



Amazon Web Services Data Engineering Immersion Day

Database Migration Services Lab
Jun 2019

Table of Contents

<i>Introduction.....</i>	<i>2</i>
<i>Create the Subnet Group.....</i>	<i>3</i>
<i>Create the Replication Instance.....</i>	<i>4</i>
<i>Create the Source Endpoint.....</i>	<i>6</i>
<i>IAM Policy for DMS->S3 Access.....</i>	<i>8</i>
<i>Create the Target Endpoint.....</i>	<i>11</i>
<i>Create a task to perform the initial full copy.....</i>	<i>12</i>
<i>(Optional) Create the CDC endpoint to replicate ongoing changes.....</i>	<i>16</i>
<i>(Optional) Create a task to perform the ongoing replication.....</i>	<i>18</i>

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.

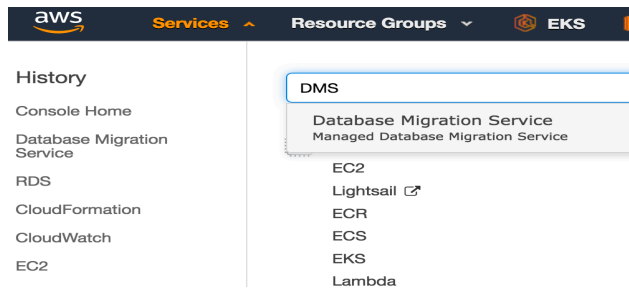
Optionally, you can add ongoing replication of data changes on the source: ***(Only one of the DMS replication instances will enable this feature.)***

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

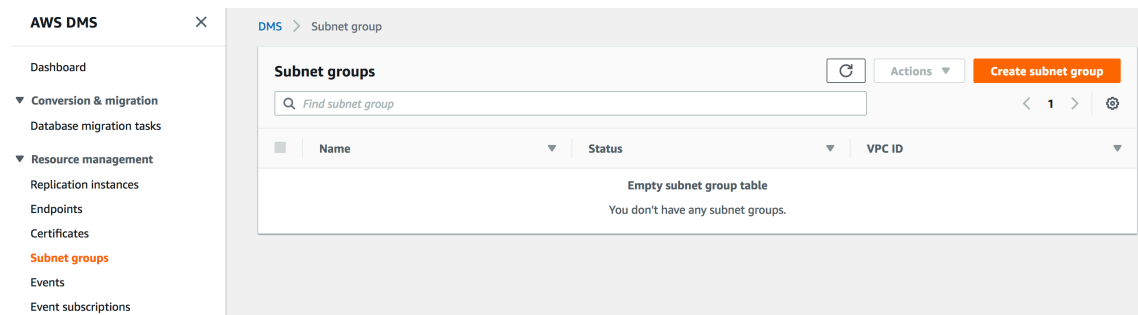
Your instructor has already created and populated the RDS Postgres database that you will use as your source endpoint in this lab.

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

The screenshot shows the AWS DMS console interface. On the left is a navigation sidebar with 'Subnet groups' highlighted. The main panel is titled 'Create replication subnet group'. It contains two sections: 'Subnet group configuration' and 'Add subnets'. In the configuration section, the 'Name' is 'dms-lab-subnet-grp', the 'Description' is 'Replication instance for production data system', and the 'VPC' is 'vpc-0314e829ba12d9481...'. The 'Add subnets' section shows three subnets being added to the group: 'subnet-006522a5bbf92c0c5', 'subnet-097495655d22962d2', and 'subnet-05f628a564cf25622'. At the bottom right are 'Cancel' and 'Create subnet group' buttons.

8. Click Create subnet group

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

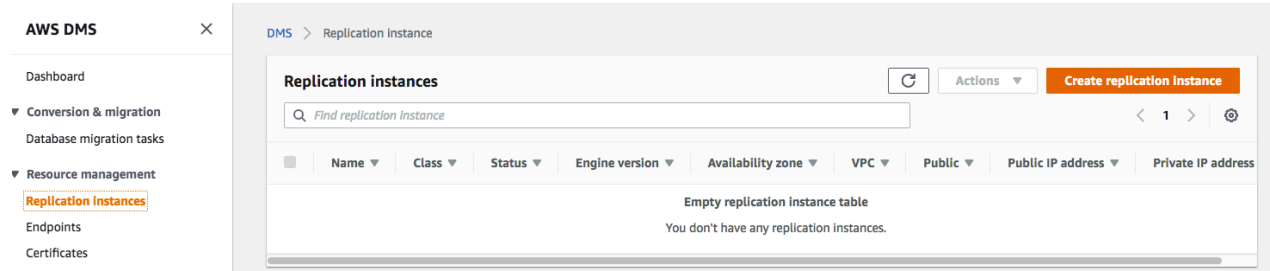
The screenshot shows the 'Subnet groups' list in the AWS DMS console. It features a search bar, a table with columns for Name, Status, and VPC ID, and a 'Create subnet group' button. The table contains one entry: 'dms-lab-subnet-grp' with a status of 'Complete' and VPC ID 'vpc-0314e829ba12d9481'.

Name	Status	VPC ID
dms-lab-subnet-grp	Complete	vpc-0314e829ba12d9481

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.)

The screenshot shows the 'Create replication instance' configuration page in the AWS DMS console. The left navigation menu is the same as in the previous image. The main content area is titled 'DMS > Create replication instance' and 'Create replication instance'. Below this is a 'Replication instance configuration' section with the following fields:

- Name:** A text input field containing 'DMS-Replication-Instance'. A note states: 'The name must be unique among all of your replication instances in the current AWS region. Replication instance name must not start with a numeric value.'
- Description:** A text input field containing 'DMS replication instance for prod'. A note states: 'The description must only have unicode letters, digits, whitespace, or one of these symbols: _/+=@. 1000 maximum character.'
- Instance class:** A dropdown menu showing 'dms.t2.micro'. A note states: 'Choose an appropriate instance class for your replication needs. Each instance class provides differing levels of compute, network and memory capacity. Billing is based on DMS pricing.'
- Engine version:** A dropdown menu showing '3.1.3'. A note states: 'Choose an AWS DMS version to run on your replication instance.'
- Allocated storage (GB):** A text input field containing '50'. A note states: 'Choose the 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.'
- VPC:** A dropdown menu showing 'vpc-0314e829ba12d9481 - dmslstudv1'. A note states: 'Choose an Amazon Virtual Private Cloud (VPC) where your replication instance should run.'
- Multi AZ:** An unchecked checkbox. A note states: '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.

Database Migration Services Lab

AWS DMS X

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 X

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 X

Dashboard

▼ Conversion & migration

Database migration tasks

▼ Resource management

Replication instances

Endpoints

DMS > Replication Instance

Replication instances (1)

Find replication instance

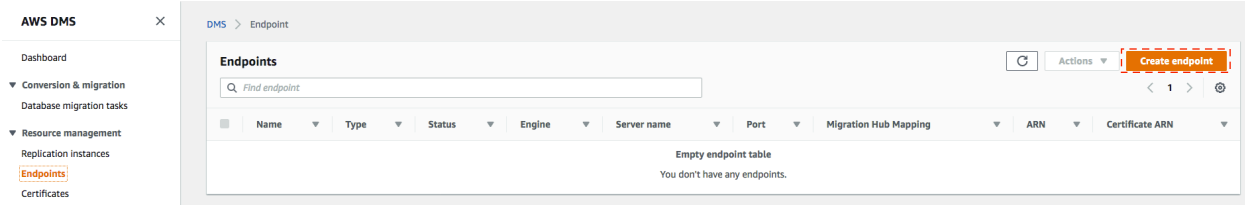
Actions Create replication instance

	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 **master**.
10. For Password, type **master123**.
11. For Database name, type **sportstickets**.

AWS DMS ×

Dashboard

▼ Conversion & migration

Database migration tasks

▼ Resource management

Replication instances

Endpoints

Certificates

Subnet groups

Events

Event subscriptions

Create endpoint

Endpoint type [Info](#)

☒ **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 S3.

☐ **Target endpoint**
A target endpoint allows AWS DMS to write data to a database, or to other data source.

☐ Select RDS DB instance

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.amazonaws.com

Port
The port the database runs on for this endpoint.

5432

Secure Socket Layer (SSL) mode
The type of Secure Socket Layer enforcement

none

User name [Info](#)

master

Password [Info](#)

Database name

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

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-0314e829ba12d9481 - dmslstdv1

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**.

AWS DMS

DMS > Endpoint

Endpoints (1)

Find endpoint

	Name	Type	Status	Engine	Server name	Port	Migration Hub Mapping	ARN
<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...

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.

Database Migration Services Lab

Search IAM

- 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.

Search IAM

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

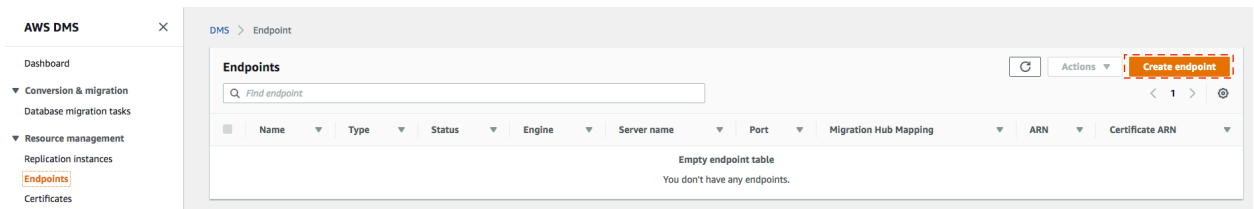
```
11      "arn:aws:s3:::dmslabstudent-dmslabs3bucket-1eegnc2tj056l/*"
12    ],
13    "Effect": "Allow"
14  },
15  {
16    "Action": [
17      "s3:ListBucket"
18    ],
19    "Resource": [
20      "arn:aws:s3:::dmslabstudent-dmslabs3bucket-1eegnc2tj056l/*"
21    ],
22    "Effect": "Allow"
23  }
24 ]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

Create the Target Endpoint

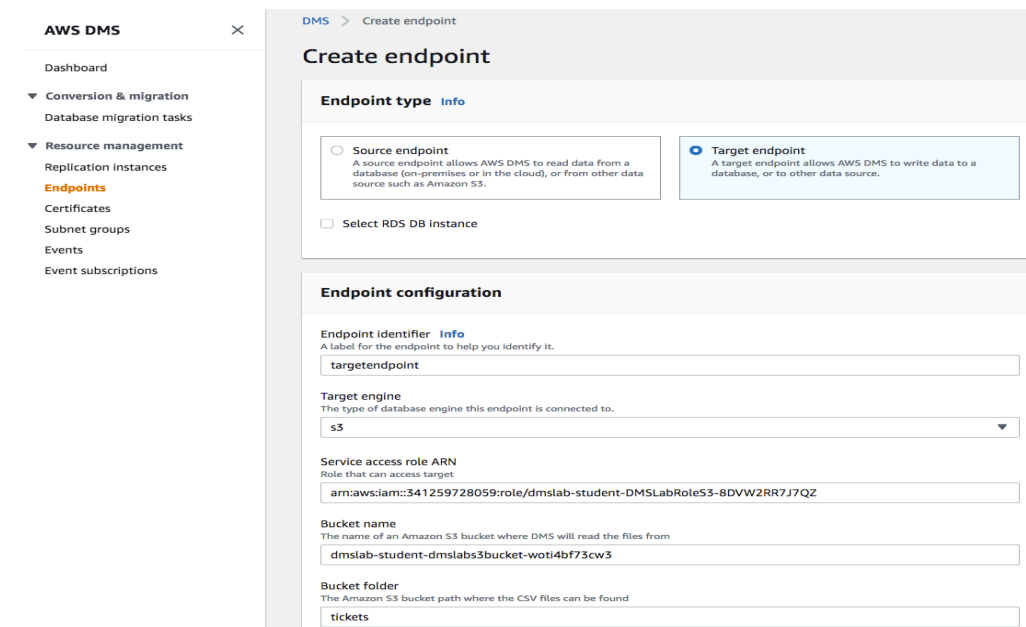
Before start, make sure you have the following information on hand:

- **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.

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
<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...

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.

The screenshot shows the AWS DMS console interface. On the left is a navigation sidebar with the following menu items: Dashboard, Conversion & migration (expanded), Database migration tasks (highlighted), Resource management, Replication instances, Endpoints, Certificates, Subnet groups, Events, and Event subscriptions. The main panel is titled 'DMS > Create replication task' and 'Create data migration task'. Under the 'Task configuration' section, the following fields are visible: Task identifier (DMS-task), Replication instance (dms-replication-instance - vpc-0314e829ba12d9481), Source database endpoint (prodendpoint-postgre), Target database endpoint (targetendpoint), Migration type (Migrate existing data), and a checked checkbox for 'Start task on create'.

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

This screenshot shows the 'Task settings' section of the AWS DMS console. The left sidebar is identical to the previous screenshot. The main panel is titled 'Task settings' and contains the following configuration options: Target table preparation mode (Drop tables on target is selected), Include LOB columns in replication (Limited LOB mode is selected), Maximum LOB size (KB) (32), and a checkbox for 'Enable validation' which is currently unchecked. The 'Enable CloudWatch logs' checkbox is checked and highlighted with a red dashed box. Below this checkbox, a blue information box states: 'CloudWatch logs usage will be charged at standard rates. See here for more details.'

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:

DMS > Database migration tasks > dms-task

dms-task

Summary

Status Load complete	Type Full load	Source prodendpoint-postgre	Target targetendpoint
-------------------------	-------------------	--------------------------------	--------------------------

Overview details

Basic configuration

Task ARN
arn:aws:dms:us-east-1:541259728059:task:MUVRCLBYT45EZESVNFNGAUL4

Type
Full load

Source
prodendpoint-postgre

Last failure message
-

Started
5/29/2019, 10:55:51 AM GMT-0700

Task settings (JSON)

Status
Load complete

Replication instance
dms-replication-instance

Target
targetendpoint

Created
5/29/2019, 10:55:15 AM GMT-0700

Migration task logs [Info](#)

[View logs](#)

Database Migration Services Lab

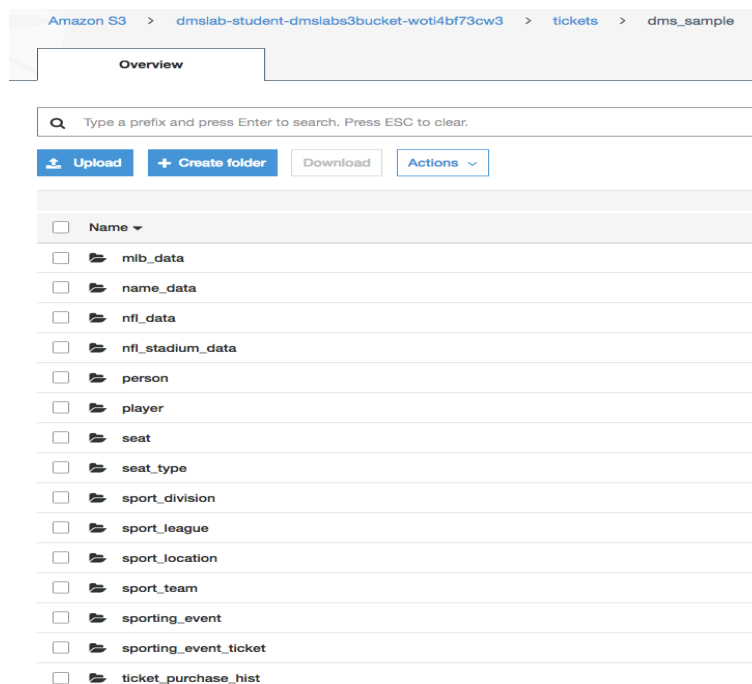
Scroll down and you can observe all table information loaded in S3 from RDS by DMS

Table statistics (16)											
<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	6	6	Not enabled	0
<input type="checkbox"/>	dms_sample	seat	Table completed	0	0	0	0	603,631	603,631	Not enabled	0
<input type="checkbox"/>	dms_sample	mlb_data	Table completed	0	0	0	0	2,230	2,230	Not enabled	0
<input type="checkbox"/>	dms_sample	player	Table completed	0	0	0	0	5,157	5,157	Not enabled	0
<input type="checkbox"/>	dms_sample	ticket_purchase_hist	Table completed	0	0	0	0	6,038,756	6,038,756	Not enabled	0
<input type="checkbox"/>	dms_sample	person	Table completed	0	0	0	0	7,025,584	7,025,584	Not enabled	0
<input type="checkbox"/>	dms_sample	name_data	Table completed	0	0	0	0	5,360	5,360	Not enabled	0
<input type="checkbox"/>	dms_sample	sport_team	Table completed	0	0	0	0	62	62	Not enabled	0
<input type="checkbox"/>	dms_sample	sport_league	Table completed	0	0	0	0	2	2	Not enabled	0
<input type="checkbox"/>	dms_sample	sporting_event	Table completed	0	0	0	0	1,158	1,158	Not enabled	0
<input type="checkbox"/>	dms_sample	sporting_event_ticket	Table completed	0	0	0	0	15,212,460	15,212,460	Not enabled	0
<input type="checkbox"/>	dms_sample	sport_division	Table completed	0	0	0	0	14	14	Not enabled	0
<input type="checkbox"/>	dms_sample	sport_location	Table completed	0	0	0	0	62	62	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	32	32	Not enabled	0
<input type="checkbox"/>	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/



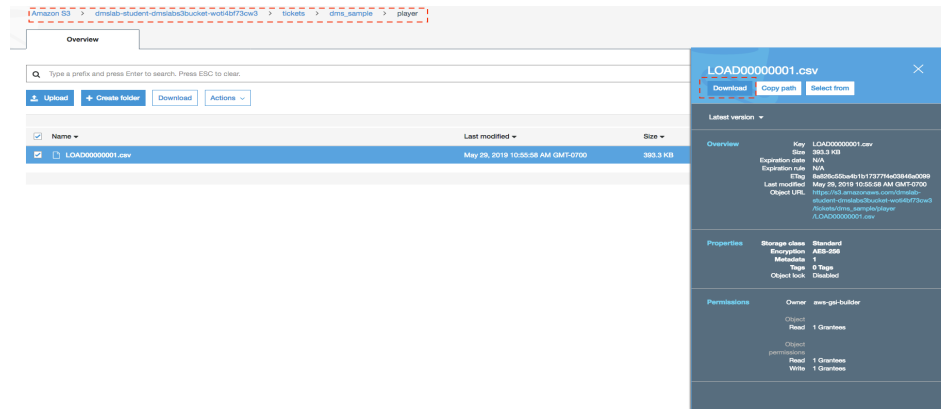
16. Download one of the files:

- Select the check box next to the object name and click Download in the pop-up window.
- Click **Save File**.
- Open the file.

You will notice that the file contains the column headers in the first row

Database Migration Services Lab

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.

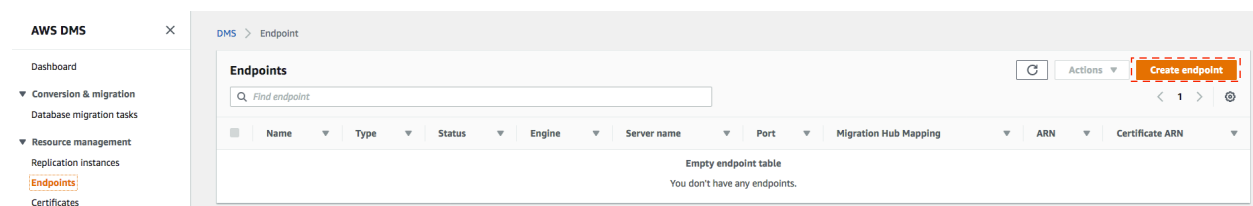
	A	B	C	D	E
1	id	sport_team_id	last_name	first_name	full_name
2	1	131	Adam Loewen	Adam	Loewen
3	11	131	A.J. Pollock	A.J.	Pollock
4	21	131	Alex Sanabia	Alex	Sanabia
5	31	131	Andrew Chafin	Andrew	Chafin
6	41	131	Andy Marte	Andy	Marte
7	51	131	Archie Bradley	Archie	Bradley
8	61	131	Ben Francisco	Ben	Francisco
9	71	131	Braden Shipley	Braden	Shipley
10	81	131	Bradin Hagens	Bradin	Hagens
11	91	131	Brandon Drury	Brandon	Drury
12	101	131	Brett Jackson	Brett	Jackson

Explore the objects in the S3 directory further.

(Optional) Create the CDC endpoint to replicate ongoing changes

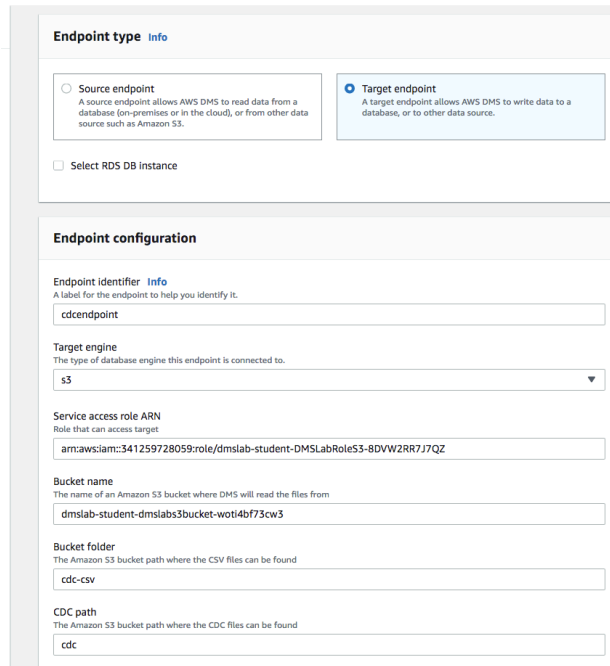
As of now we are enabling only one schema replication for CDC

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



Database Migration Services Lab

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 **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



Endpoint type [Info](#)

☐ 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 S3.

☒ Target endpoint
A target endpoint allows AWS DMS to write data to a database, or to other data source.

☐ Select RDS DB instance

Endpoint configuration

Endpoint identifier [Info](#)
A label for the endpoint to help you identify it.
cdcendpoint

Target engine
The type of database engine this endpoint is connected to.
s3

Service access role ARN
Role that can access target
arn:aws:iam::341259728059:role/dmslab-student-DMSLabRoleS3-8DVW2RR7J7QZ

Bucket name
The name of an Amazon S3 bucket where DMS will read the files from
dmslab-student-dmslabs3bucket-wot4bf73cw3

Bucket folder
The Amazon S3 bucket path where the CSV files can be found
cdc-csv

CDC path
The Amazon S3 bucket path where the CDC files can be found
cdc

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.

Database Migration Services Lab

▼ 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**

13. Click **Create endpoint**.

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

DMS > Endpoint

Endpoints (3) Find endpoint Actions Create endpoint

	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

(Optional) Create a task to perform the ongoing replication

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

AWS DMS ×

Database migration tasks Find task Actions Create 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.											

Database Migration Services Lab

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 ×

Dashboard

▼ Conversion & migration

Database migration tasks

▼ Resource management

Replication instances

Endpoints

Certificates

Subnet groups

Events

Event subscriptions

DMS > Create replication task

Create data migration task

Task configuration

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.

Database Migration Services Lab

AWS DMS ×

Dashboard

▼ Conversion & migration

Database migration tasks

▼ Resource management

Replication instances

Endpoints

Certificates

Subnet groups

Events

Event subscriptions

Task settings

Target table preparation mode [Info](#)

☐ Do nothing

☒ Drop tables on target

☐ Truncate

Include LOB columns in replication [Info](#)

☐ Don't include LOB columns

☐ Full LOB mode

☒ Limited LOB mode

Maximum LOB size (KB) [Info](#)

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.

☒ Enable CloudWatch logs [Info](#)

[i](#) 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

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

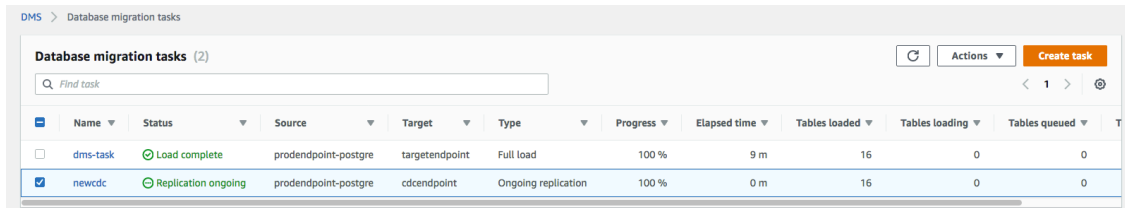
► Advanced task settings

Cancel Create task

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.



Database migration tasks (2)										
Find task										
	Name	Status	Source	Target	Type	Progress	Elapsed time	Tables loaded	Tables loading	Tables queued
<input type="checkbox"/>	dms-task	Load complete	prodendpoint-postgre	targetendpoint	Full load	100 %	9 m	16	0	0
<input checked="" type="checkbox"/>	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:

newcdc

Summary

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

Overview details

Basic configuration

Task ARN
arn:aws:dms:us-east-1:341259728059:task:Y5L3X5DAFT6B7FSEQYFGJ45TUQ

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
-

Database Migration Services Lab

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

Table statistics (16)											
<input type="text" value="Find schema"/>											
<div><div>Validate again</div><div>Reload table data</div></div>											
<div>< 1 > ⚙</div>											
<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

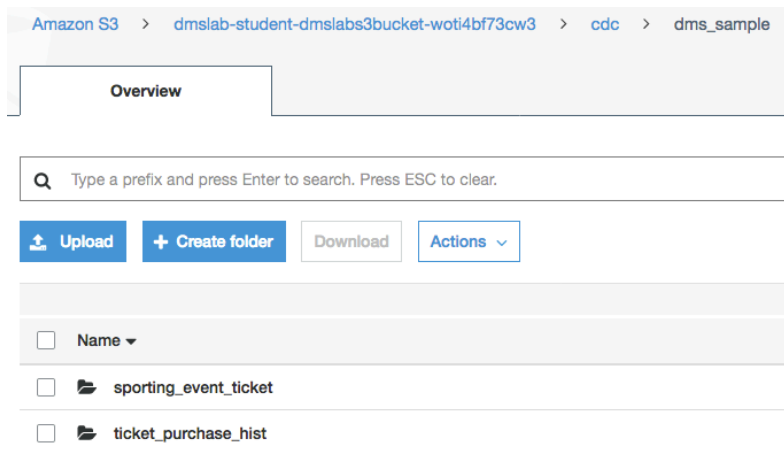
17. Open the S3 console and view the CDC data that was copied by DMS.

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)



Database Migration Services Lab

18. Download one of the files:

- Select the check box next to the object name and click Download in the pop-up window.
- Click **Save File**.
- Open the file.

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.

Amazon S3 > dmslab-student-dmslab3bucket-wot4b73cw3 > cdc > dms_sample > sporting_event_ticket

Overview

20190529-230604667.csv

Download Copy path Select from

Latest version

Overview

Key: 20190529-230604667.csv

Size: 19.3 MB

Expiration date: N/A

Expiration rule: N/A

ETag: ad6b57f453c0e49745035e8d50dfb3bc-4

Last modified: May 29, 2019 4:06:05 PM GMT-0700

Object URL: https://s3.amazonaws.com/dmslab-student-dmslab3bucket-wot4b73cw3/cdc/dms_sample/sporting_event_ticket/20190529-230604667.csv

Name	Last modified	Size
<input checked="" type="checkbox"/> 20190529-230604667.csv	May 29, 2019 4:06:05 PM GMT-0700	19.3 MB
<input type="checkbox"/> 20190529-230729693.csv	May 29, 2019 4:07:30 PM GMT-0700	53.9 MB

Note that file name has date time - 20190529-230604667.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.

	A	B	C	D	E	F	G	H	I	J
	Op	id	sporting_eve	sport_location_id	seat_level	seat_section	seat_row	seat	ticketholder_id	ticket_price
1	U	145192591	3931	4	2	10 A		2	2898028	98
2	U	145192601	3931	4	2	10 A		1	2898028	98
3	U	145192581	3931	4	2	10 A		3	2898028	98
4	U	145192501	3931	4	2	10 B		1	2898028	98
5	U	145187751	3931	4	2	13 B		2	2898028	49
6	U	145187741	3931	4	2	13 B		3	2898028	49
7	U	145187721	3931	4	2	13 C		2	2898028	49
8	U	145187711	3931	4	2	13 C		3	2898028	49
9	U	145187731	3931	4	2	13 C		1	2898028	49
10	U	145187701	3931	4	2	14 A		1	2898028	49
11	U	145187681	3931	4	2	14 A		3	2898028	49
12	U	145187691	3931	4	2	14 A		2	2898028	49
13	U	145187471	3931	4	2	14 B		3	2898028	49
14	U	145187671	3931	4	2	14 B		1	2898028	49
15	U	145187481	3931	4	2	14 B		2	2898028	49
16	U	145187451	3931	4	2	14 C		2	2898028	49
17	U	145187461	3931	4	2	14 C		1	2898028	49
18	U	145190341	3931	4	2	14 C		3	2898028	49
19	U	145183201	3931	4	2	15 A		4	2898028	49
20	U	145179691	3931	4	2	15 A		1	2898028	49
21	U	145179661	3931	4	2	15 A		4	2898028	49
22	U	145179671	3931	4	2	15 A		3	2898028	49
23	U	145179681	3931	4	2	15 A		2	2898028	49
24	U	145190321	3931	4	2	15 A		2	2898028	49

Explore the objects in the S3 directory further.