**January 2018**



**Intelligent Data Lake Workshop**

*Lab 4 – DMS (Optional)*

Table of Contents

[Overview 3](#_Toc507510149)

[Preparing AWS Resources Using CloudFormation 4](#_Toc507510150)

[Create a Key Pair 4](#_Toc507510151)

[Launch Resources using AWS CloudFormation 5](#_Toc507510152)

[Connecting to your EC2 Instance 10](#_Toc507510153)

[Preparing SQL Server for DMS replication 12](#_Toc507510154)

[Setting up the replication instance in DMS, source database and target s3 endpoints and the DMS migration tasks 16](#_Toc507510155)

[1. Create DMS replication instance 16](#_Toc507510156)

[2. Create SQL Server source endpoint 18](#_Toc507510157)

[3. Create the S3 Target Endpoint 19](#_Toc507510158)

[4. Create the required DMS tasks to move the data 22](#_Toc507510159)

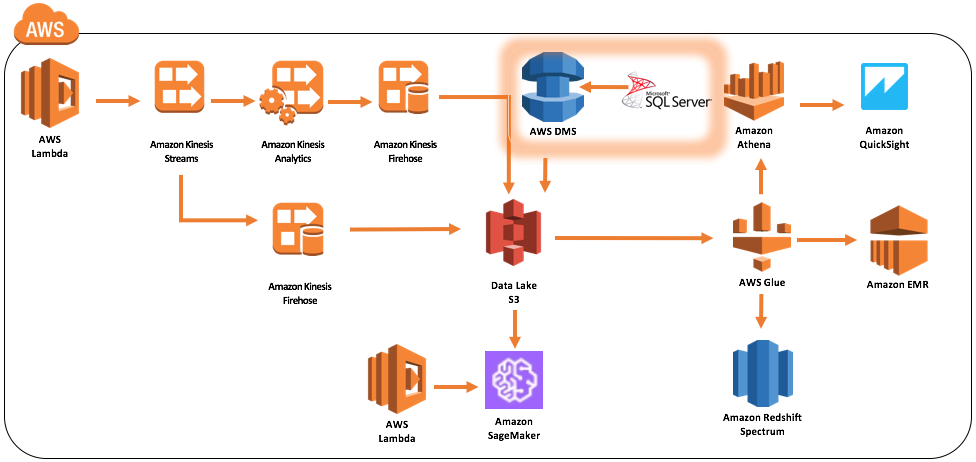
[Conclusion 25](#_Toc507510160)

# Overview

AWS [Database Migration Service](https://aws.amazon.com/dms/getting-started/) helps you migrate databases to AWS quickly and securely. The source database remains fully operational during the migration, minimizing downtime to applications that rely on the database. The AWS Database Migration Service can migrate your data to and from most widely used commercial and open-source databases.

In this lab, you will utilize Data Migration Service(DMS) to replicate data from SQL Server database to S3 Data lake.

Diagram below with highlighted area depicts what you will be building in this lab.



# Preparing AWS Resources Using CloudFormation

In this exercise, you will launch a template that creates the AWS resources needed to complete the replication using AWS Database Migration Service. You’ll launch the template using AWS CloudFormtaion.

AWS CloudFormation gives developers and system administrators an easy way to create and manage a collection of related AWS resources, provisioning and updating AWS resources by defining infrastructure as code.

This template will be run in the **us-east-1 (Virginia) Region.**

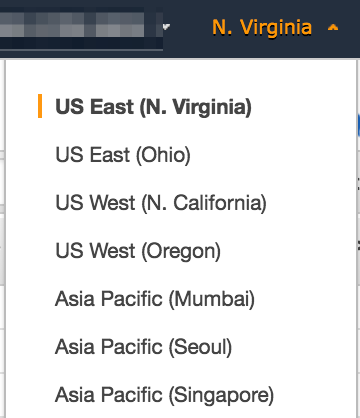
Please use Chrome or Firefox browser to ensure smooth lab experience.

## Create a Key Pair

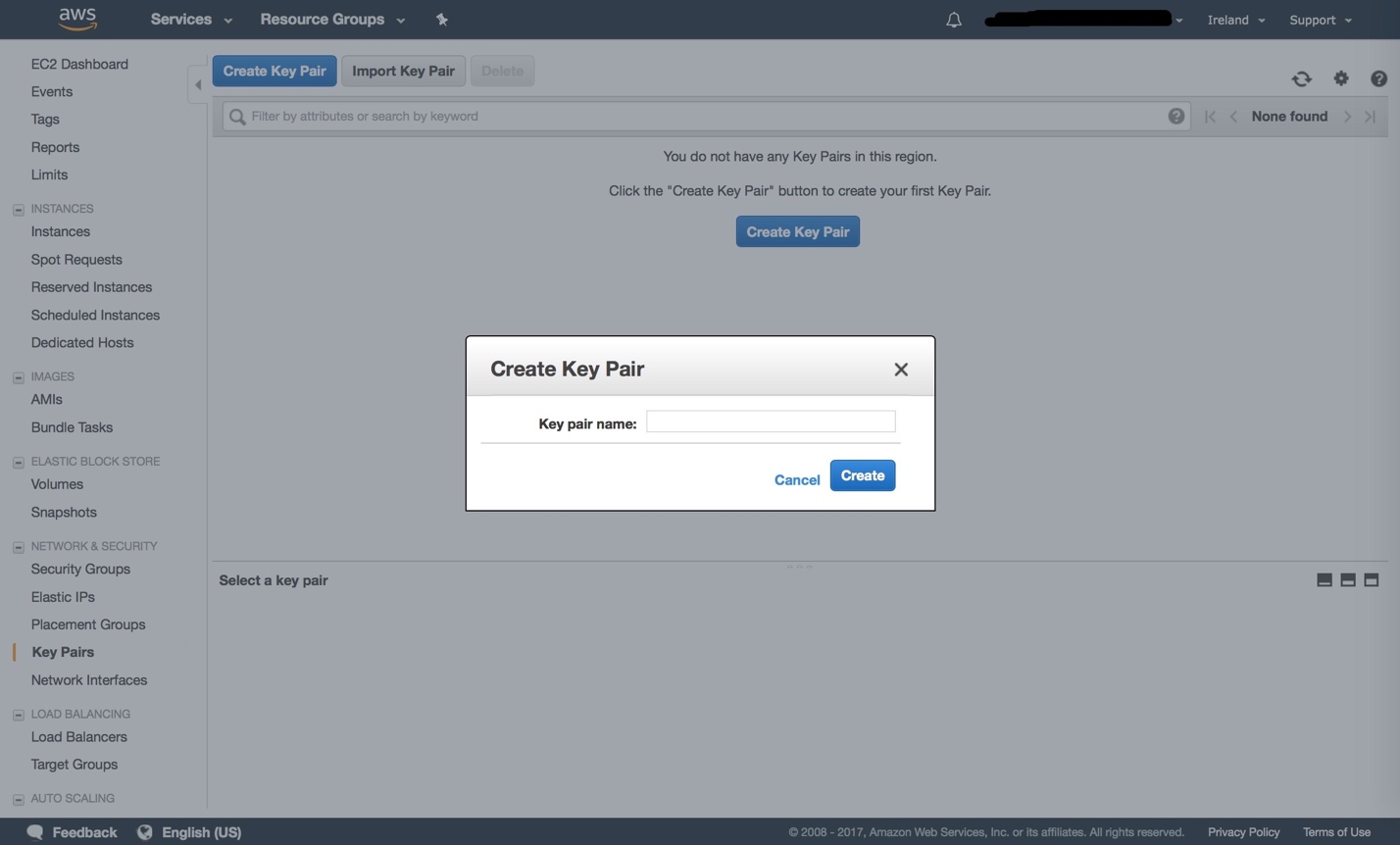
Before launching CloudFormation, you will need to create an EC2 Key Pair which will be used to allow administrative access into the EC2 Instances created by CloudFormation. If you already have one or created one from Lab 2, you can use the same key and go to **“Launch Resources using AWS CloudFormation”** section below.

Sign in to the AWS management console. Open the EC2 page by clicking the searching for EC2 or going to <https://console.aws.amazon.com/ec2/home>.

Once on the EC2 page, look in the upper right corner for the AWS Region and change this to **US East (N. Virginia)**



* Under **Network & Security**, select **Key Pairs** on the left and select **Create Key Pair**. Recommended name “DMSlab”. Please append your initials if multiple people are running in the same account, such as “DMSlab\_ws”.

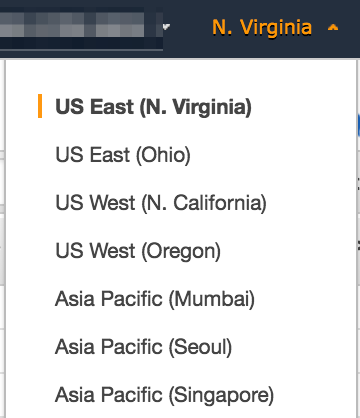


* + Mac users: You will receive a download of a pem.txt file. Rename the file and remove the .txt to make it a .pem file. For example: DMSlab\_ws.pem. Note where this file was downloaded.
  + Windows users: You will be prompted to download a .pem file. Save this, and note where this was downloaded.

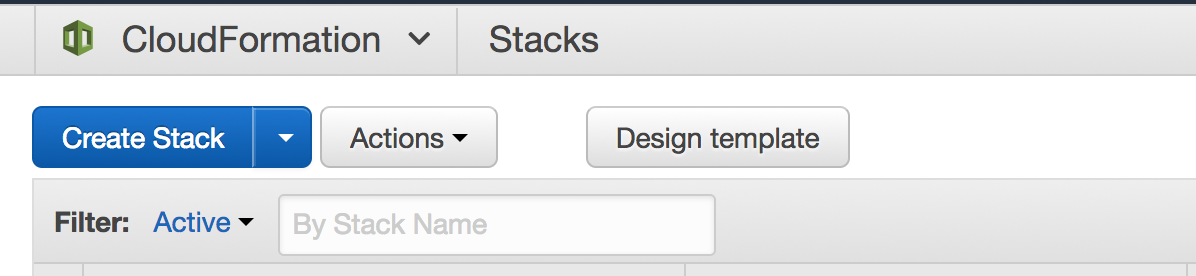
You have completed creating a new EC2 Key Pair.

# Launch Resources using AWS CloudFormation

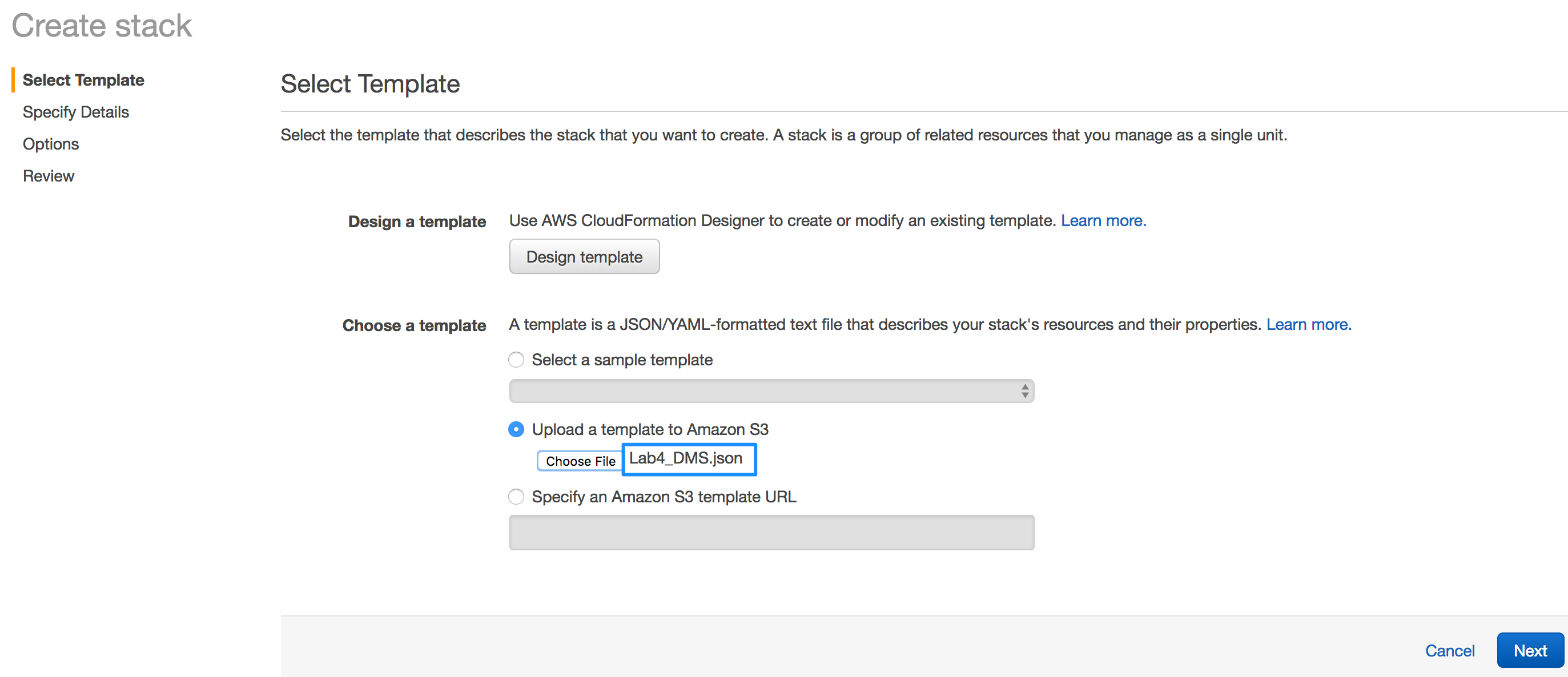
* Return to the AWS management console home by clicking the AWS icon in the upper left of the EC2 page or by going to <https://console.aws.amazon.com/cloudformation/home>.
* Once on the CloudFormation home page, look in the upper right corner for the AWS Region and ensure this is **US East (N. Virginia).**



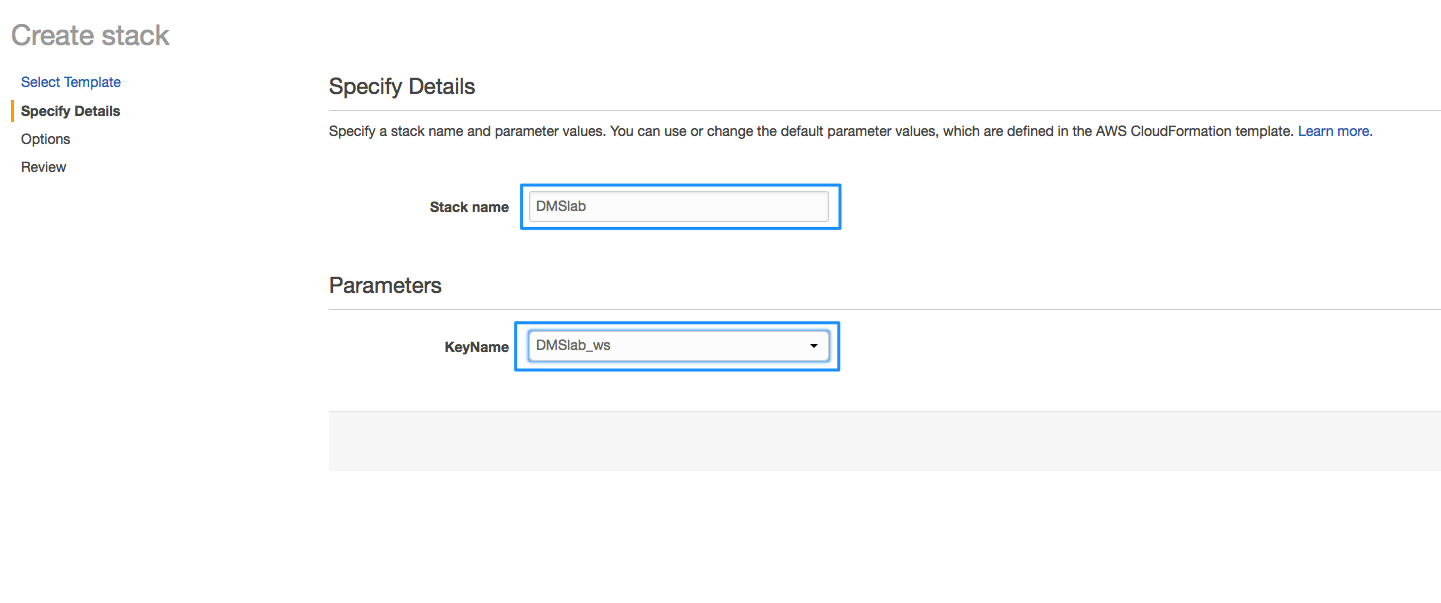
* In CloudFormation, click on CreateStack.



* On the create stack page, choose Upload a template to Amazon S3 and use the Choose file option to choose the Lab4\_DMS.json template in the zip package



* Enter the following**:**
  + **Stack name**: Enter any name you would like to give the stack. A suggested name is “DMSlab”. Please append your initials if multiple people are running in the same account such as “DMSlabws”.
  + **KeyName**: Enter the name of the key pair you created previously. Enter the name without the .pem extension, such as “DMSlab\_ws”.



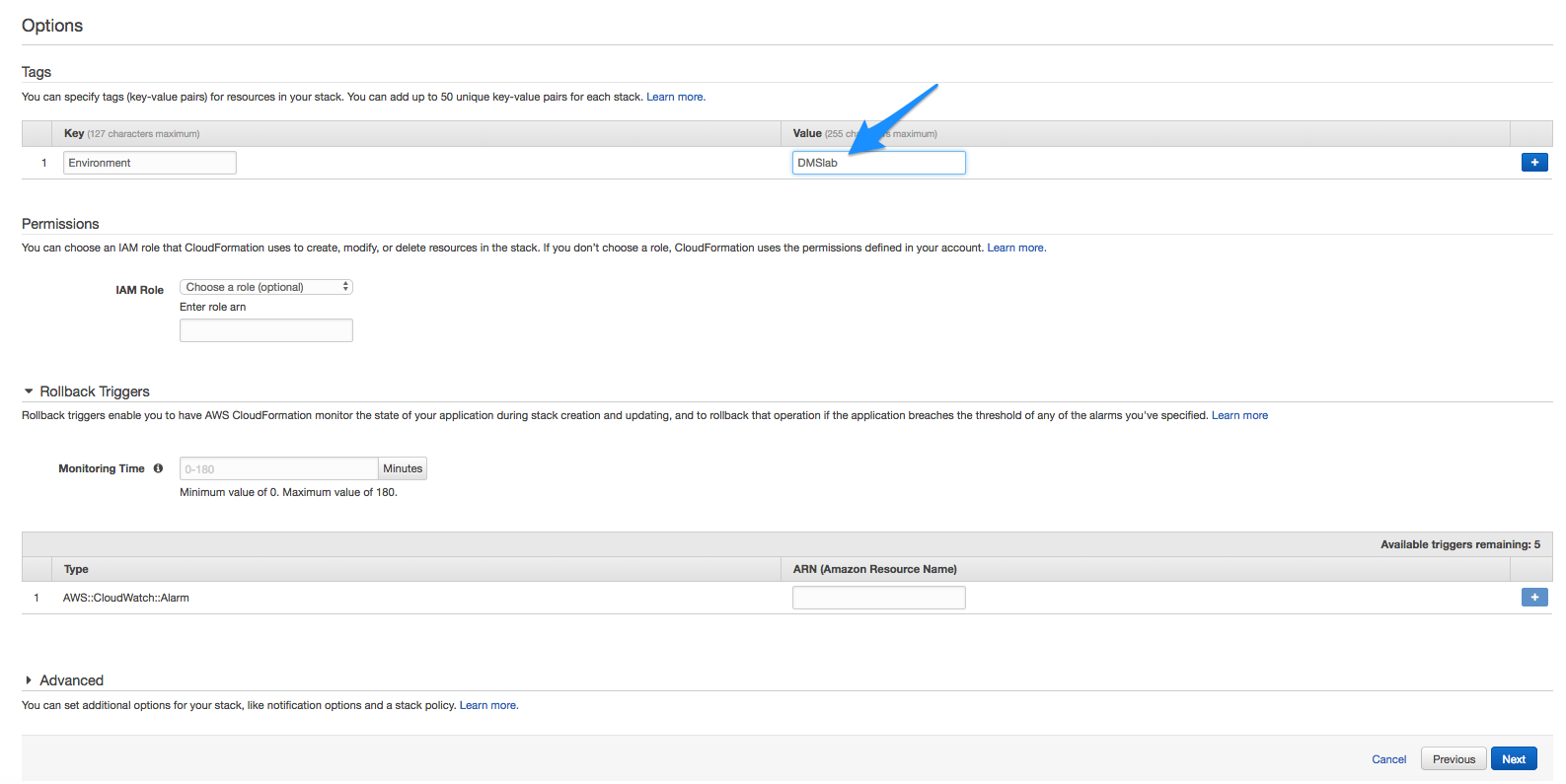
* On the **Options** page, put the following values for Tags and click **Next**.

Key = ‘Environment’  
Value = ‘DMSlab’

Under **Permissions**, you do not need to specify a role.

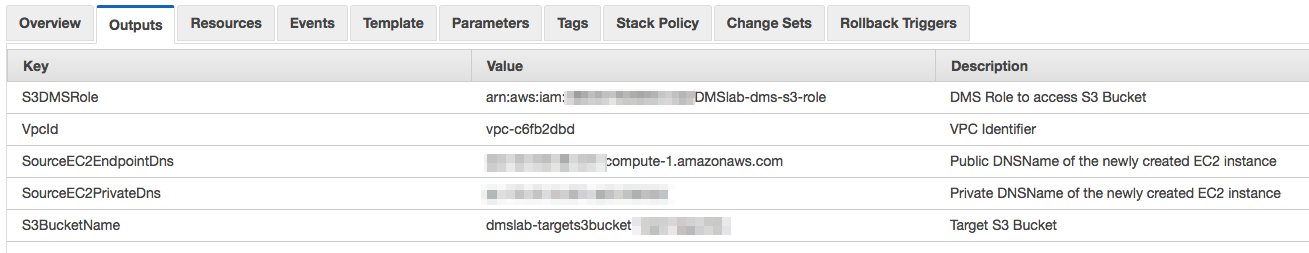
You do not need to specify any **Advanced** settings.

Make sure you acknowledge the Checkbox that states that AWS CloudFormation might create IAM resources with customer names



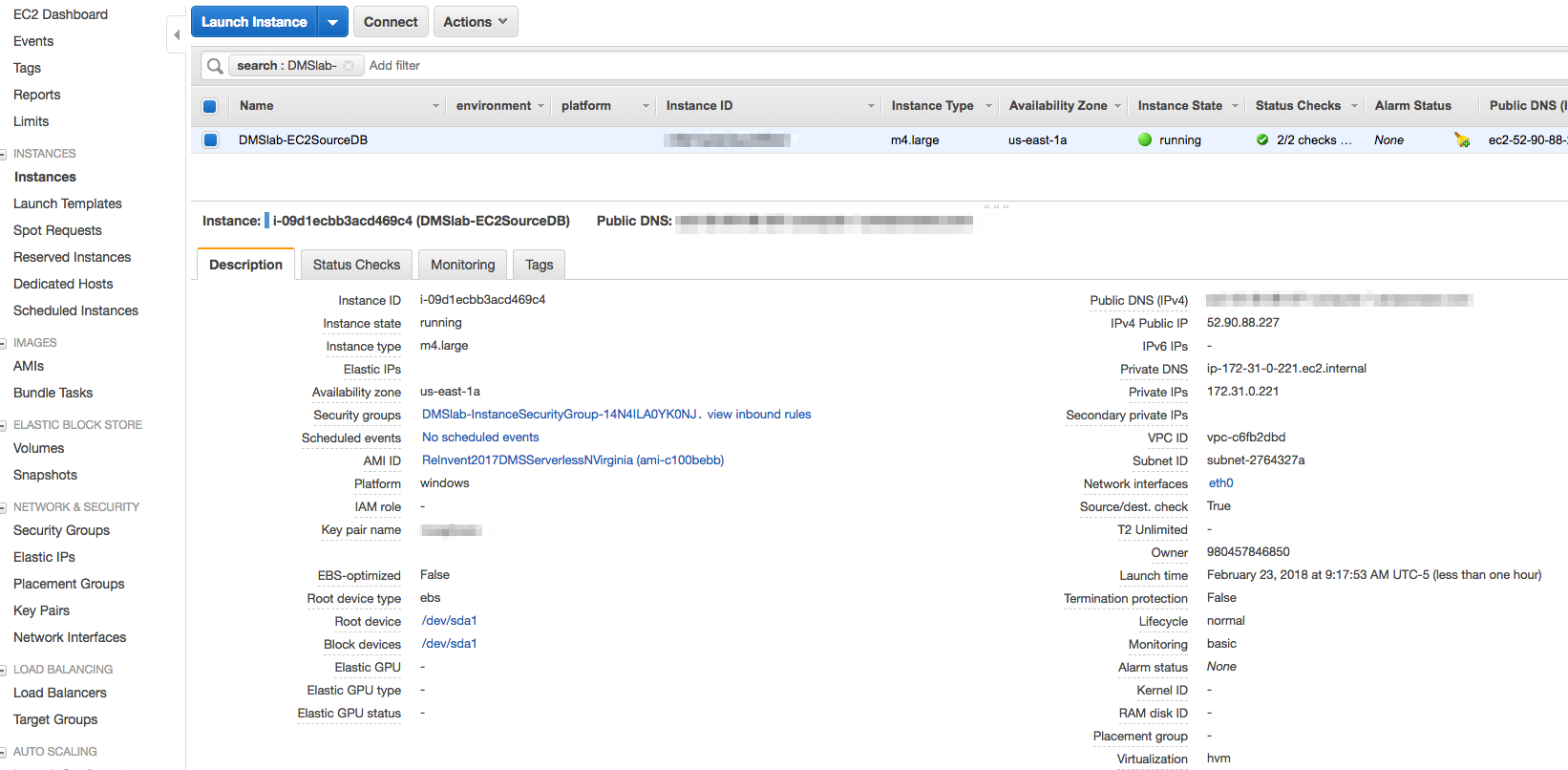
* Check through the final **Review** page and click **Create.**
* At this point, you will be directed back to the CloudFormation console and will see a status of **CREATE\_IN\_PROGRESS**. The stack will take approximately 10 minutes to launch. You can check the progress in the **Events** tab on the CloudFormation home page**.** Do not continue until the status changes to **CREATE\_COMPLETE**
* Once CloudFormation has completed building the stack, the status will change to complete. Make sure to note the values for the following keys in the **OUTPUT** tab. These will be needed as input for the Database Migration Service setup.

|  |  |
| --- | --- |
| VpcId | VPC ID of the newly created VPC |
| SourceEC2EndpointDNS | Public name of the EC2 instance that holds the Source database |
| S3BucketName | Target S3 Bucket Name |
| S3DMSRole | DMS Role to access S3 Bucket |

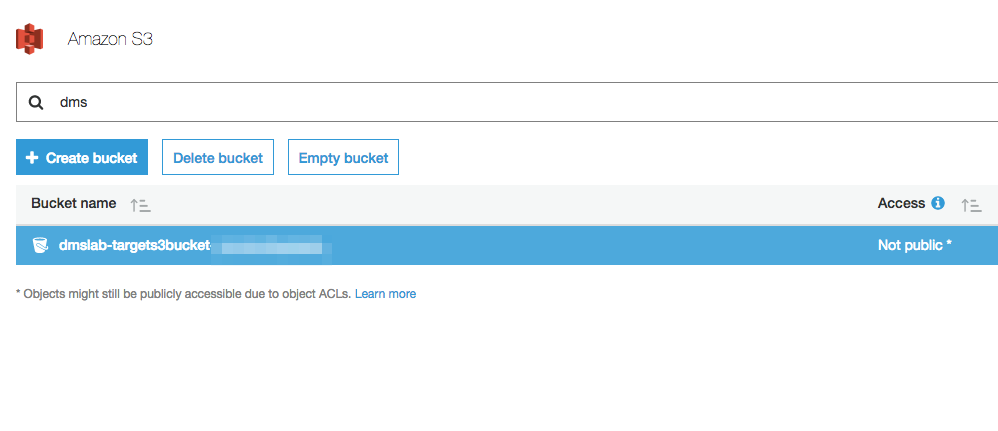


Once the Stack is complete, view the newly created resources in the AWS Management Console.

* From the Management Console home page, search for EC2 and go to the EC2 home page to view the EC2 instance that holds the source database. Look for the instance with “DMSlab” in the name:



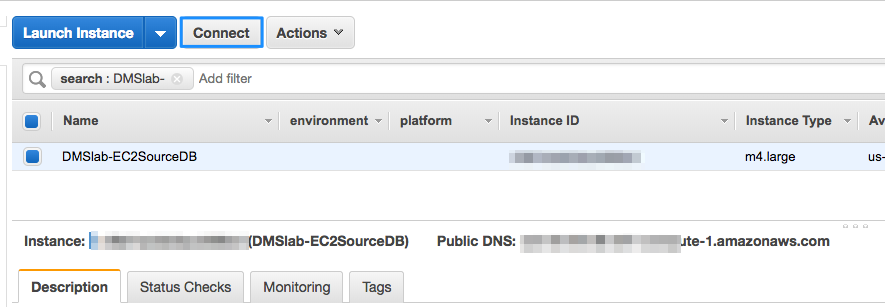
* From the Management Console home page, search for S3 and go to the S3 home page. Find and view the target S3 bucket that will be a target for the Database Migration Service.



# Connecting to your EC2 Instance

Once CloudFormation is has completed building the environment, the next step is to connect to the EC2 instance.

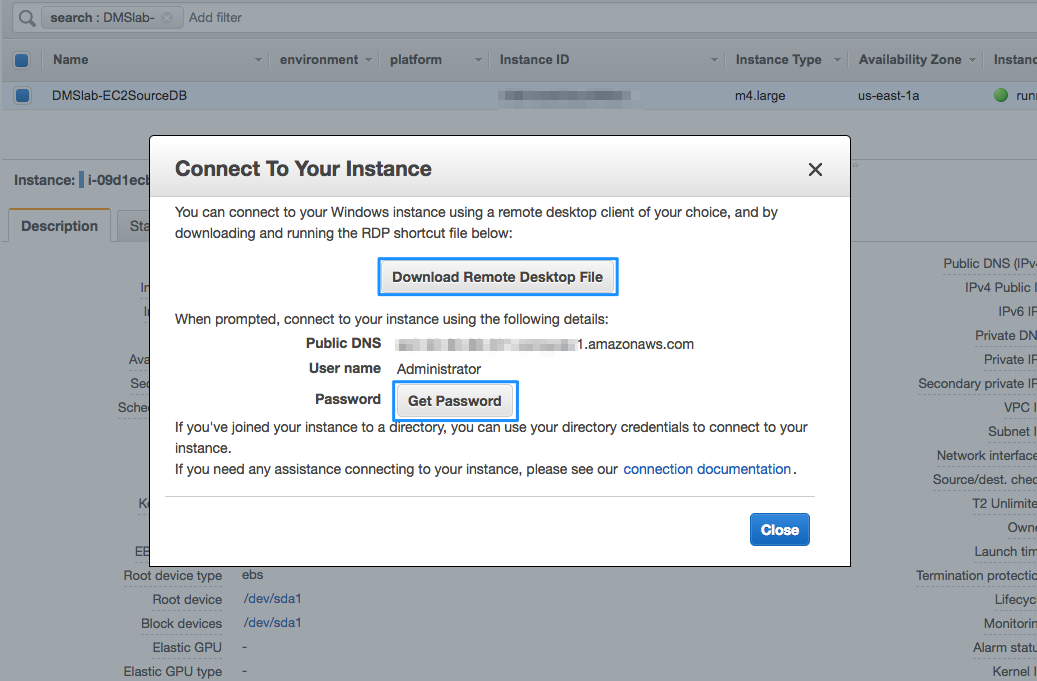
* Search for EC2 or go to <https://console.aws.amazon.com/ec2/v2/home> and click **Instances** in the left column.
* Select the instance with *DMSlab* in its name, which you viewed previously, and then click the **Connect** button at the top of the page:



* In this step, you will perform 3 tasks:
  1. Click **Get Password** and navigate to the key file that you specified when launching the CloudFormation template. Please take note of the automatically generated administrator password.
  2. Click **Download Remote Desktop File** to download the RDP file for this instance.
  3. Connect to the instance using your RDP client.

For more information on connecting to a Windows instance from a Windows or Mac PC see this document:

<http://docs.aws.amazon.com/AWSEC2/latest/WindowsGuide/connecting_to_windows_instance.html>

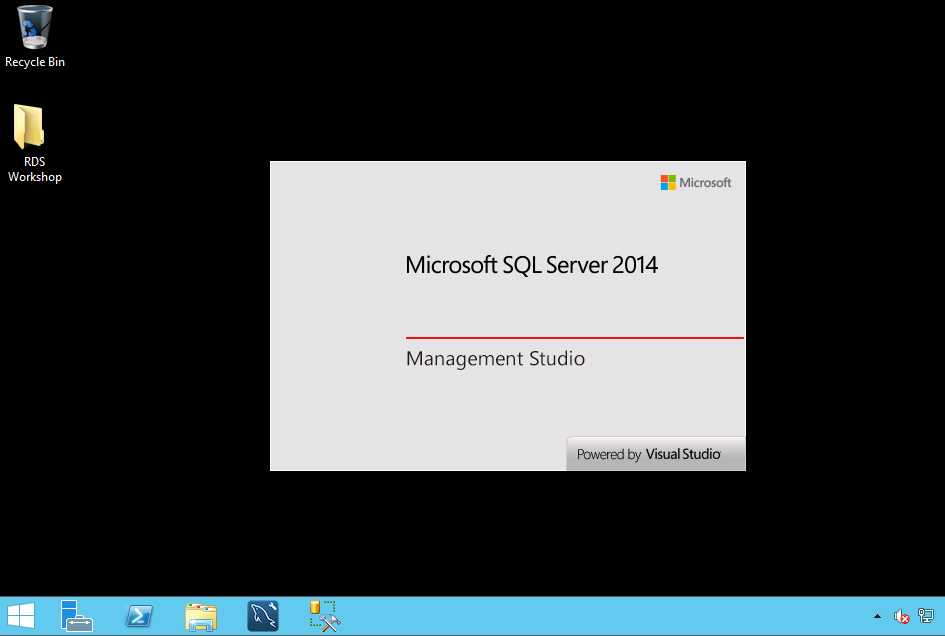


* You may see a certificate warning as you begin the RDP session. This is expected when connecting using the Public DNS name. Click Continue.
* You are now connected to the EC2 Instance.

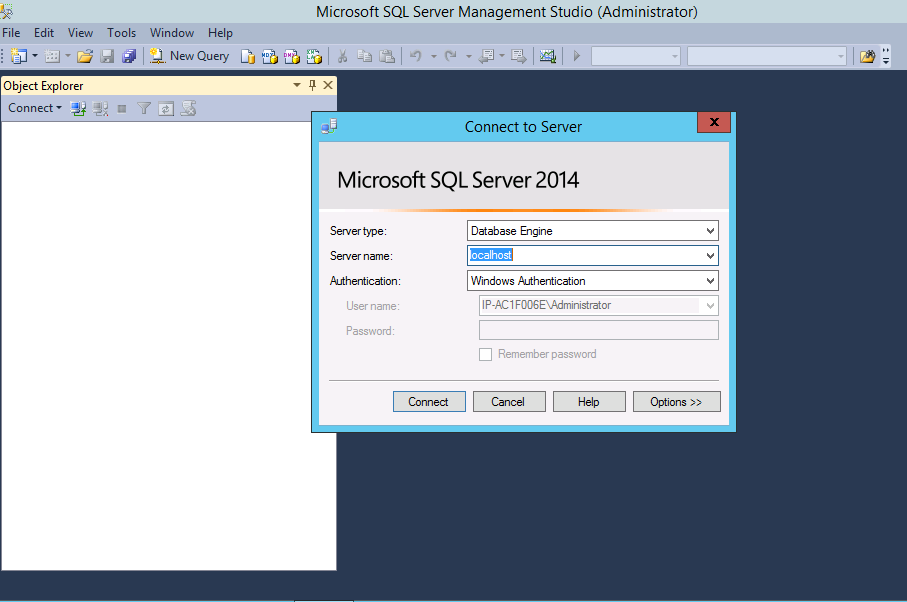


# Preparing SQL Server for DMS replication

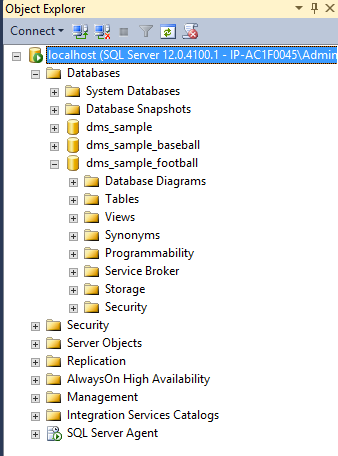
* Once logged into the server, launch SSMS (SQL Server Management 2014 Studio) using the icon displayed on the tool bar



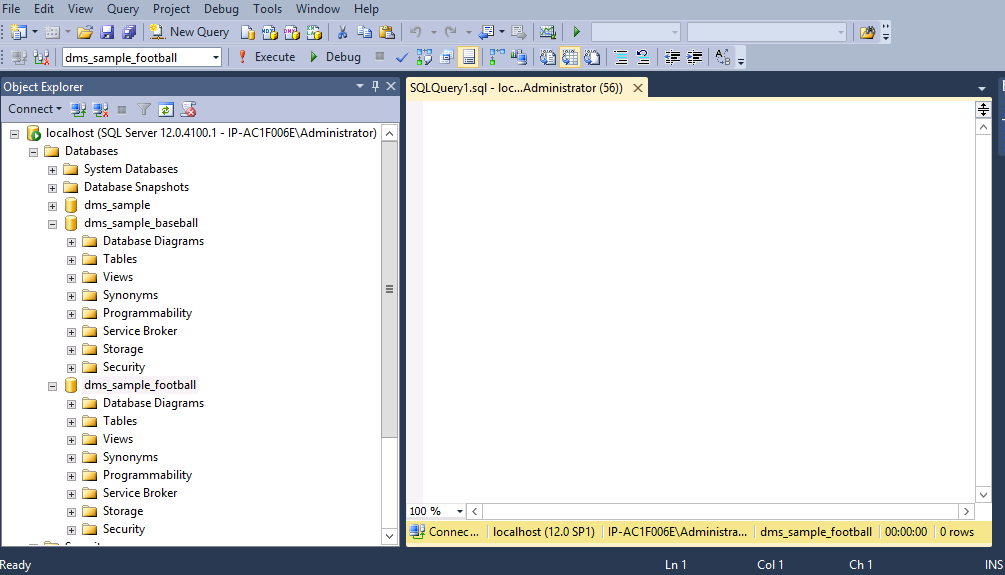
* On the SQL server authentication screen, use localhost as Server name and Windows Authentication and click Connect



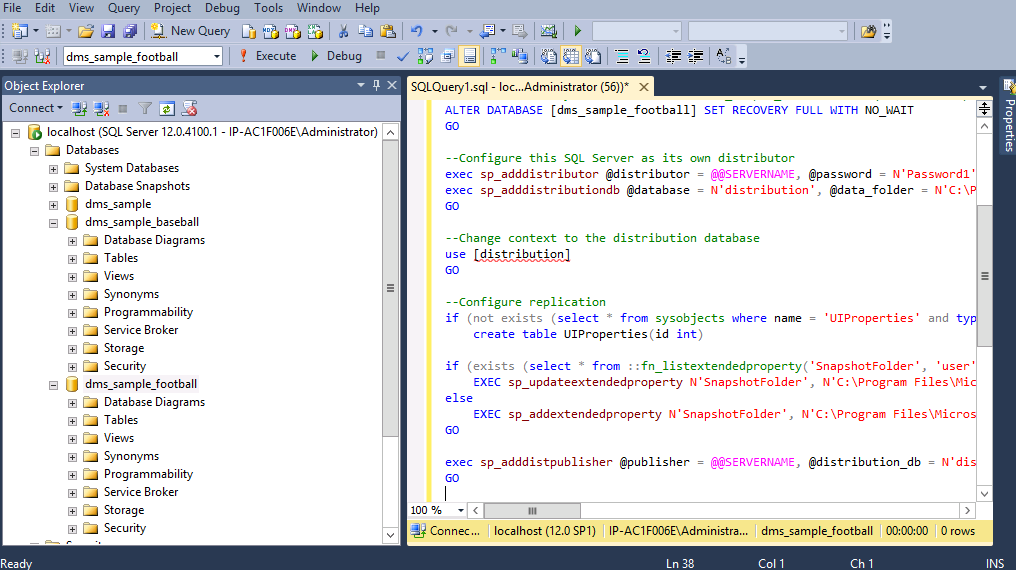
* Expand Databases folder on Object Explorer to inspect databases that are available. You should see 3 databases: dms\_sample, dms\_sample\_football, dms\_sample\_baseball. For the remainder of this workshop, we will be working on dms\_sample\_football database.
* We will use dms\_sample\_football database, convert schema and apply schema to target S3 data lake



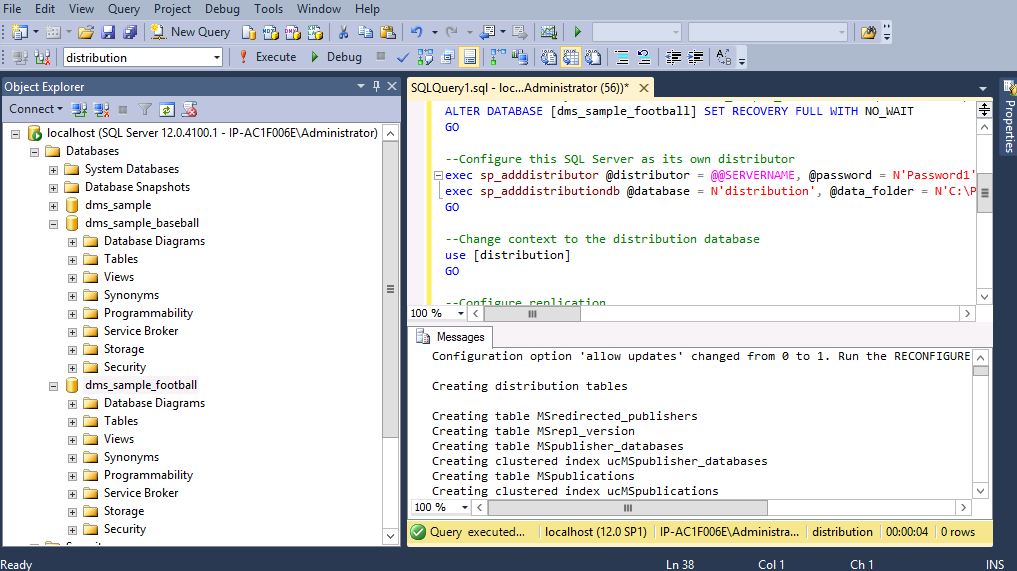
* In order to replicate football databases to Amazon S3, we need to prepare the source database for replication. This is done by configuring the database for full recovery and setting it as distribution database. For additional detailed information, please refer to documentation link here: <http://docs.aws.amazon.com/dms/latest/userguide/CHAP_Source.SQLServer.html#CHAP_Source.SQLServer.Configuration>
* Right click dms\_sample\_football database, select New Query. A blank query window should open up to the right as shown:



* We will use the provided SQL server script to configure for replication. Open the provided ConfigureSQLServer.sql script using a text editor, copy and paste the script into the blank Query editor on the SQL server



* Click on Execute button on the top bar to execute the commands for replication setup
* Once the query is executed, you should see successful query execution message on the right window as shown below:



Minimize SSMS window. Now SQL Server database is ready for replication to S3 Data lake.

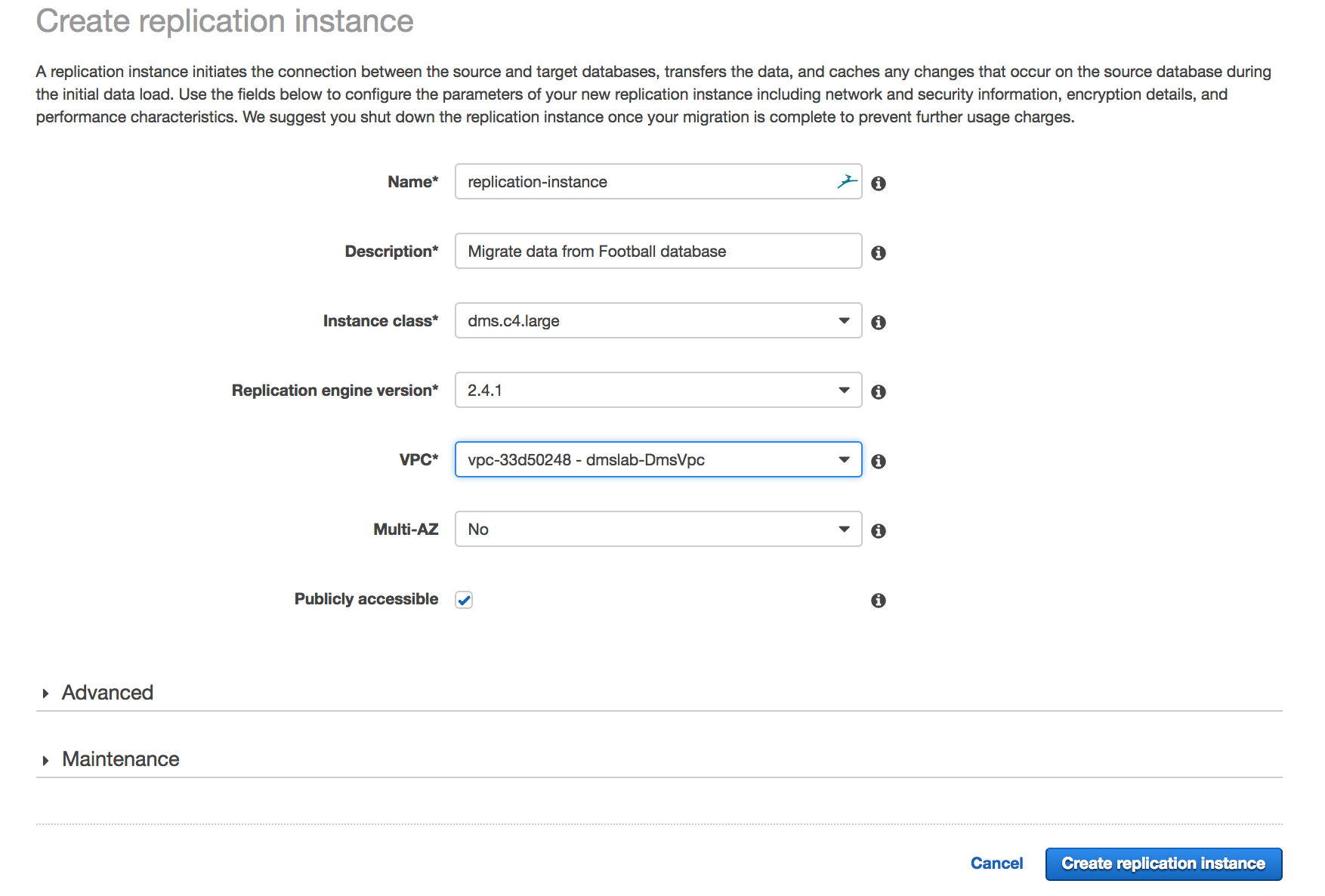
# Setting up the replication instance in DMS, source database and target s3 endpoints and the DMS migration tasks

In this exercise, you will do the following:

1. Create DMS replication instance
2. Create SQL Server source endpoint
3. Create a S3 target endpoint
4. Create SQL Server to S3 replication task

## Create DMS replication instance

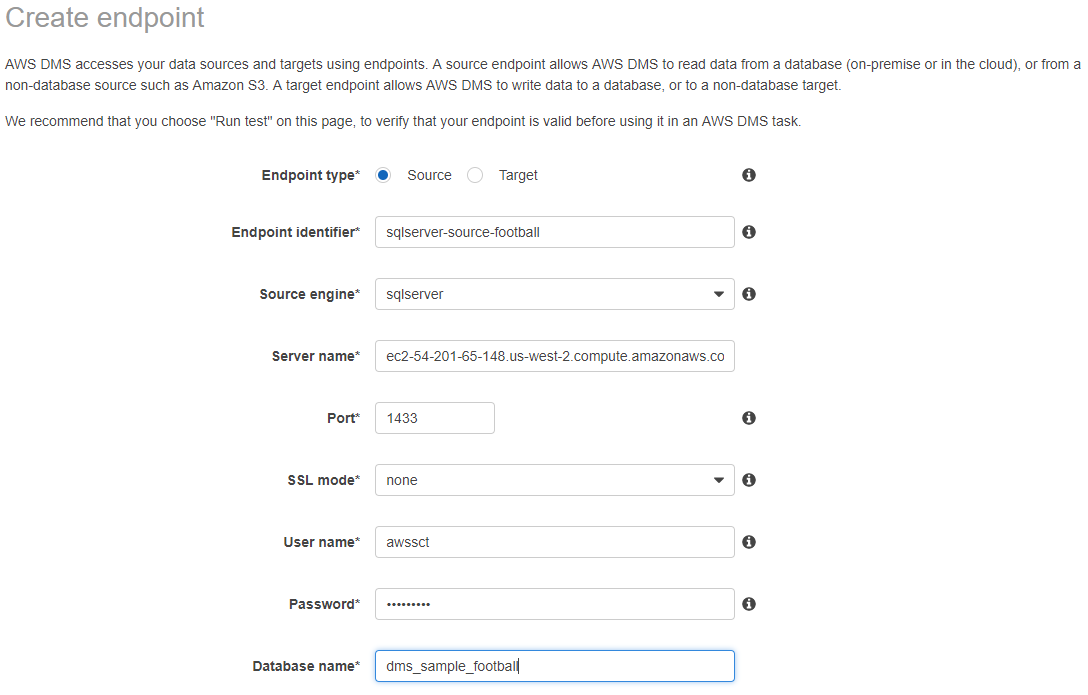
* Navigate to the Database Migration Service console
* Choose the replication instances tab (on the left)
* Choose create replication instance and fill out required details. The page should look like this after filling out the details –



* Name: **replication-instance**
* Description: **Migrate data from Football database**
* Instance Class: **dms.c4.large**
* Replication engine version: **2.4.1**
* VPC: Choose the VPC that the CloudFormation template created from the previous exercise (you will see it from the CloudFormation output)
* Note: Leave everything else default and click “Create replication instance”. Wait for the replication instance to become available before proceeding to the next step.

## Create SQL Server source endpoint

* Once the replication instances are available, click on the endpoints tab
* Click create endpoint
* Fill out the required info,

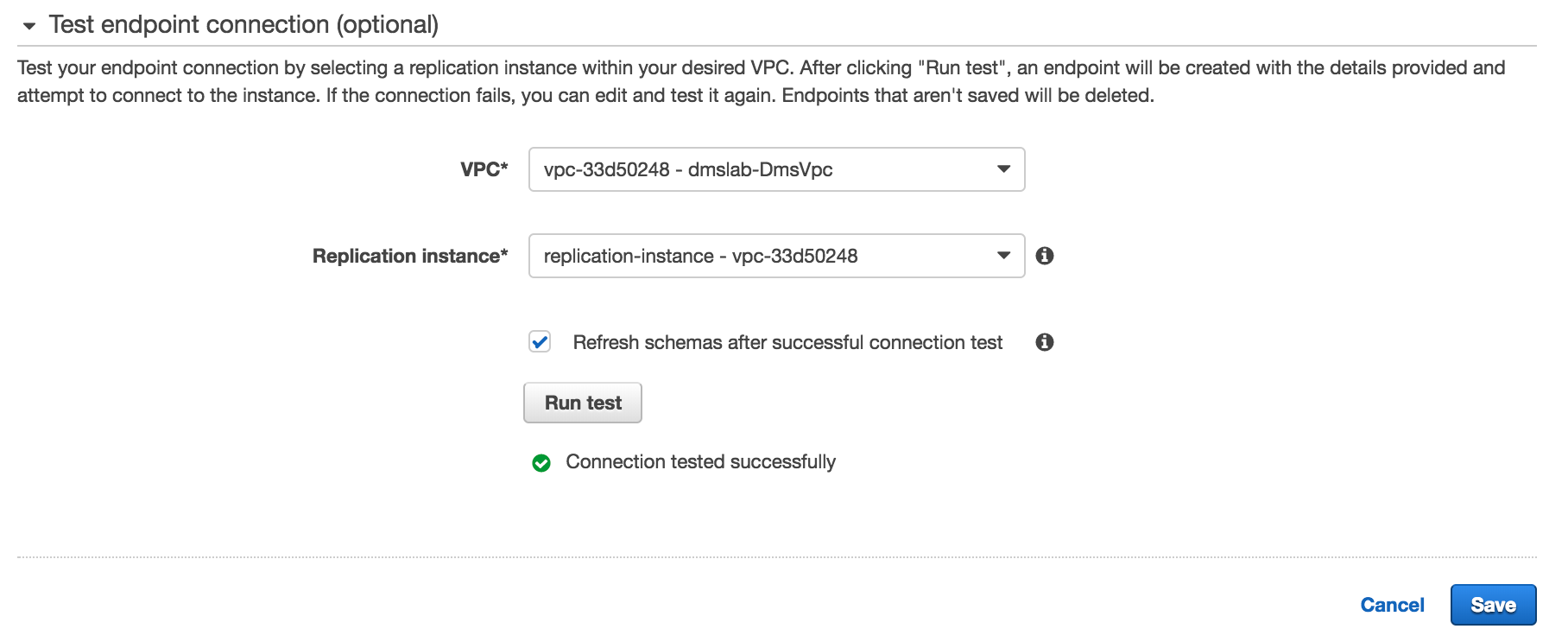


* Endpoint type: choose source (as this is going to be the source endpoint)
* Endpoint identifier: **sqlserver-source-football**
* Source engine: choose SQL Server
* Server name: put in the DNS name of the SQL Server EC2 instance
* Port: **1433**
* SSL mode: **none**
* User name: **awssct**
* Password: **Password1**
* Database name: **dms\_sample\_football**

Scroll down to get the optional endpoint test screen. Endpoint tests help you understand if the endpoint has been set up following all DMS pre-requisites and also checks if DMS can make connections to the endpoint.

* Choose the VPC and the replication instance created in step 1.
* Click run test

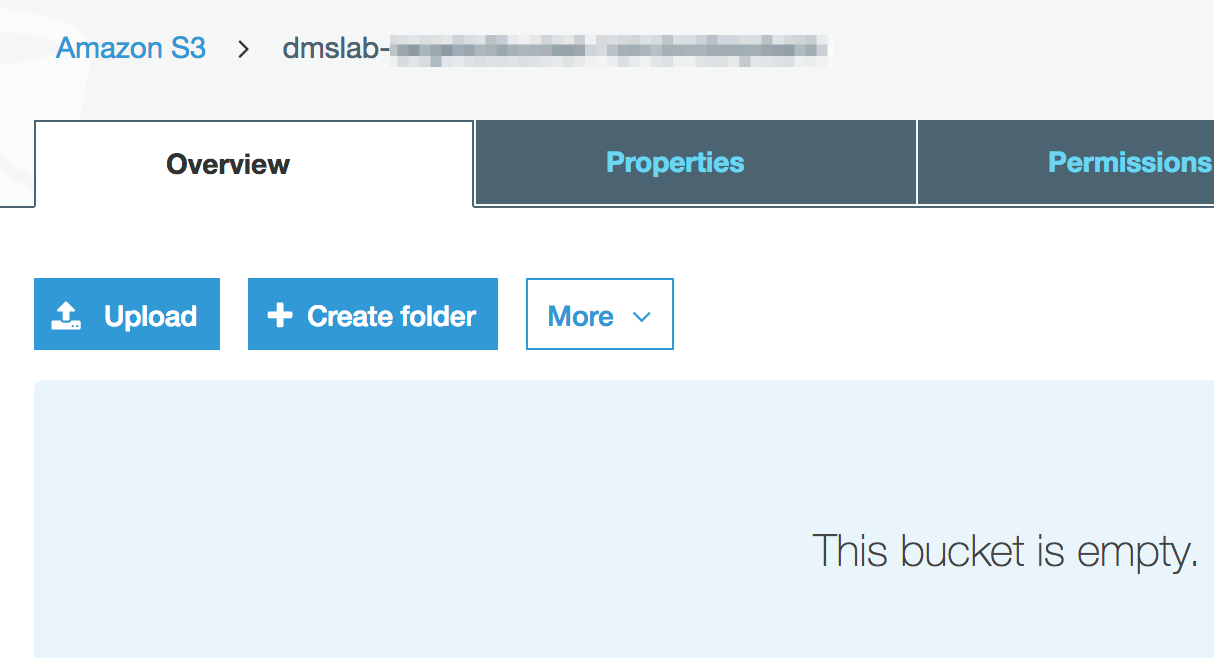
If everything has been set up well, you should see screenshot below after the endpoint connection test completes. Click Save.



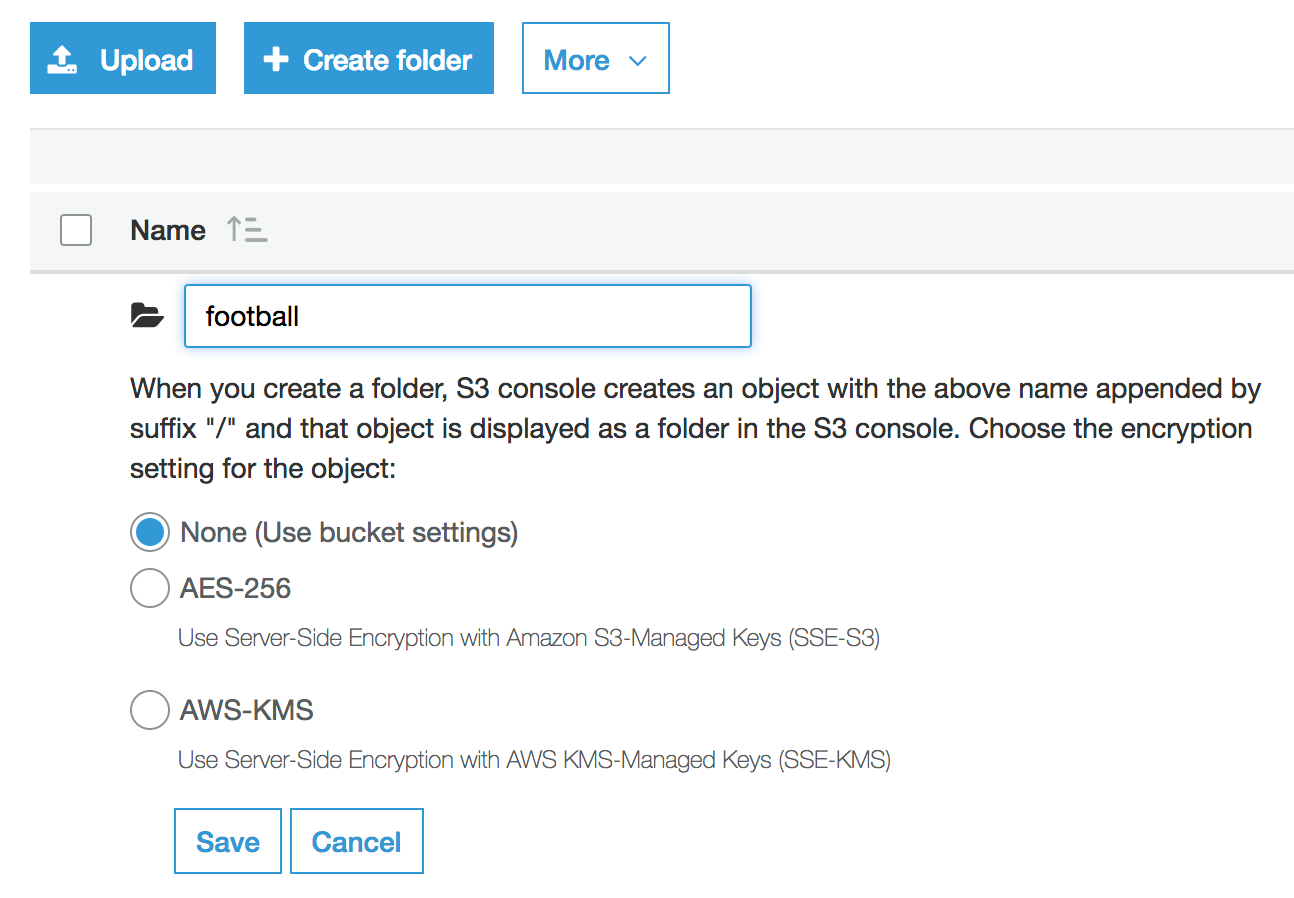
## Create the S3 Target Endpoint

* Open the S3 page from the AWS Management Console by clicking on or searching for “S3”. Locate the S3 Bucket created in this workshop. The S3 Bucket name can be found in the CloudFormation outputs.

Click Create Folder.

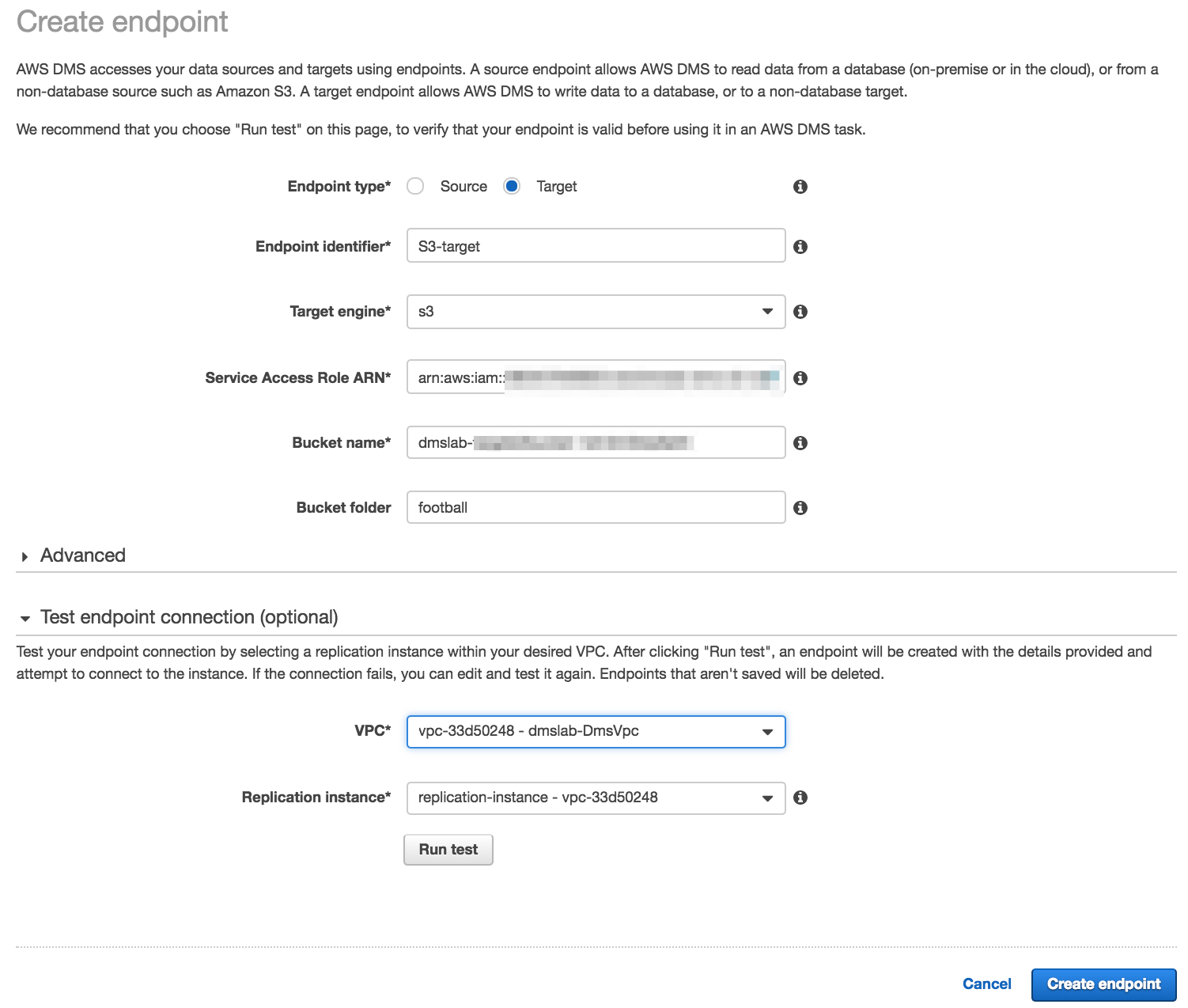


* Create a folder called “football”. Use bucket settings for encryption and click Save.



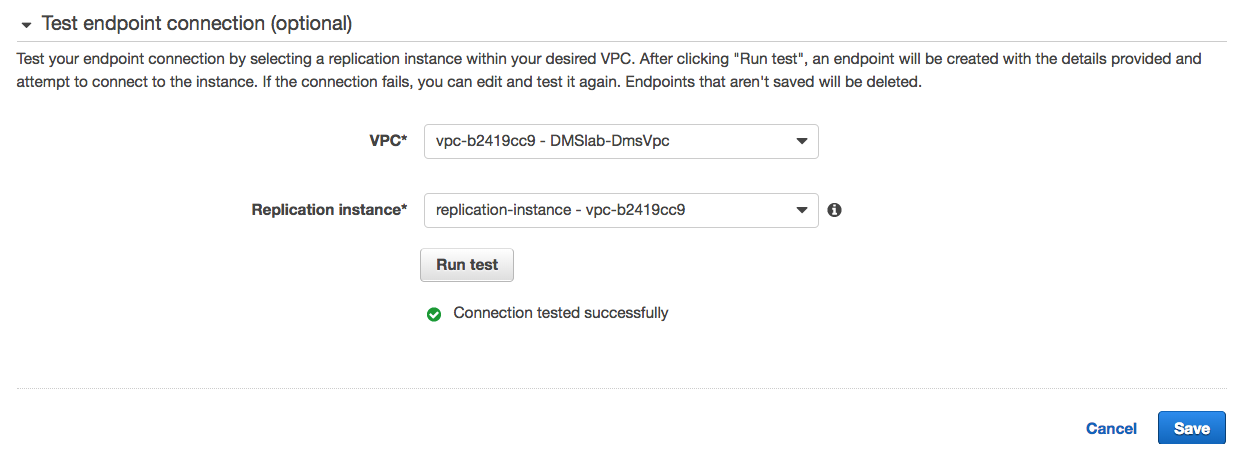
* Create S3 target endpoint using the IAM role CloudFormation created

Go to DMS console and click on endpoints on the left. Click create endpoint and fill out required info below.



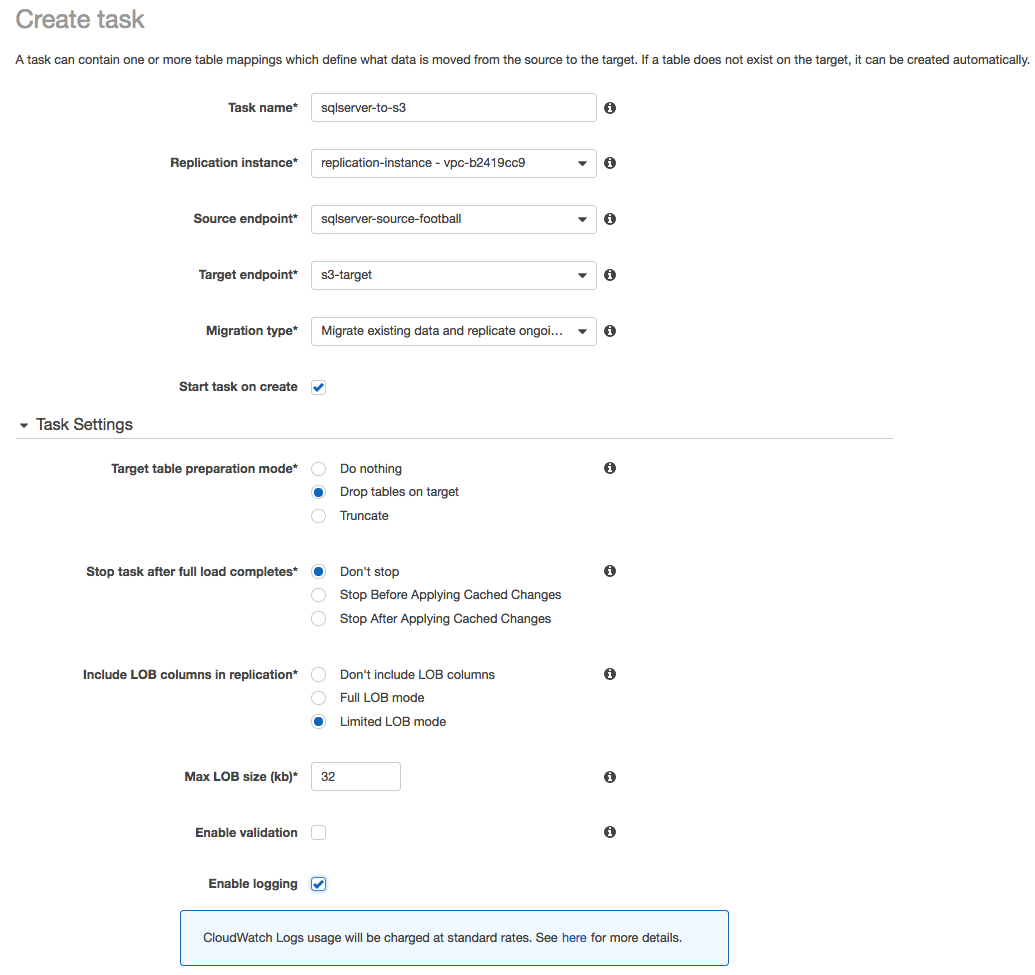
* Endpoint type: set this to target
* Endpoint identifier: **S3-target**
* Target engine: choose **s3**
* Service Access Role ARN: Paste the ARN of the IAM role from the output of the CloudFormation stack. Alternatively, you can also get this ARN from the IAM console under roles. Check to make sure there are no spaces in or after the ARN when you copy and paste.
* Bucket name: Paste the bucket name from the output of the CloudFormation stack
* Bucket folder: **football**
  + Scroll down to get the optional endpoint test screen. Endpoint tests help you understand if the endpoint has been set up following all DMS pre-requisites and also checks if DMS can make connections to the endpoint.
* Choose the VPC and the replication instance created in step 1.
* Click run test

If everything has been set up well, you should see screenshot below after the endpoint connection test completes. Click Save.

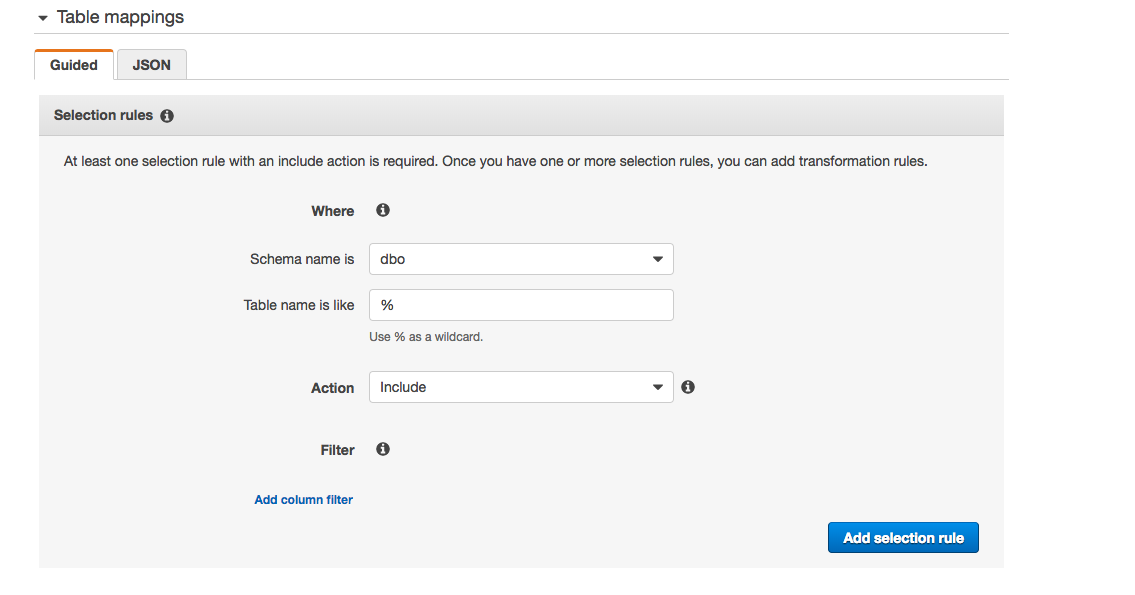


## Create the required DMS tasks to move the data

We are going to create a task to replicate SQL Server Football database to S3,Go to the tasks console in DMS and create a new task

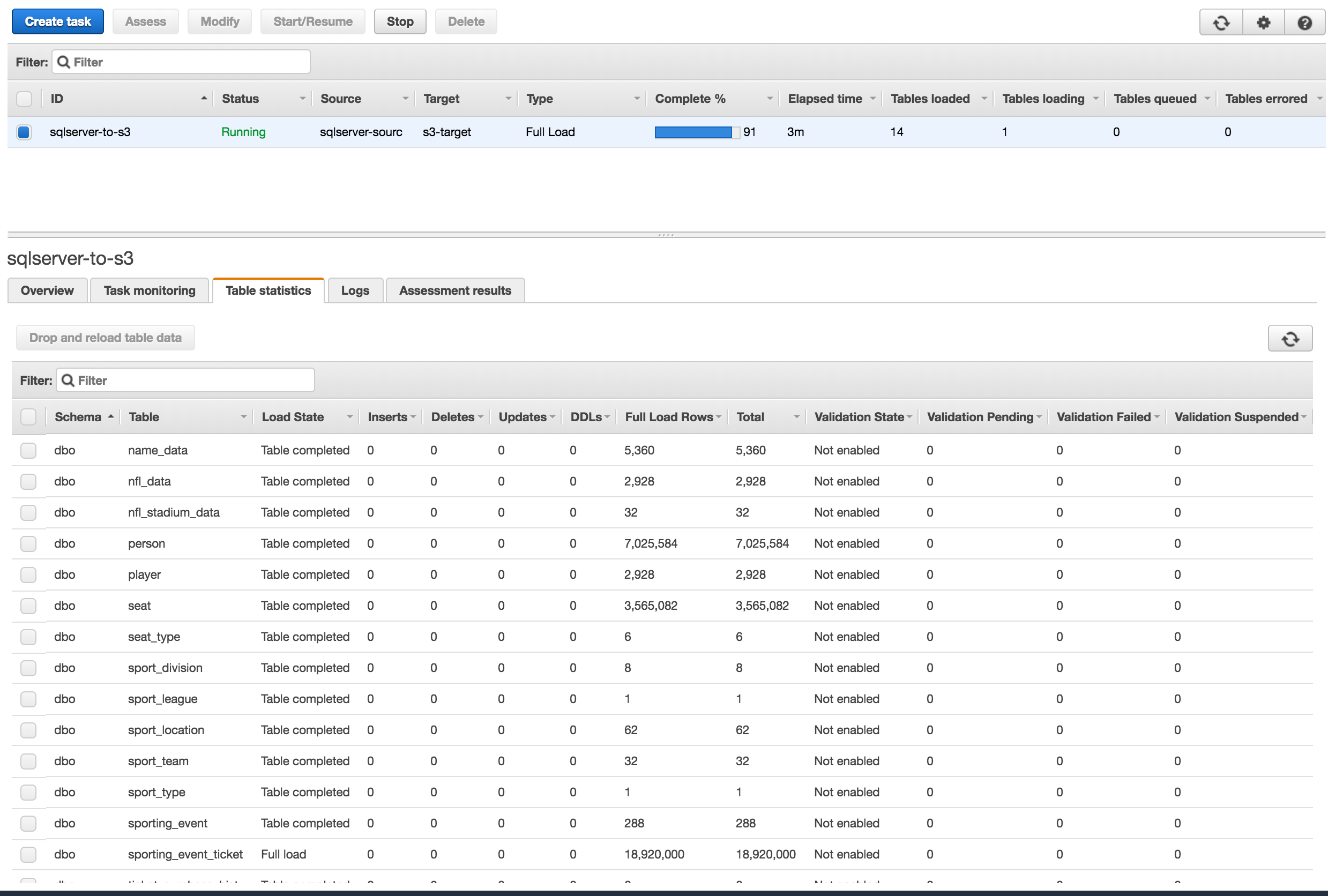


* Task name: **sqlserver-to-s3**
* Replication instance: **replication-instance**
* Source endpoint: **sqlserver-source-football**
* Target endpoint: **s3-target**
* Migration type: **Migrate existing data** **and replicate ongoi…**
* Target table preparation mode: **Drop tables on target**
* Include LOB columns in replication: **Limited LOB mode**
* Check **enable logging** and leave everything else default before scrolling down
* In the table mappings section, select the **dbo** schema and click add selection rule

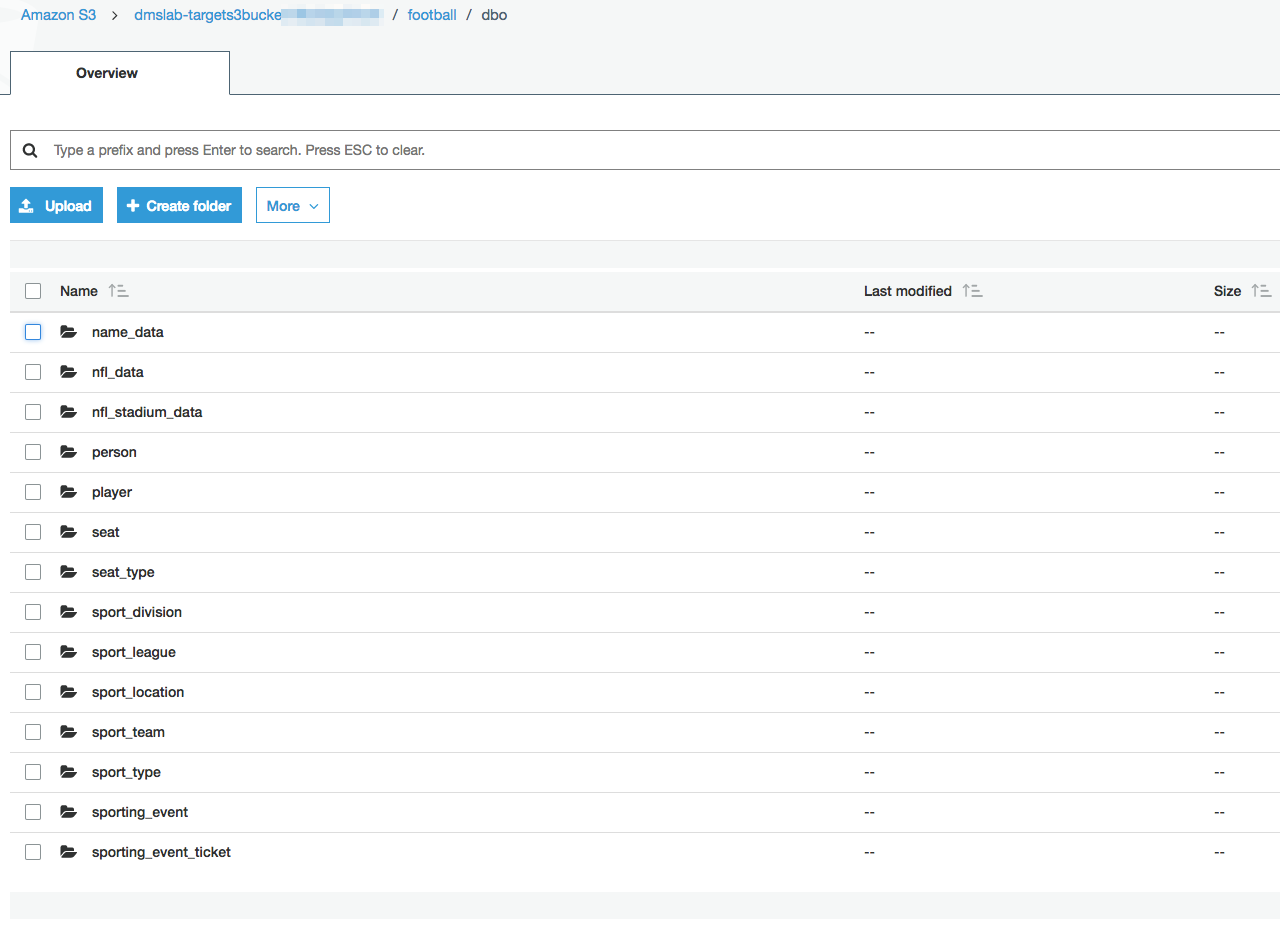


* Click Create task

The replication process will take a while, and below is a screen shot of the table statistics tab when the migration is in progress,



After Task “completion %” status changes to **100,** browse target S3 bucket in s3 console to verify data is replicated to “football / dbo” folder with every subfolder underneath representing a table containing data in CSV files.



# Conclusion

In this exercise, you learned to launch DMS replication instance, Create source and target endpoints in DMS and Run migration tasks using those endpoints to replicate data from a Microsoft SQL relational database to your data lake.