Amazon Redshift- Learning doc

For bringing your own data to Amazon Redshift follow the below steps.

➢ **Create an IAM role**

- For any operation that accesses data from another AWS resource, your cluster needs permission to access the resource and the data on the resource on your behalf. An example is using a COPY command to load data from Amazon Simple Storage Service (Amazon S3).
- You provide those permissions by using AWS Identity and Access Management (IAM). You can do this through an IAM role that is attached to your cluster. Or you can provide the AWS access key for an IAM user that has the necessary permissions.
- To best protect your sensitive data and safeguard your AWS access credentials, it is recommended to create an IAM role and attaching it to your cluster. In this step, you create a new IAM role that allows Amazon Redshift to load data from Amazon S3 buckets. An IAM role is an IAM identity that you can create in your account that has specific permissions. In the next step, you have to attach the role to your cluster.

✓ **To create an IAM role for Amazon Redshift**

1. Sign in to the AWS Management Console and open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane, choose Roles.
3. Choose Create role.
4. In the AWS Service group, choose Redshift.
5. Under Select your use case, choose Redshift - Customizable, then choose Next: Permissions.
6. On the Attach permissions policies page, choose AmazonS3ReadOnlyAccess. You can leave the default setting for Set permissions boundary. Then choose Next: Tags.
7. The Add tags page appears. You can optionally add tags. Choose Next: Review.
8. For Role name, enter a name for your role.
9. Review the information, and then choose Create Role.
10. Choose the role name of the role that you just created.
11. Copy the Role ARN value to your clipboard—this value is the Amazon Resource Name (ARN) for the role that you just created. You use that value when you use the COPY command to load data in later steps.
12. Now that you have created the new role, your next step is to attach it to your cluster. You can attach the role when you launch a new cluster or you can attach it to an existing cluster.

➢ Create a sample Amazon Redshift cluster

1. After you have the prerequisites completed, you can launch an Amazon Redshift cluster.
2. To create an Amazon Redshift cluster **s**ign in to the AWS Management Console and open the Amazon Redshift console at https://console.aws.amazon.com/redshift/.
3. At upper right, choose the AWS Region where you want to create the cluster.
4. On the navigation menu, choose CLUSTERS, then choose Create cluster. The Create cluster page appears.
5. In the Cluster configuration section, specify values for Cluster identifier, Node type, Nodes, and how you plan to use the cluster. For Cluster identifier, it must be unique. Also, the identifier must be from 1–63 characters using as valid characters a–z (lowercase only) and - (hyphen).
6. Choose Help me choose if you don't know how large to size your cluster. Doing this starts a sizing calculator that asks you questions about the size and query characteristics of the data that you plan to store in your data warehouse.
7. Choose ‘I'll choose’ option if you know the required size of your cluster (that is, the node type and number of nodes). Then choose a Node type value and number for Nodes to size your cluster.
8. In the Database configuration section, specify values for Database name (optional), Database port (optional), Admin username, and Admin user password. Or choose Generate password to use a password generated by Amazon Redshift.
9. In the Cluster permissions section, for Available IAM roles choose the IAM role that you previously created as per service requirement.
10. In the Additional configurations section, turn off Use defaults to modify Network and security, Database configuration, Maintenance, Monitoring, and Backup settings.
11. Choose Create cluster.

✓ Configure inbound rules for SQL clients

If you use an SQL client from outside your firewall to access the cluster, make sure that you grant inbound access. This step can be skipped if you plan to access the cluster with the Amazon Redshift query editor from within your VPC.

✓ To check your firewall and grant inbound access to your cluster

a) Check your firewall rules if your cluster needs to be accessed from outside a firewall. For example, your client might be an Amazon Elastic Compute Cloud (Amazon EC2) instance or an external computer.

b) To access from an Amazon EC2 external client, add an ingress rule to the security group attached to your cluster that allows inbound traffic. You add Amazon EC2 security group rules in the Amazon EC2 console.

c) For e.g., a CIDR/IP of 192.0.2.0/24 allows clients in that IP address range to connect to the cluster. Similarly, check the correct CIDR/IP for your environment.

➢ Grant access to the query editor and run queries

To query databases hosted by the Amazon Redshift cluster, we have two options:

- Connect to your cluster and run queries on the AWS Management Console with the query editor. If you use the query editor, you don't have to download and set up an SQL client application.
- Connect to your cluster through an SQL client tool, such as SQL Workbench/J. Using the Amazon Redshift query editor is the easiest way to run queries on databases hosted by your Amazon Redshift cluster. After creating your cluster, you can immediately run queries using the Amazon Redshift console.
- To use the Amazon Redshift query editor, you need permission. To set access, attach the AmazonRedshiftQueryEditor and AmazonRedshiftReadOnlyAccess IAM policies to the IAM user that you use to access your cluster.
- If you have already created an IAM user to access Amazon Redshift, you can attach the AmazonRedshiftQueryEditor and AmazonRedshiftReadOnlyAccess policies to that user. If you haven't created an IAM user yet, create one and attach the policies to the IAM user.

✓ To attach the required IAM policies for the query editor

1. Sign into the AWS Management Console and open the IAM console at https://console.aws.amazon.com/iam/.
2. Choose Users.
3. Choose the user that needs access to the query editor.
4. Choose Add permissions.
5. Choose Attach existing policies directly.
6. For Policy names, choose AmazonRedshiftQueryEditor and AmazonRedshiftReadOnlyAccess.
7. Choose Next: Review.
8. Choose Add permissions.
9. Select the query editor and run SQL commands, view details about how queries run, save a query and download a query result set.

✓ To use the query editor

1. Sign in to the AWS Management Console and open the Amazon Redshift console at https://console.aws.amazon.com/redshift/.
2. On the navigation menu, choose EDITOR, then connect to a database in your cluster.

3. On the Connect to database page, there are two ways to authenticate, namely, Temporary credentials and AWS Secrets Manager. Choose one from Create a new connection and Temporary credentials, then enter the values that you used when you created the cluster, then choose Connect.
4. For Schema, choose public to create a new table based on that schema.
5. Enter SQL queries in editor window and choose Run to create a new table. Choose Clear to remove current query.
6. Enter the new SQL command in the query editor window and choose Run to add rows to the table. Similarly, you can try other SQL queries.
7. Choose Execution to view the run details.
8. Choose Export to download the query results as a file. The supported file formats are CSV, TXT, and HTML.

➢ **Load sample data from Amazon S3**

1. At this point, you have a database and you are connected to it. Next, you create some tables in the database, upload data to the tables, and try a query.
2. For your convenience, the sample data that you load is available in an Amazon S3 bucket. If you're using a SQL client tool, ensure that your SQL client is connected to the cluster.
3. To try querying data in the query editor without loading your own data, choose Load sample data. If you do, Amazon Redshift loads its sample dataset to your Amazon Redshift cluster automatically during cluster creation.
4. Create tables- If you are using the Amazon Redshift query editor, individually copy and run the create table statements to create tables in your database.
5. Load sample data from Amazon S3 by using the COPY command. It is recommended to use the COPY command to load large datasets into Amazon Redshift from Amazon S3 or Amazon DynamoDB.
6. Provide authentication for your cluster to access Amazon S3 on your behalf to load the sample data. You can provide either role-based authentication or key-based authentication. It is recommended to use role-based authentication.
7. For this step, you provide authentication by referencing the IAM role that you created and then attached to your cluster in previous steps.
8. If you don't have proper permissions to access Amazon S3, you receive the following error message when running the COPY command: S3ServiceException: Access Denied.

✓ **To delete a cluster**

1. Sign in to the AWS Management Console and open the Amazon Redshift console at https://console.aws.amazon.com/redshift/.
2. On the navigation menu, choose CLUSTERS to display your list of clusters.
3. Choose your cluster. For Actions, choose Delete. The Delete cluster page appears.

4. Confirm the cluster to be deleted, then choose Delete cluster. On the cluster list page, the cluster status is updated as the cluster is deleted.