Adding Amazon Redshift Data Warehouses

An Amazon Redshift data warehouse is an Amazon Redshift cluster and database that contains the tables and views that you want to access as cube facts and dimensions. It also contains aggregate-table instances that either you or the AtScale engine creates in a schema that you specify.

Before You Begin

- Permissions

  Ensure that your user ID in the Design Center is assigned the Super User role, or is assigned the Manage Data Warehouses role permission.

- Aggregate schema

  Ensure that you know the schema to use for aggregate tables to be built in the data warehouse.
  AtScale reads and writes aggregate data to this schema. The AtScale service account user must have ALL privileges for this schema. BI tool user accounts should not have the select permission for this schema.

- SSL configuration

  For detailed information on Redshift SSL configuration options and certificates see Configuring security options for connections. AtScale routinely updates the CA certificates that it ships with, making it unlikely that you will have to update the AtScale JVM with the Amazon ACM certificates. However, if the certificate changes in the future before you can install another AtScale upgrade, follow the instructions below to import the new certificate.

- At-rest encryption

  To apply at-rest encryption to your data see the Amazon Redshift database encryption documentation.

- Known limitation

  Redshift doesn't support use of LN or LOG functions with NUMERIC or DECIMAL columns as described in LN function. If this limitation affects your dashboard design or AtScale cube, then please open a ticket with AWS Redshift.

Add The Data Warehouse

1. Choose Settings from the top navigation menu, select **Data Warehouses**.
2. Click on CREATE DATA WAREHOUSE.
3. Under **Type of Data Warehouse**, select **Redshift**.
4. Specify a name to identify your data warehouse. AtScale displays this name in the Design Center and uses it in log files.

5. (Optional) Enter the **External Connection ID** (formerly Query Name). The External Connection ID defaults to the data warehouse name. Override by entering a different name and enabling **Override generated value**.
6. In the **Database** field, specify the name of the database that appears in the **Cluster Database Properties** section of the overview page for the cluster.
7. Specify a schema that AtScale can use when creating aggregate tables. All types of aggregate tables will be created in this schema. See Types of Aggregate Table in AtScale for a list of the different types.

   You must create a schema; AtScale does not create a schema if one does not exist.

8. Specify the Data Warehouse as a **Read-only source**.

   - Select this option if AtScale will be configured to access this database with credentials that do not have permission to create new tables, or if you need to prevent AtScale from placing any aggregate tables in this data warehouse.
   - If **Read-only source** is enabled upon editing an existing data warehouse with aggregates present, all existing aggregates stored in this data warehouse will be deactivated.

9. **Optional:** In the section **Bucket Details**, specify the path of an S3 bucket and the region in which it is located.

   You must specify a bucket and region if you want to be able to load sample data from AtScale, use trigger files to initiate the build of aggregate tables for published cubes, or both.

   Also, you must specify an access key and secret key of your Amazon S3 bucket to read from and write to the bucket.

   Public visibility of the S3 bucket is not required, but atscale services must have access to it.

10. **Optional:** Click **Test s3 Connection** to check whether the connection parameters that you specified are valid.

    The name must be in the format `s3a://<name>`. For example, your bucket might be named `s3a://atscale`.

    If you choose to load sample data, AtScale stages the data in the bucket before loading it. There are about 70,000 rows of data. After the process of loading the sample data has finished, AtScale no longer needs access to the bucket, unless you plan to use trigger files to initiate the build of aggregate tables for published cubes.

Add A Redshift Connection

After setting up your data warehouse, you must add a SQL engine connection to the Redshift data warehouse before you can run queries. Expand the Redshift Data Warehouse, and select **Create Connection**.

1. Enter the unique name of the Redshift connection.
2. For **Host**, enter the Host for which your Redshift server resides.

3. If you are using a non-standard port number for connections to your Amazon Redshift database, specify the port number in the **Port** field. The port number that you are using appears in the **Cluster Database Properties** section of the overview page for the cluster. The default port number is 5439.
4. An IAM Role is needed for AtScale only for customers using Redshift. In that case we need the following Redshift privileges to be defined through the GRANT command:

   For Aggregate Schema: SELECT, UPDATE, CREATE, DROPFor Customer tables/schema: SELECT

   AtScale authenticates to Redshift using Password or Access Key and Secret key.

   Optional: if you want to be able to load sample data from AtScale, use trigger files to initiate the build of aggregate tables for published cubes, or both you must specify a bucket and region.

   Bucket name format: s3a://<name>

   In this case AtScale IAM role needs to have read/write access to this S3 bucket:

   ```
   {
      "Version": "2012-10-17",
      "Statement": [
         {
            "Sid": "ListObjectsInBucket",
            "Effect": "Allow",
            "Action": ["s3:ListBucket"],
            "Resource": ["arn:aws:s3:::bucket-name"]
         },
         {
            "Sid": "AllObjectActions",
            "Effect": "Allow",
            "Action": "s3:*Object",
            "Resource": ["arn:aws:s3:::bucket-name/*"]
         }
      ]
   }
   ```

   The only supported authentication for this particular use case is using access key and secret key.

5. Provide the Access Keys and Secret key, or the Password of the administrative user using the radio button. Note the following:

   - The access key and secret key are saved in the AtScale database:
     - Access key is saved in plain text format
     - Secret key is encrypted before being saved
     - If you configure AtScale to use password when connecting to your data warehouse, the password will be encrypted
   - If you have external secret manager enabled and configured, enter the corresponding credentials.

6. Enter the desired values for EXTRA JDBC FLAGS to cutomize the Redshift driver behavior. For example, see Configuring security options for connections for optional settings you may enter to customize SSL behavior.
7. Test and save the connection.

Connection Troubleshooting

If you have trouble making an SSL connection because the Amazon certificate chain has changed, then do the following:

1. Download the latest certificate from Amazon. For information see Configuring security options for connections
2. On each AtScale Server, load the certificate into the Java certificate store.

    1. Change to the following directory on the AtScale node:

        ```
        {atscale_install_directory}/current/pkg/jdk/lib/security/
        ```

    2. Import the certificate:

        ```
        ../../bin/keytool -import -trustcacerts -file /<path to your .cert file>/<filename>.cert -alias RSHIFT_CERT -keystore cacerts
        ```

    3. Enter the password from the Database.
    4. Execute configurator.sh with the `--apply` option to apply the new configuration as the Atscale installation user (user atscale in this example):

        ```
        su - atscale
        cd /opt/atscale/versions/<package_version>
        ./bin/configurator.sh --apply
        ```

3. Restart AtScale

    1. To stop all AtScale processes, run the following command on the AtScale node (if you have a single-instance installation of AtScale) or on all of the nodes in your AtScale cluster:

        ```
        /opt/atscale/bin/atscale_stop_apps
        ```

    2. To start all AtScale processes, run the following command where you ran the previous command:

        ```
        /opt/atscale/bin/atscale_start_apps
        ```