpoint	Full load	100 %	9 m	16	0	0
	newcdc	Replication ongoing	prodendpoint-postgre	cdcendpoint	Ongoing replication	100 %	0 m	16	0	0

Once complete, the console displays 100% complete.

15. Your instructor will generate CDC activity which above migration task will capture, if you ran instructor setup by own, then make sure to follow "Generate the CDC Data" section from instructor lab.

You may need to wait 5 to 10 minutes for CDC data to first reflect in your RDS postgre database and then picked up by DMS CDC migration task.

16. Select your CDC task and explore the summary:

newcdc			
Summary			
Status O Replication ongoing	Type Ongoing replication	Source prodendpoint-postgre	Target cdcendpoint
Overview details			
Basic configuration			
Task ARN arn:aws:dms:us-east-1:341259728059:task:Y5L3X5DAFT6B;	7FSEQYFGJ45TUQ 🗇	Status	
Type Ongoing replication		Replication instance dms-replication-instance	
Source prodendpoint-postgre		Target cdcendpoint	
Last failure message -		Created 5/29/2019, 3:15:19 PM GMT-0700	
Started 5/29/2019, 4:01:28 PM GMT-0700		Migration task logs Info View logs [건	
Change data capture (CDC)			
Change data capture (CDC) start position -		Change data capture (CDC) stop position	
Change data capture (CDC) recovery checkpoint			

Scroll down and you will see all table changes impacted by CDC:

Tabl	e statistics (16)									7 Validate again	Reload table data
Q	Find schema										< 1 > ©
	Schema name 🔻	Table 🔻	Load state 🛛 🔻	Inserts 🔻	Deletes 🔻	Updates 🔻	DDLs V	Full load rows 🔻	Total 🔻	Validation state 🔻	Validation pending <b>v</b>
	dms_sample	seat_type	Table completed	0	0	0	0	0	0	Not enabled	0
	dms_sample	seat	Table completed	0	0	0	0	0	0	Not enabled	0
	dms_sample	mlb_data	Table completed	0	0	0	0	0	0	Not enabled	0
	dms_sample	player	Table completed	0	0	0	0	0	0	Not enabled	0
	dms_sample	ticket_purchase_hist	Table completed	680,218	0	0	0	0	680,218	Not enabled	0
	dms_sample	person	Table completed	0	0	0	0	0	0	Not enabled	0
	dms_sample	name_data	Table completed	0	0	0	0	0	0	Not enabled	0
	dms_sample	sport_team	Table completed	0	0	0	0	0	0	Not enabled	0
	dms_sample	sport_league	Table completed	0	0	0	0	0	0	Not enabled	0
	dms_sample	sporting_event	Table completed	0	0	0	0	0	0	Not enabled	0
	dms_sample	sporting_event_ticket	Table completed	0	0	680,218	0	0	680,218	Not enabled	0
	dms_sample	sport_division	Table completed	0	0	0	0	0	0	Not enabled	0
	dms_sample	sport_location	Table completed	0	0	0	0	0	0	Not enabled	0
	dms_sample	sport_type	Table completed	0	0	0	0	0	0	Not enabled	0
	dms_sample	nfl_stadium_data	Table completed	0	0	0	0	0	0	Not enabled	0
	dms_sample	nfl_data	Table completed	0	0	0	0	0	0	Not enabled	0

17. Open the S3 console and view the CDC data that was copied by DMS.

Your S3 bucket name will look like below : BucketName/bucket\_folder\_name/schema\_name/table\_name/objects/

In our lab example this becomes:

"/dmslab-student-dmslabs3bucket-woti4bf73cw3/cdc/dms\_sample" with a separate path for each table\_name)

Amazon S3 > mod-3fccddd609114925-dmslabs3bucket-1u	t2vprjqnoe1 >	cdc/ > dms_sample/			
dms_sample/					🗇 Copy S3 URI
Objects Properties					
Objects (2)         Objects are the fundamental entities stored in Amazon S3. You can u grant them permissions. Learn more 2         C       C Copy S3 URI	se Amazon 53 invent	tory 🕻 to get a list of all objects in y	our bucket. For others to Actions	access your objects, yo Create folder	u'll need to explicitly 다 Upload
<b>Q</b> Find objects by prefix					< 1 > ©
Name 🔺	Туре 🗸	Last modified	⊽ Size		:lass ⊽
sporting_event_ticket/	Folder	-			
ticket_purchase_hist/	Folder	-			

- 17. Navigate to one of the files and review it using <u>S3 Select</u>:
  - a. Navigate in to the directory named \*\*player\*\* and select the check box next to the file name.
  - b. Click the Actions dropdown button and choose Query with S3 Select. Amazon S3 > mod-3fccddd609114925-dmslabs3bucket-1ut2vprjqnoe1 > cdc/ > dms\_sample/ > sporting\_event\_ticket/

porting_event_ticket/		① Copy S3 URI
Objects Properties		
Objects (1)         Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory [2] to get a list of all objects in grant them permissions. Learn more [2]         Of       Image: Comparison of the permission of the	your bucket. For others to access your obje	cts, you'll need to explicitly
Q     Find objects by prefix	Create fold Download as Calculate total size	< 1 > ©
✓ Name ▲ Type マ Last modified	Сору	Storage class 🛛 🗸
20210915-085904798.csv csv September 15, 2021, 10:59:05 (UTC+	0. Initiate restore	Standard
	Query with S3 Select	
	Edit actions	
	Rename object	
	Edit storage class	
	Edit server-side encryption	
	Edit metadata	
	Edit tags	
	Make public	

c. In the Query with S3 Select page, leave the default value for Input Settings and SQL Query and click Run SQL query.

Query with S3 Select Info
Ise Amazon 53 Select to retrieve a subset of data from an object using standard SQL queries. Pricing is based on the size of the input, query results, and data transferred. Learn more 🖄 or se mazon 53 pricing 🔀
Input settings
Path
s3://mod-3fccddd609114925-dmslabs3bucket-1ut2vprjqnoe1/cdc/dms_sample/sporting_event_ticket/20210915-085904798.csv
Size
5.8 MB (6092630.0 B)
Format
O CSV
Apache Parquet
CSV delimiter
Exclude the first line of CSV data Enable this setting if CSV contains a header row.
Compression
O None
○ GZIP
O BZIP2
Output settings
Format
© CSV
Nozi 🔾
CSV delimiter
SOL query
Amazon 53 Select supports only the SELECT SQL command. Using the 53 console, you can extract up to 40 MB of records from an object that is up to 128 MB in size. To work with larger files or more records, use the
ATT 2 CLI,
Add SQL from templates Run SQL query
1 /* To create reference point for writing SQL queries, you can display the first 5 records of input data by running the following SQL query: SELECT * FROM s3object s LIMIT 5 */
2 SELECT * FROM s3object s LIMIT 5

d. It will execute the specified SQL query and return the first 5 lines from the CSV file.

Query results Query results are not available after you choose Close or navigate away. Choose Download results to download a copy of the following query results.	Download results
Status Stacessfully returned 5 records in 225 ms Bytes returned: 561 B	
Raw Formatted	
Op,id,sporting_event_id,sport_location_id,seat_level,seat_section,seat_row,seat,ticketholder_id,tick U,+1.2655711000000000e+07,+1.441000000000000e+03,+9.0000000000000000e+00,2,10,A,2,+4.91217500000000 U,+1.2655721000000000e+07,+1.4410000000000000e+03,+9.000000000000000e+00,2,10,A,1,+4.91217500000000 U,+1.265781000000000e+07,+1.4410000000000000e+03,+9.000000000000000e+00,2,10,A,2,+4.91217500000000 U,+1.265208100000000e+07,+1.4410000000000000e+03,+9.00000000000000e+00,2,10,A,1,+4.91217500000000	et_price 00e+06,43.23 00e+06,43.23 00e+06,86.46 00e+06,43.23

You will notice that the file contains the column headers in the first row as requested by the "addColumnName=true" connection attribute we included when we created the s3 target endpoint.

Note that file name has date time - 20210915-085904798.csv

You can see the header is included and the operation column is added at the beginning of each row. The file below shows updates (U) to the table along with the values after the update. Inserts (I) show data after the insert and Deletes (D) show data before the delete.

Explore the objects in the S3 directory further.