



Amazon Web Services Data Engineering Immersion Day

Database Migration Services Lab
September 2021

Table of Contents

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

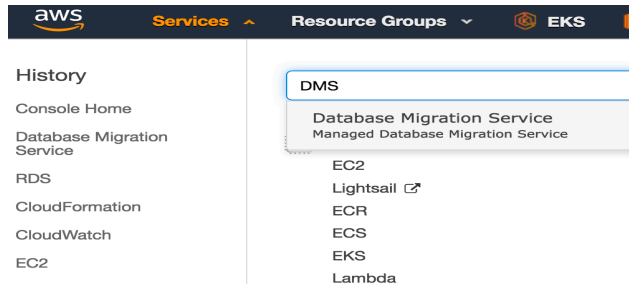
Optionally, you can add ongoing replication of data changes on the source: *(Only one of the DMS replication instances will enable this feature.)*

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

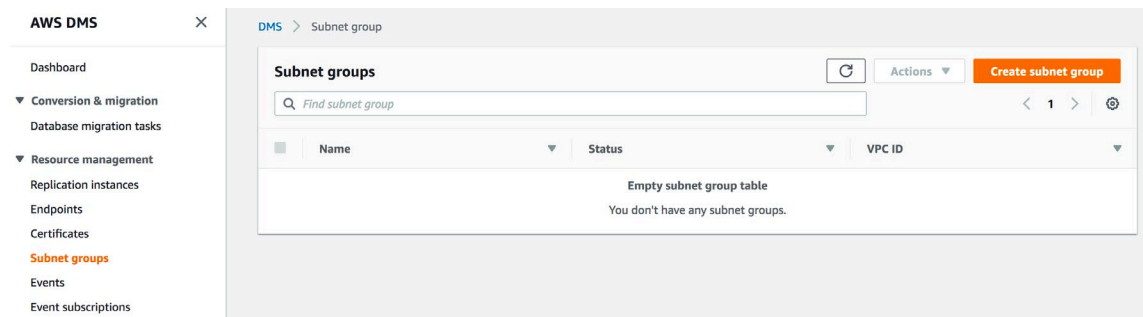
Your instructor has already created and populated the RDS Postgres database that you will use as your source endpoint in this lab.

Create the Subnet Group

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

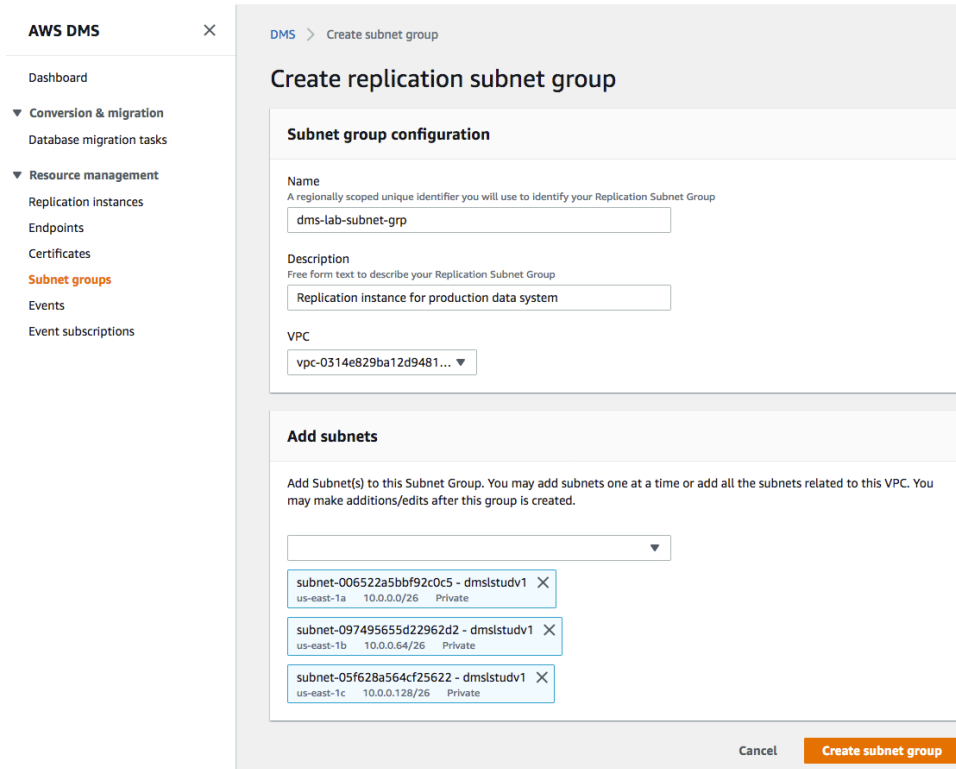


2. On the DMS console, select **Subnet Groups**.



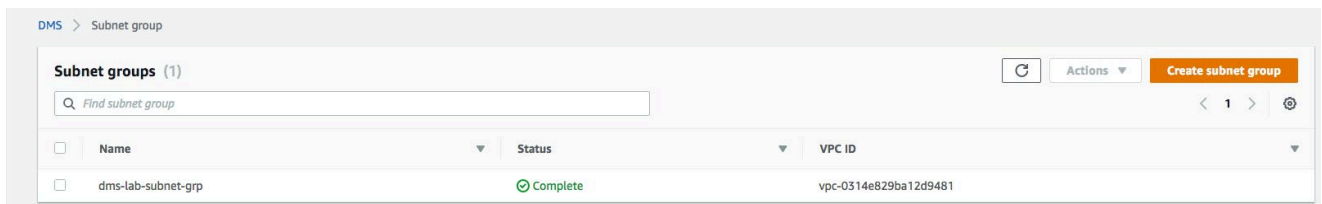
3. Click **Create subnet group**.
4. In the Identifier box, type a descriptive name that you will easily recognize (e.g., dms-lab-subnet-grp).
5. In the Description box, type an easily recognizable description (e.g., Replication instance for production data system).
6. For VPC, select the name of the 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.

Database Migration Services Lab



8. Click Create subnet group

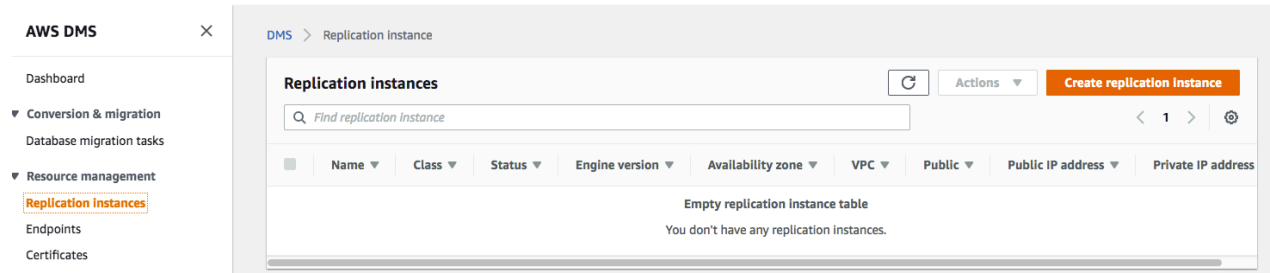
9. On the DMS console, the subnet group status displays Complete.



Create the Replication Instance

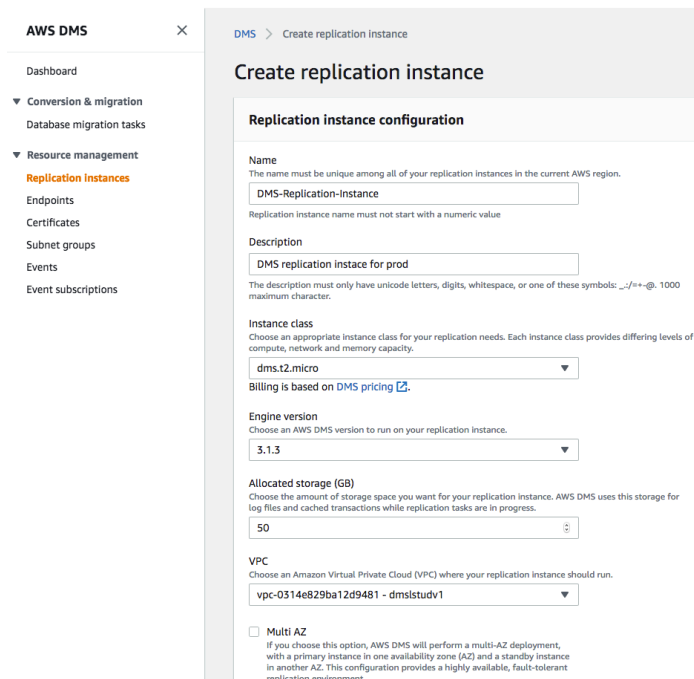
1. On the DMS console, select **Replication instances**.
2. Click **Create replication instance**.

Database Migration Services Lab



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 **dmslstudv1** 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.)



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

Database Migration Services Lab

AWS DMS ×

Dashboard

▼ Conversion & migration
Database migration tasks

▼ Resource management
Replication Instances
Endpoints
Certificates
Subnet groups
Events
Event subscriptions

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.

▼ **Advanced security and network configuration**

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

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.
No Preference

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 ×

KMS master key [Info](#)
(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 ×

Dashboard

▼ Conversion & migration
Database migration tasks

▼ Resource management
Replication Instances
Endpoints

DMS > Replication Instance

Replication instances (1) [Refresh](#) [Actions](#) [Create replication instance](#)

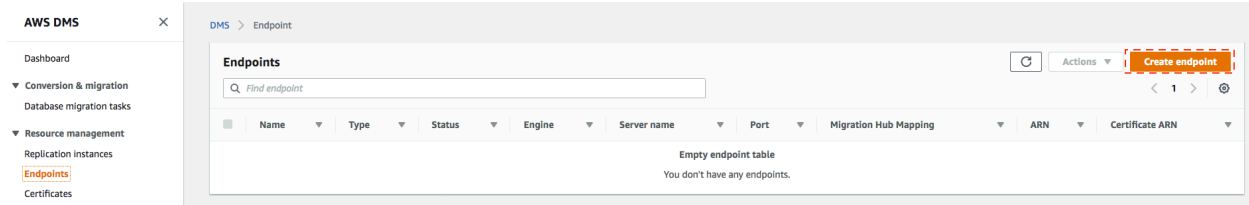
Find replication instance

<input type="checkbox"/>	Name	Class	Status	Engine version	Availability zone	VPC	Public	Public IP address	Private IP address
<input type="checkbox"/>	dms-replication-instance	dms.t2.micro	Available	3.1.3	us-east-1a	vpc-0314e829ba12d9481	Yes	3.213.132.171	10.0.0.7

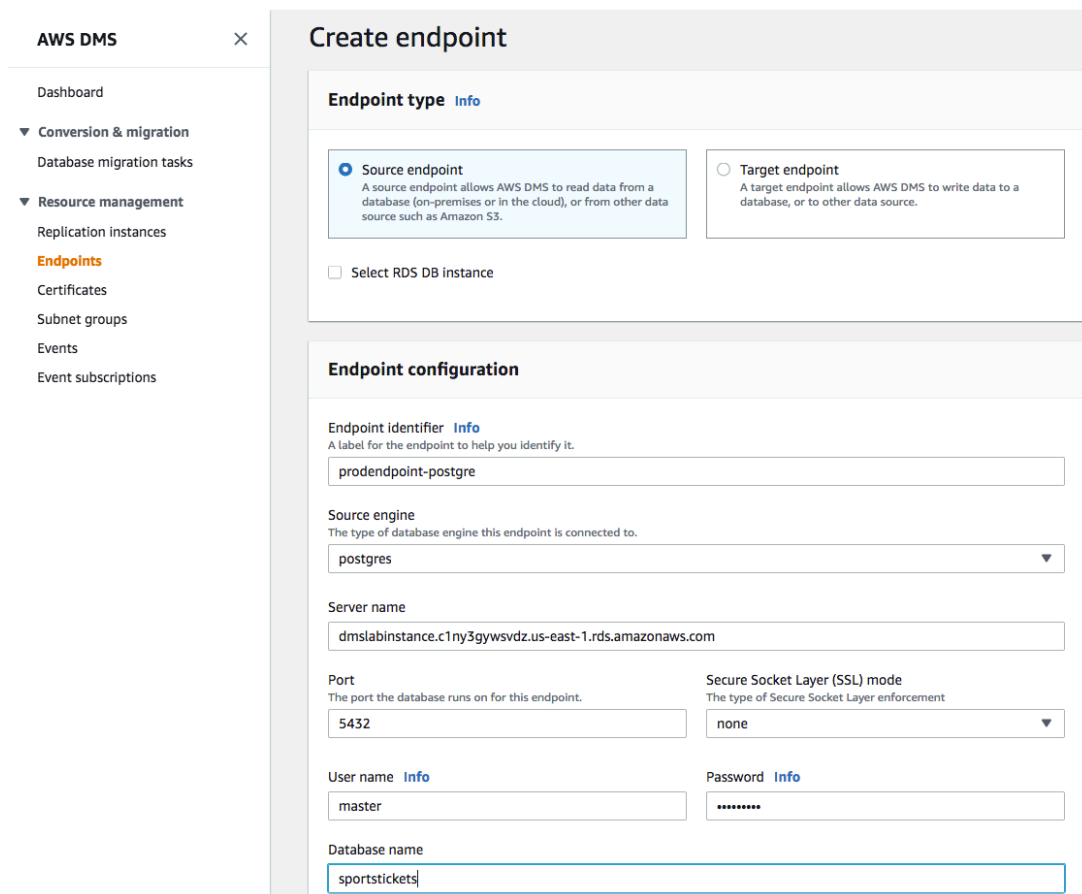
Create the Source Endpoint

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

Database Migration Services Lab



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**.



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

Database Migration Services Lab

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

Database name: sportstickets

Endpoint-specific settings

KMS master key

Test endpoint connection (optional)

Test your endpoint connection by selecting a replication instance within your desired VPC. After clicking "Run test", an endpoint will be created with the details provided and attempt to connect to the instance. If the connection fails, you can edit and test it again. Endpoints that aren't saved will be deleted.

VPC: vpc-G314e829ba12d9481 - dmslstudv1

Replication instance: dms-replication-instance

Run test

After clicking "Run test", an endpoint will be created with the details provided and attempt to connect to the instance. If the connection fails, you can edit and test it again. Endpoints that aren't saved will be deleted.

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**.

Endpoints (1)

Name	Type	Status	Engine	Server name	Port	Migration Hub Mapping	ARN
prodendpoint-postgre	Source	Active	PostgreSQL	dmslabinstance.c1ny3gywsvdz.us-east-1.rds.amazonaws.com	5432		amaws:dmsus-east-1:341259728059:endp...

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.

Database Migration Services Lab

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.

Database Migration Services Lab

Search IAM

- Dashboard
- Groups
- Users
- Roles**
- Policies
- Identity providers
- Account settings
- Credential report
- Encryption keys

• IAM Roles Documentation
• Tutorial: Setting Up Cross Account Access
• Common Scenarios for Roles

Create role Delete role

Q dmslab

Role name	Description	Trusted entities
<input type="checkbox"/> dmslab-student-DMSLabRo...		AWS service: dms

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

Roles > dmslab-student-DMSLabRoleS3-1SR7IR2GC5VD5

Summary Delete role

Role ARN arn:aws:iam::341259728059:role/dmslab-student-DMSLabRoleS3-1SR7IR2GC5VD5

Role description Edit

Instance Profile ARNs

Path /

Creation time 2018-09-29 21:51 PDT

Maximum CLI/API session duration 1 hour Edit

Permissions Trust relationships Access Advisor Revoke sessions

Permissions policies (1 policy applied)

Attach policies Add inline policy

Policy name	Policy type
DMSLabS3Policy	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::'

Attach policies Add inline policy

Policy name	Policy type
DMSLabS3Policy	Inline policy

Policy summary {} JSON Edit policy Simulate policy

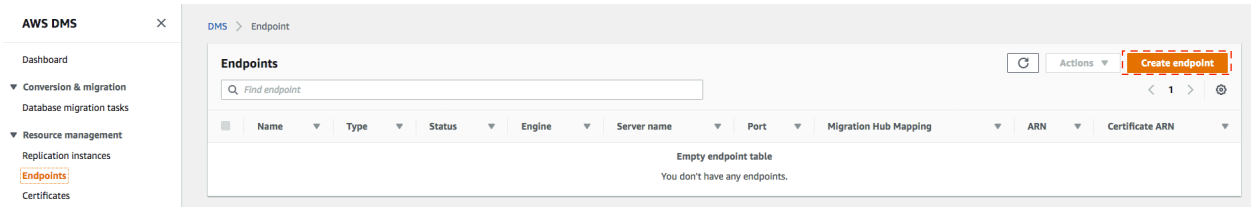
```
11     "arn:aws:s3:::dmslabstudent-dmslabs3bucket-1eegnc2tj056l/*"  
12   ],  
13   "Effect": "Allow"  
14 },  
15 {  
16   "Action": [  
17     "s3:ListBucket"  
18   ],  
19   "Resource": [  
20     "arn:aws:s3:::dmslabstudent-dmslabs3bucket-1eegnc2tj056l"  
21   ],  
22   "Effect": "Allow"  
23 }  
24 ]  
25 }
```

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**.



2. Click **Create endpoint**.

3. For Endpoint type, select **Target**.

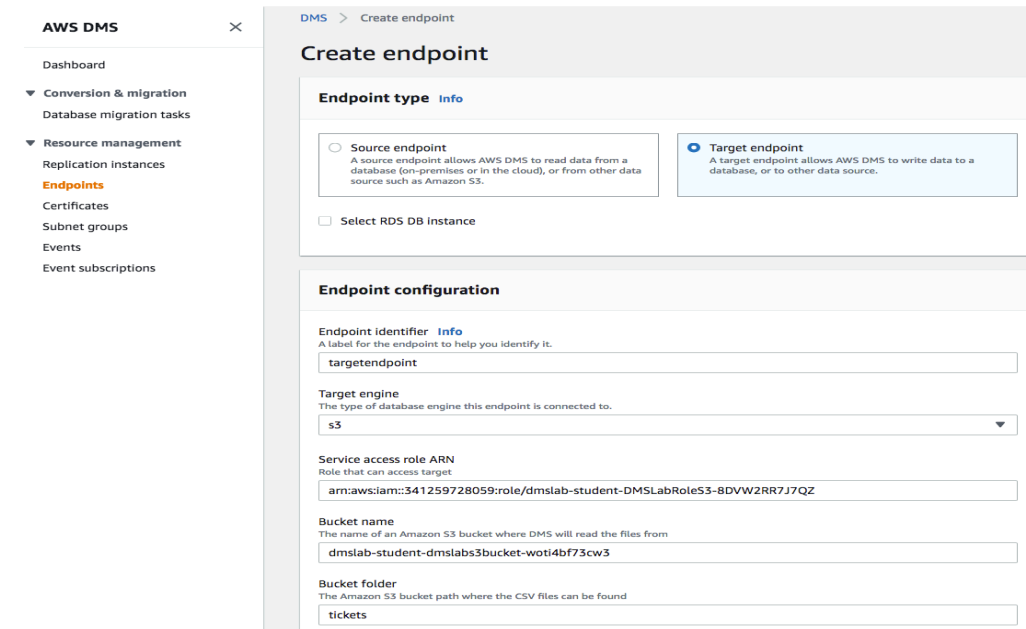
4. For Endpoint identifier, type an easily recognized name.

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**.



9. Click **Endpoint-specific settings** to expand the section.

10. In the **Extra connection attributes** box, type **addColumnNames=true**. This attribute includes the column names in the files in the S3 bucket.

Database Migration Services Lab

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

▼ Conversion & migration

Database migration tasks

▼ Resource management

Replication Instances

Endpoints

Certificates

Subnet groups

Events

Event subscriptions

▼ **Endpoint-specific settings**

Extra connection attributes
Type any additional connection parameters here. See the documentation for more information.

addColumnName=true

▼ **Test endpoint connection (optional)**

Test your endpoint connection by selecting a replication instance within your desired VPC. After clicking "Run test", an endpoint will be created with the details provided and attempt to connect to the instance. If the connection fails, you can edit and test it again. Endpoints that aren't saved will be deleted.

VPC
vpc-0314e829ba12d9481 - dmslstudv1

Replication instance
A replication instance performs the database migration
dms-replication-instance

Run test

After clicking "Run test", an endpoint will be created with the details provided and attempt to connect to the instance. If the connection fails, you can edit and test it again. Endpoints that aren't saved will be deleted.

Endpoint Identifier	Replication Instance	Status	Message
targetendpoint	dms-replication-instance	successful	

Cancel **Create endpoint**

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

AWS DMS

DMS > Endpoint

Endpoints (2)

Find endpoint

Name	Type	Status	Engine	Server name	Port	Migration Hub Mapping	ARN
prodendpoint-postgre	Source	Active	PostgreSQL	dmslabinstance.c1ny3gywsvdz.us-east-1.rds.amazonaws.com	5432		arn:aws:dms:us-east-1:341259728059:endp...
targetendpoint	Target	Active	Amazon S3	-	-		arn:aws:dms:us-east-1:341259728059:endp...

Create a task to perform the initial full copy

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

AWS DMS

DMS > Database migration tasks

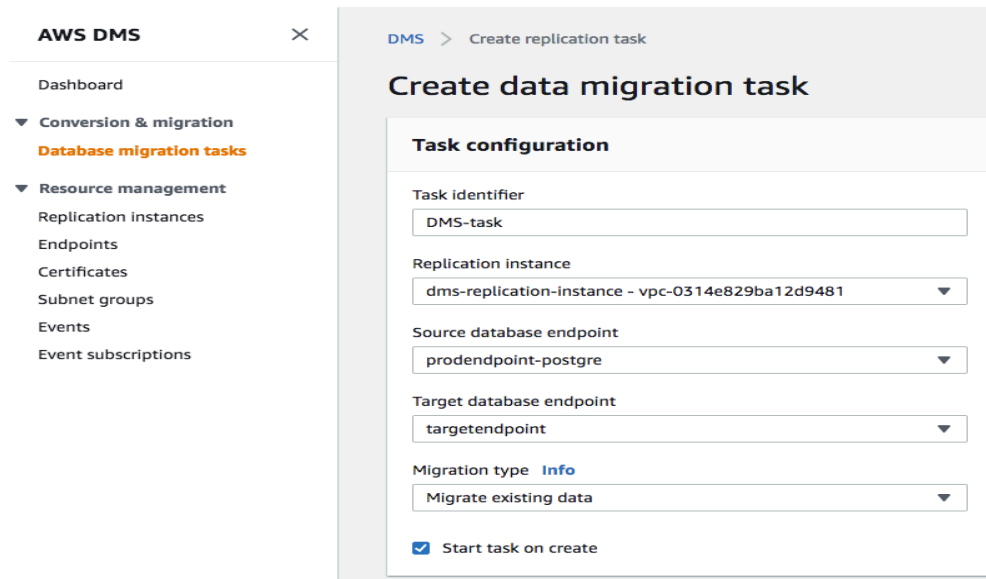
Database migration tasks

Find task

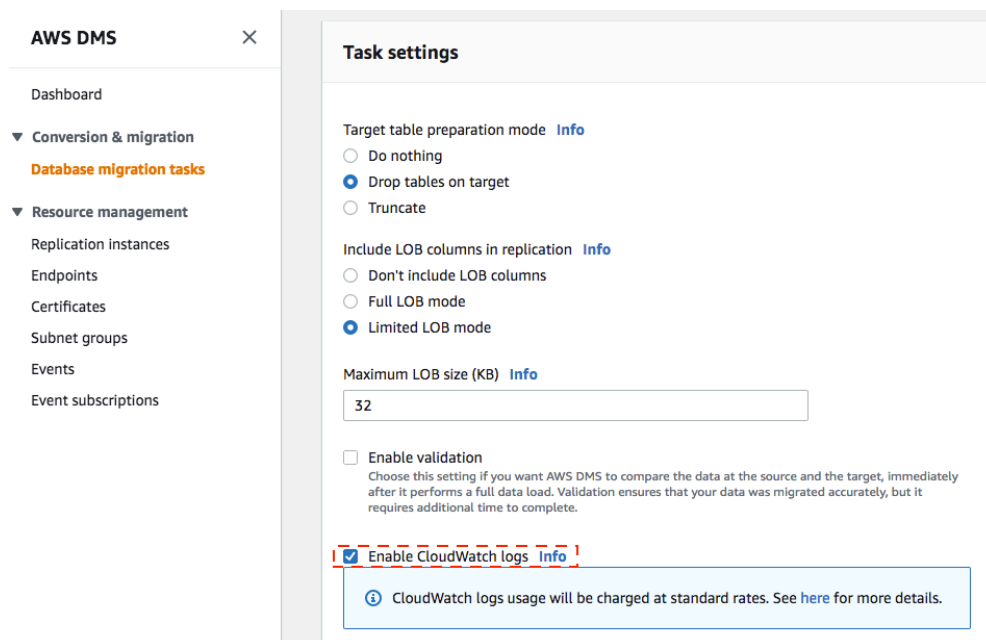
Name	Status	Source	Target	Type	Progress	Elapsed time	Tables loaded	Tables loading	Tables queued	Tables errored
Empty replication task table You don't have any replication tasks.										

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.



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



Database Migration Services Lab

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

Editing mode

Guided UI
Set up your table mapping rules using a step-by-step guided interface.

JSON editor [Learn more](#)
Enter your table mapping rules directly, in JSON format.

Specify at least one selection rule with an include action. After you do this, you can add one or more transformation rules.

▼ **Selection rules**

Choose the schema and/or tables you want to include with, or exclude from, your migration task. [Info](#) **Add new selection rule**

▼ where schema name is like 'dms_sample' and table name is like '%', include

Schema
dms_sample

Table name
Use the % character as a wildcard
%

Action
Choose "Include" to migrate your selected objects, or "Exclude" to ignore them during the migration.
Include

Source filters [Info](#) Add column filter

► Transformation rules

► **Advanced task settings**

Cancel **Create task**

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 > Database migration tasks

Database migration tasks (1)

Find task

	Name	Status	Source	Target	Type	Progress	Elapsed time	Tables loaded	Tables loading	Tables queued	Tables errored
<input type="checkbox"/>	dms-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:

Database Migration Services Lab

The screenshot shows the AWS DMS console interface for a specific task named 'dms-task'. The 'Summary' section indicates the task is 'Load complete'. The 'Overview details' section provides basic configuration information, including the task ARN, type (Full load), source (prodendpoint-postgre), and target (targetendpoint). It also shows the task was created on 5/29/2019 at 10:55:15 AM GMT-0700 and started on the same date at 10:55:51 AM GMT-0700. A 'View logs' link is available.

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

The screenshot displays the 'Table statistics (16)' section of the AWS DMS console. It features a search bar for finding schemas and a table with 16 columns: Schema name, Table, Load state, Inserts, Deletes, Updates, DDLs, Full load rows, Total, Validation state, and Validation pending. The table lists 16 tables, all with a 'Load state' of 'Table completed' and a 'Validation state' of 'Not enabled'. The 'Total' column shows the number of rows loaded for each table, ranging from 2 to 15,212,460.

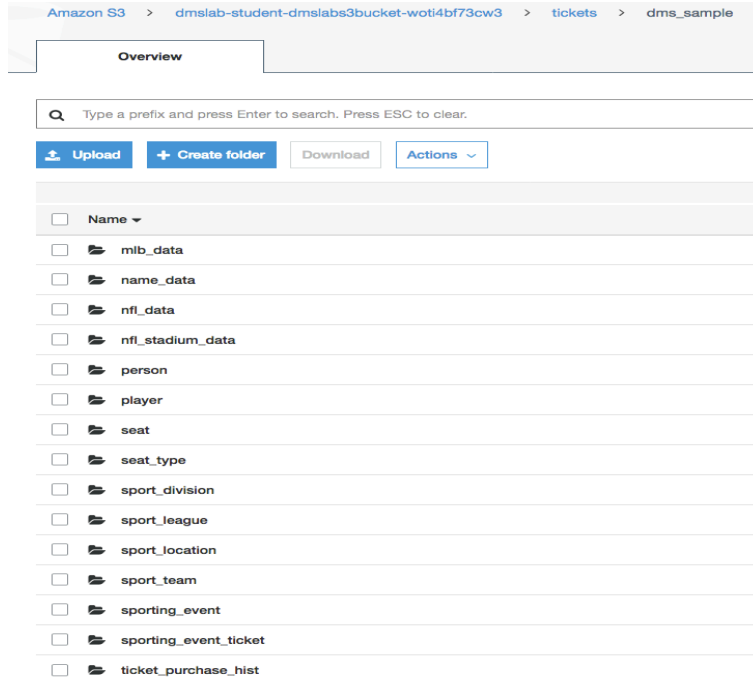
Schema name	Table	Load state	Inserts	Deletes	Updates	DDLs	Full load rows	Total	Validation state	Validation pending
dms_sample	seat_type	Table completed	0	0	0	0	6	6	Not enabled	0
dms_sample	seat	Table completed	0	0	0	0	603,631	603,631	Not enabled	0
dms_sample	mtb_data	Table completed	0	0	0	0	2,230	2,230	Not enabled	0
dms_sample	player	Table completed	0	0	0	0	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	0	0	0	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	0	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	0	0	0	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	0	0	0	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	0	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	0	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 :

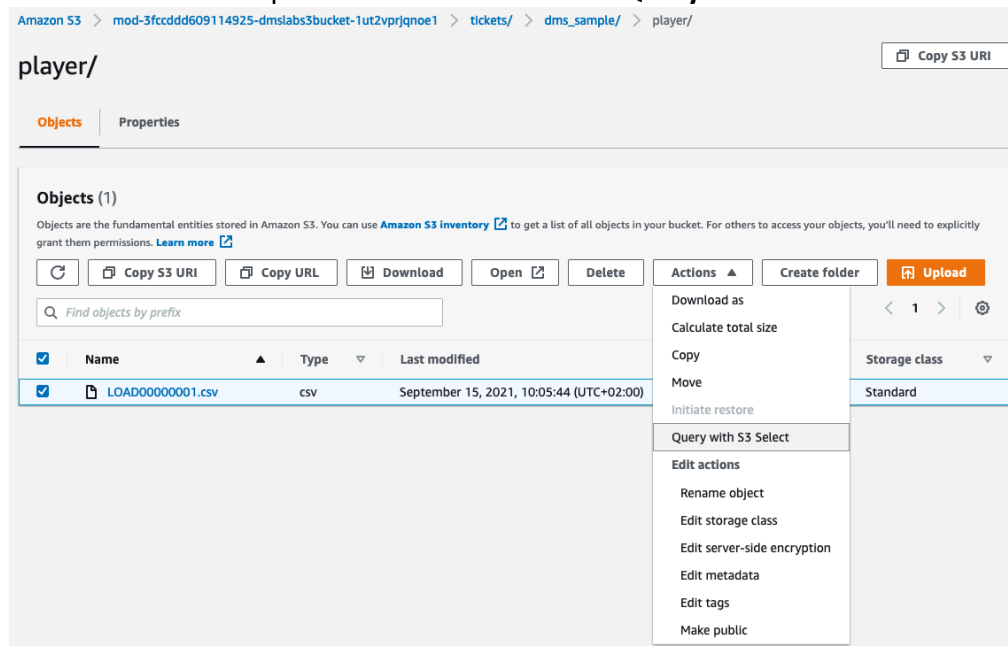
BucketName/bucket_folder_name/schema_name/table_name/objects/

Database Migration Services Lab



16. Navigate to one of the files and review it using [S3 Select](#):

- Navigate in to the directory named ****player**** and select the check box next to the file name.
- Click the **Actions** dropdown button and choose **Query with S3 Select**.



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

Database Migration Services Lab

Query with S3 Select [Info](#)

Use Amazon S3 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 see [Amazon S3 pricing](#)

Input settings

Path
s3://mod-3fccddd609114925-dmslabs3bucket-1ut2vrjqnoe1/tickets/dms_sample/player/LOAD00000001.csv

Size
393.3 KB (402738.0 B)

Format
 CSV
 JSON
 Apache Parquet

CSV delimiter
 Comma
 Tab
 Custom

Exclude the first line of CSV data
Enable this setting if CSV contains a header row.

Compression
 None
 GZIP
 BZIP2

Output settings

Format
 CSV
 JSON

CSV delimiter
 Comma
 Tab
 Custom

SQL query

Amazon S3 Select supports only the SELECT SQL command. Using the S3 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 AWS CLI, AWS SDK, or Amazon S3 REST API. For more complex SQL queries, use [Amazon Athena](#)

```
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.

Status
✔ Successfully returned 5 records in 208 ms
Bytes returned: 352 B

Raw | Formatted

```
id,sport_team_id,last_name,first_name,full_name
+1.0000000000000000e+00,+1.3100000000000000e+02,Adam Loewen,Adam , Loewen
+1.1000000000000000e+01,+1.3100000000000000e+02,A.J. Pollock,A.J. , Pollock
+2.1000000000000000e+01,+1.3100000000000000e+02,Alex Sanabia,Alex , Sanabia
+3.1000000000000000e+01,+1.3100000000000000e+02,Andrew Chafin,Andrew , Chafin
```

You will notice that the file contains the column headers in the first row as requested by the **“addColumnNames=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.

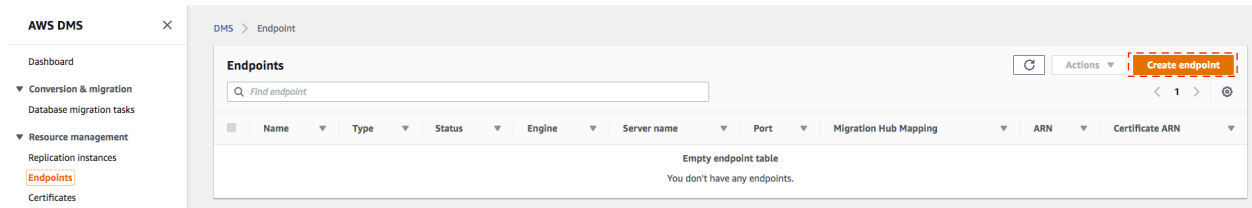
Database Migration Services Lab

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**.



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

Database Migration Services Lab

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

▼ Endpoint-specific settings

Extra connection attributes
Type any additional connection parameters here. See the documentation for more information.

▼ Test endpoint connection (optional)

Test your endpoint connection by selecting a replication instance within your desired VPC. After clicking "Run test", an endpoint will be created with the details provided and attempt to connect to the instance. If the connection fails, you can edit and test it again. Endpoints that aren't saved will be deleted.

VPC

Replication instance
A replication instance performs the database migration

Run test

After clicking "Run test", an endpoint will be created with the details provided and attempt to connect to the instance. If the connection fails, you can edit and test it again. Endpoints that aren't saved will be deleted.

Endpoint identifier	Replication instance	Status	Message
cdcendpoint	dms-replication-instance	successful	

Cancel **Create endpoint**

- Click **Create endpoint**.

- When available, the endpoint status changes to active.

DMS > Endpoint

Endpoints (3) Refresh Actions Create endpoint

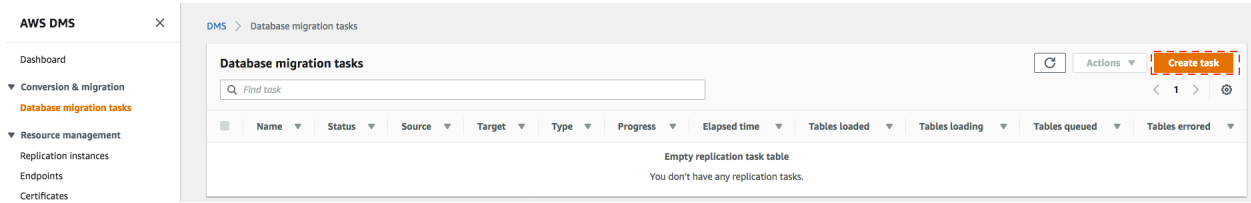
< 1 > ⚙

Name	Type	Status	Engine	Server name	Port	Migration Hub Mapping	ARN
<input checked="" type="checkbox"/> cdcendpoint	Target	Active	Amazon S3	-	-	-	arn:aws:dms:us-east-1:341259728059:endp
<input type="checkbox"/> prodendpoint-postgre	Source	Active	PostgreSQL	dmslabinstance.c1ny3gywsvdz.us-east-1.rds.amazonaws.com	5432	-	arn:aws:dms:us-east-1:341259728059:endp
<input type="checkbox"/> targetendpoint	Target	Active	Amazon S3	-	-	-	arn:aws:dms:us-east-1:341259728059:endp

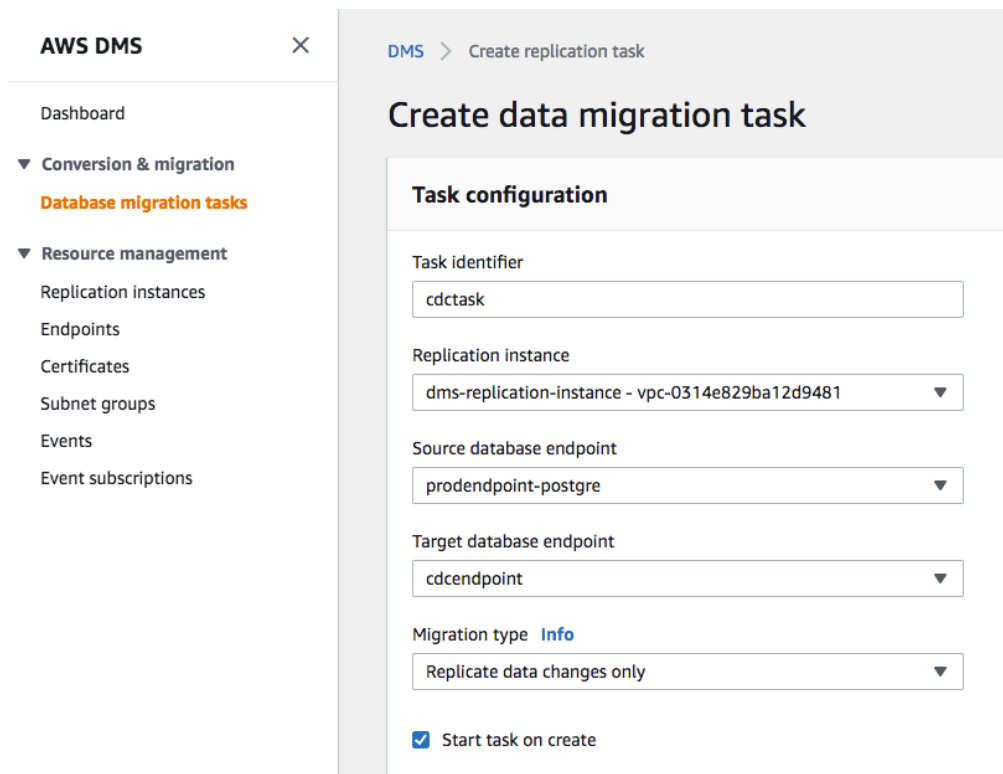
Database Migration Services Lab

(Optional) Create a task to perform the ongoing replication

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



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.



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

Database Migration Services Lab

The screenshot shows the AWS DMS console interface. On the left is a navigation menu with 'AWS DMS' at the top, followed by 'Dashboard', 'Conversion & migration' (with 'Database migration tasks' highlighted), and 'Resource management'. The main area is titled 'Task settings' and contains several configuration options: 'Target table preparation mode' with radio buttons for 'Do nothing', 'Drop tables on target' (selected), and 'Truncate'; 'Include LOB columns in replication' with radio buttons for 'Don't include LOB columns', 'Full LOB mode', and 'Limited LOB mode' (selected); 'Maximum LOB size (KB)' with an input field containing '32'; and 'Enable validation' (unchecked). A red dashed box highlights the 'Enable CloudWatch logs' checkbox, which is checked. Below it, a blue information box states: 'CloudWatch logs usage will be charged at standard rates. See here for more details.'

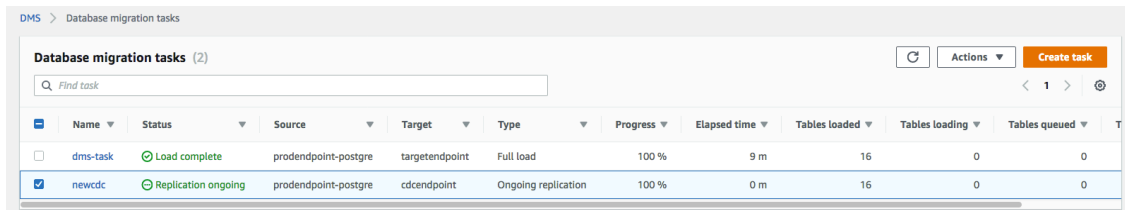
10. Go to **Table Mappings**.
11. Click on **Add new selection rule**
12. For **Schema name**, select **dms_sample** from drop down. Keep the settings for the remaining fields

The screenshot shows the 'Editing mode' for a table mapping rule. It offers two options: 'Guided UI' (selected) and 'JSON editor'. Below, it instructs to 'Specify at least one selection rule with an include action'. Under 'Selection rules', there is a red dashed box around the 'Add new selection rule' button. The rule configuration shows: 'where schema name is like 'dms_sample' and table name is like %, include'. The 'Schema' dropdown is set to 'dms_sample' (highlighted with a red dashed box), and the 'Table name' is set to '%'. The 'Action' is set to 'Include'. At the bottom right, there is a red dashed box around the 'Create task' button.

Database Migration Services Lab

13. Click **Create task**.

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



The screenshot shows the 'Database migration tasks' interface. At the top, there is a search bar labeled 'Find task' and a 'Create task' button. Below the search bar is a table with columns: Name, Status, Source, Target, Type, Progress, Elapsed time, Tables loaded, Tables loading, and Tables queued. Two tasks are listed: 'dms-task' with status 'Load complete' and 'newcdc' with status 'Replication ongoing'.

Name	Status	Source	Target	Type	Progress	Elapsed time	Tables loaded	Tables loading	Tables queued
dms-task	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 postgres database and then picked up by DMS CDC migration task.

16. Select your CDC task and explore the summary:

Database Migration Services Lab

newcdc

Summary

Status 🟢 Replication ongoing	Type Ongoing replication	Source prodendpoint-postgre	Target cdcendpoint
---------------------------------	-----------------------------	--------------------------------	-----------------------

Overview details

Basic configuration

Task ARN
arn:aws:dms:us-east-1:341259728059:task:Y5L3X5DAFT6B7F5EQYFGJ45TUQ

Type
Ongoing replication

Source
prodendpoint-postgre

Last failure message
-

Started
5/29/2019, 4:01:28 PM GMT-0700

Change data capture (CDC)

Change data capture (CDC) start position
-

Change data capture (CDC) recovery checkpoint
-

Status
🟢 Replication ongoing

Replication instance
dms-replication-instance

Target
cdcendpoint

Created
5/29/2019, 3:15:19 PM GMT-0700

Migration task logs [Info](#)

[View logs](#)

Change data capture (CDC) stop position
-

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

Table statistics (16)												Validate again	Reload table data
<input type="text" value="Find schema"/>													
<input type="checkbox"/>	Schema name	Table	Load state	Inserts	Deletes	Updates	DDLs	Full load rows	Total	Validation state	Validation pending		
<input type="checkbox"/>	dms_sample	seat_type	Table completed	0	0	0	0	0	0	Not enabled	0		
<input type="checkbox"/>	dms_sample	seat	Table completed	0	0	0	0	0	0	Not enabled	0		
<input type="checkbox"/>	dms_sample	mlb_data	Table completed	0	0	0	0	0	0	Not enabled	0		
<input type="checkbox"/>	dms_sample	player	Table completed	0	0	0	0	0	0	Not enabled	0		
<input type="checkbox"/>	dms_sample	ticket_purchase_hist	Table completed	680,218	0	0	0	0	680,218	Not enabled	0		
<input type="checkbox"/>	dms_sample	person	Table completed	0	0	0	0	0	0	Not enabled	0		
<input type="checkbox"/>	dms_sample	name_data	Table completed	0	0	0	0	0	0	Not enabled	0		
<input type="checkbox"/>	dms_sample	sport_team	Table completed	0	0	0	0	0	0	Not enabled	0		
<input type="checkbox"/>	dms_sample	sport_league	Table completed	0	0	0	0	0	0	Not enabled	0		
<input type="checkbox"/>	dms_sample	sporting_event	Table completed	0	0	0	0	0	0	Not enabled	0		
<input type="checkbox"/>	dms_sample	sporting_event_ticket	Table completed	0	0	680,218	0	0	680,218	Not enabled	0		
<input type="checkbox"/>	dms_sample	sport_division	Table completed	0	0	0	0	0	0	Not enabled	0		
<input type="checkbox"/>	dms_sample	sport_location	Table completed	0	0	0	0	0	0	Not enabled	0		
<input type="checkbox"/>	dms_sample	sport_type	Table completed	0	0	0	0	0	0	Not enabled	0		
<input type="checkbox"/>	dms_sample	nfl_stadium_data	Table completed	0	0	0	0	0	0	Not enabled	0		
<input type="checkbox"/>	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.

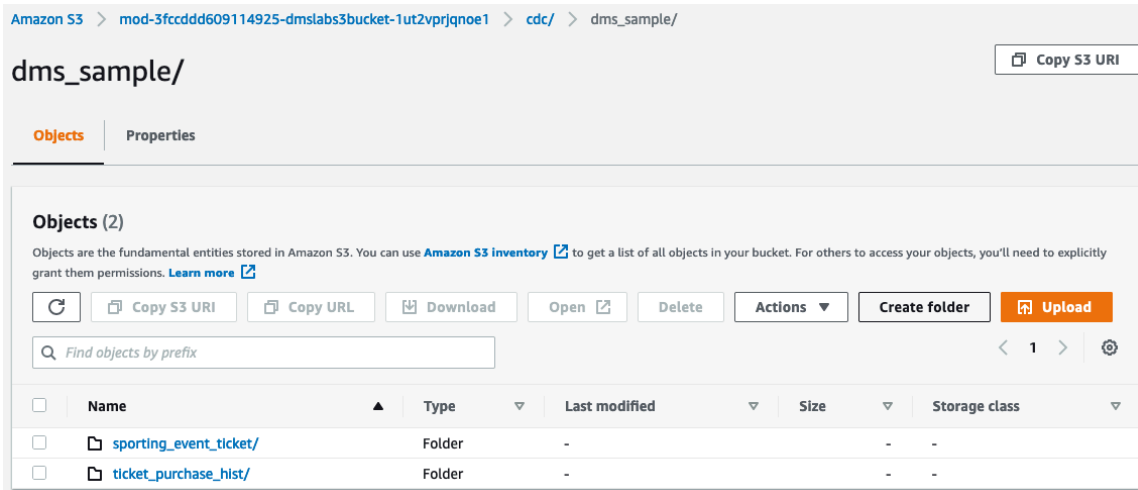
Database Migration Services Lab

Your S3 bucket name will look like below :

BucketName/bucket_folder_name/schema_name/table_name/objects/

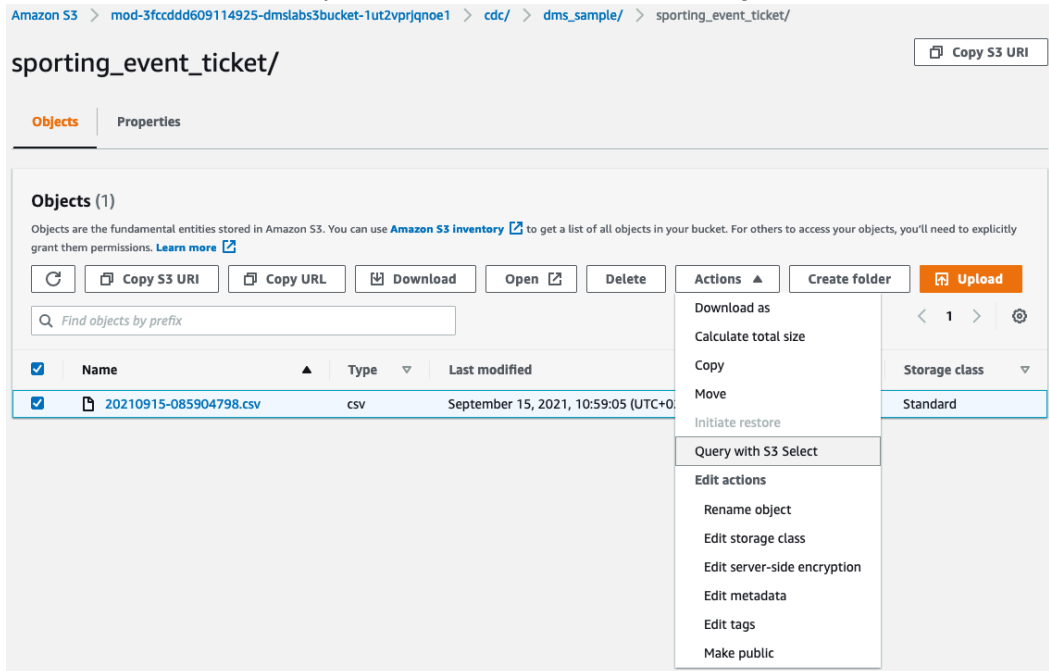
In our lab example this becomes:

“/dmslab-student-dmslabs3bucket-woti4bf73cw3/cdc/dms_sample” with a separate path for each table_name)



17. Navigate to one of the files and review it using [S3 Select](#):

- Navigate in to the directory named ****player**** and select the check box next to the file name.
- Click the **Actions** dropdown button and choose **Query with S3 Select**.



Database Migration Services Lab

- 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](#)

Use Amazon S3 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 [see Amazon S3 pricing](#)

Input settings

Path
s3://mod-3fccddd609114925-dmsslabs3bucket-1ut2vprjqnoe1/cdc/dms_sample/sporting_event_ticket/20210915-085904798.csv

Size
5.8 MB (6092630.0 B)

Format
 CSV
 JSON
 Apache Parquet

CSV delimiter
 Comma
 Tab
 Custom

Exclude the first line of CSV data
Enable this setting if CSV contains a header row.

Compression
 None
 GZIP
 BZIP2

Output settings

Format
 CSV
 JSON

CSV delimiter
 Comma
 Tab
 Custom

SQL query

Amazon S3 Select supports only the SELECT SQL command. Using the S3 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 AWS CLI, AWS SDK, or Amazon S3 REST API. For more complex SQL queries, use [Amazon Athena](#)

```
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.

Status
✔ Successfully 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,ticket_price
U,+1.2655711000000000e+07,+1.4410000000000000e+03,+9.0000000000000000e+00,2,10,A,2,+4.9121750000000000e+06,43.23
U,+1.2655721000000000e+07,+1.4410000000000000e+03,+9.0000000000000000e+00,2,10,A,1,+4.9121750000000000e+06,43.23
U,+1.2667891000000000e+07,+1.4410000000000000e+03,+9.0000000000000000e+00,2,10,A,2,+4.9121750000000000e+06,86.46
U,+1.2652081000000000e+07,+1.4410000000000000e+03,+9.0000000000000000e+00,2,10,A,1,+4.9121750000000000e+06,43.23
```

Database Migration Services Lab

You will notice that the file contains the column headers in the first row as requested by the “addColumnNames=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.