**Manually Migrating Data Between Redshift Clusters**

**Introduction**

In this lab, we'll utilize the Redshift UNLOAD and COPY commands to migrate data between an existing Redshift cluster, which we will launch in the course of completing the lab.

**Scenario**

You have been presented with a few pain points to solve around your company's Redshift solution. The original Redshift cluster that was launched for the company's analytics stack has become underpowered over time. Several groups wish to create incremental backups of certain tables to S3 in a format that can be plugged into data lake solutions, as well as other groups wishing to have select pieces of the main Redshift schema splintered to new department-specific clusters.

You've come up with a plan to utilize the UNLOAD and COPY commands to facilitate all of the above, and need to test a proof of concept to ensure that all pain points above can be addressed in this manner.

You've requisitioned a test AWS account with a Redshift cluster containing a relatively small test table in it. Also provided in the AWS account is an S3 bucket which will function as an intermediary storage point between Redshift clusters, as well as a point to test backup/data lake solutions.

The existing table has the following definition:

users\_data -

id\_value varchar(64),

name\_first varchar(64),

name\_last varchar(64),

location\_country varchar(32),

dob\_age int,

picture\_large varchar(64),

primary key(id\_value)

distkey(location\_country)

compound sortkey(id\_value);

**Additional Resources**

* [Redshift UNLOAD](https://docs.aws.amazon.com/redshift/latest/dg/r_UNLOAD.html)
* [Redshift COPY](https://docs.aws.amazon.com/redshift/latest/dg/r_COPY.html)
* [Lab Solution Github](https://github.com/linuxacademy/Content-AWS-Certified-Data-Analytics---Speciality/tree/master/Lab_Assets/manually_migrating_data_between_redshift_clusters)

**Solution**

Log in to the AWS Management Console using the credentials provided for the lab. Make sure you're in the us-east-1 region.

**Investigate the Lab Environment**

1. Navigate to S3 using the *Services* menu or the unified search bar.
2. Select the users-data-<ACCOUNT\_NUMBER> bucket and note the bucket name. There are currently no objects in the bucket.
3. Navigate to Amazon Redshift using the *Services* menu or the unified search bar.
4. Select the **users-cluster** link to review the cluster's configuration. You will recreate this cluster with the ID users-cluster-2.
5. Select **Editor** from the sidebar menu to open the query editor.
6. Click **Connect to Database** and fill in the connection details:
   * Ensure the *Connection* is set to **Create a new connection**.
   * Ensure the *Authentication* is set to **Temporary credentials**. This lets you use your IAM permissions to log in to the Redshift cluster.
   * Use the *Cluster* dropdown to select **users-cluster (Available)**.
   * In the *Database name* field, enter users.
   * In the *Database user* field, enter users\_admin.
7. After the database connection details are filled in, click **Connect**.
8. In the *Resources* pane on the left, use the *Select schema* dropdown to select **public**. You should see two schemas: *users\_data* and *users\_data\_pkey*.
9. In the query editor, run the following query to review some sample data:

select \* from users\_data limit 10;

**Launch the Target Redshift Cluster**

1. Select **Clusters** from the sidebar menu, then click **Create cluster**.
2. Configure the *Cluster configuration* settings:
   * In the *Cluster identifier* field, enter users-cluster-2.
   * Select the **Free trial** option to automatically set your cluster configuration.
3. Configure the *Database configurations* settings:
   * In the *Database name (optional)* field, enter users.
   * In the *Master user name* field, enter users\_admin.
   * In the *Master user password* field, enter a password (e.g. Strongpass1).
4. Expand the *Cluster permissions (optional)* section.
5. Use the *Available IAM roles* dropdown to select **RedshiftS3**, then click **Associate IAM role** on the right.
6. Leave all other default settings and click **Create cluster**. The cluster takes some time to create.
7. After the cluster is created, open it and select the **Properties** tab. Verify the IAM role is associated.

**Copy the Existing Redshift Table to S3**

1. Select **Editor** from the sidebar menu to open the query editor.
2. Copy the UNLOAD statement provided in the [GitHub repository](https://github.com/linuxacademy/Content-AWS-Certified-Data-Analytics---Speciality/tree/master/Lab_Assets/manually_migrating_data_between_redshift_clusters) and paste it into the query editor.
3. UNLOAD ('select \* from users\_data')
4. TO '<users-data-bucket>'
5. IAM\_ROLE '<RedshiftS3 ARN>'

FORMAT AS PARQUET;

1. Replace <users-data-bucket> with your S3 bucket name and preface the bucket name with s3://, so bucket name follows the format s3://<users-data-bucket>.
2. Replace <RedshiftS3 ARN> with the ARN provided in the lab resources.
3. Click **Run**. If your session times out, re-enter users in the *Database name* field and users\_admin in the *Database user* field, then click **Connect**.
4. Open S3 in a new browser tab and select the users-data-<ACCOUNT\_NUMBER> bucket name. You should see two *.parquet* objects.
5. Go back to the query editor and click **Change connection**.
6. Update the connection details:
   * Ensure the *Connection* field is set to **Create a new connection**.
   * Ensure the *Authentication* field is set to **Temporary credentials**.
   * Use the *Cluster* dropdown to select **users-cluster-2 (Available)**.
   * In the *Database name* field, enter users.
   * In the *Database user* field, enter users\_admin.
7. Click **Connect** to connect to the cluster.
8. In the *Resources* pane on the left, use the *Select schema* dropdown to select **public**. Note there are currently no tables in the schema. You must create a table before you can copy data to the new cluster.
9. Copy the create table statement provided in the [GitHub repository](https://github.com/linuxacademy/Content-AWS-Certified-Data-Analytics---Speciality/tree/master/Lab_Assets/manually_migrating_data_between_redshift_clusters) and paste it into the query editor.
10. create table users\_data(
11. id\_value varchar(64),
12. name\_first varchar(64),
13. name\_last varchar(64),
14. location\_country varchar(32),
15. dob\_age int,
16. picture\_large varchar(64),
17. primary key(id\_value)
18. )
19. distkey(location\_country)

compound sortkey(id\_value);

1. Click **Run**. You should now see two *public* schemas: *users\_data* and *users\_data\_pkey*.

**Copy Data from S3 to the Newly Launched Redshift Cluster**

1. Copy the COPY statement provided in the [GitHub repository](https://github.com/linuxacademy/Content-AWS-Certified-Data-Analytics---Speciality/tree/master/Lab_Assets/manually_migrating_data_between_redshift_clusters) and paste it into the query editor.
2. COPY users\_data
3. FROM '<users-data-bucket>'
4. IAM\_ROLE '<RedshiftS3 ARN>'

FORMAT AS PARQUET;

1. Replace <users-data-bucket> with your S3 bucket name and preface the bucket name with s3://, so the bucket name follows the format s3://<users-data-bucket>.
2. Replace <RedshiftS3 ARN> with the ARN provided in the lab resources.
3. Click **Run** to run the query.
4. After the COPY query is complete, run the following query to view some sample data:

select \* from users\_data limit 10;

This data should be identical to the data you originally saw in *users-cluster*.

**Check Your Data**

1. To verify the data is identical for both clusters, review the data in your *users-cluster-2* cluster and note a couple of the entries.
2. Click **Change connection** and fill in the connection details:
   * In the *Connection* field, select *Use a recent connection*.
   * Use the *Recent connection* dropdown to select **users-cluster (Available)**.
3. Click **Connect** to connect to the cluster. Your select \* from users\_data limit 10; query should still be available in the query editor.
4. Click **Run** to run the query again. Your data should be the same as the data you noted from *users-cluster-2*.

**Conclusion**

Congratulations — you've completed this hands-on lab!