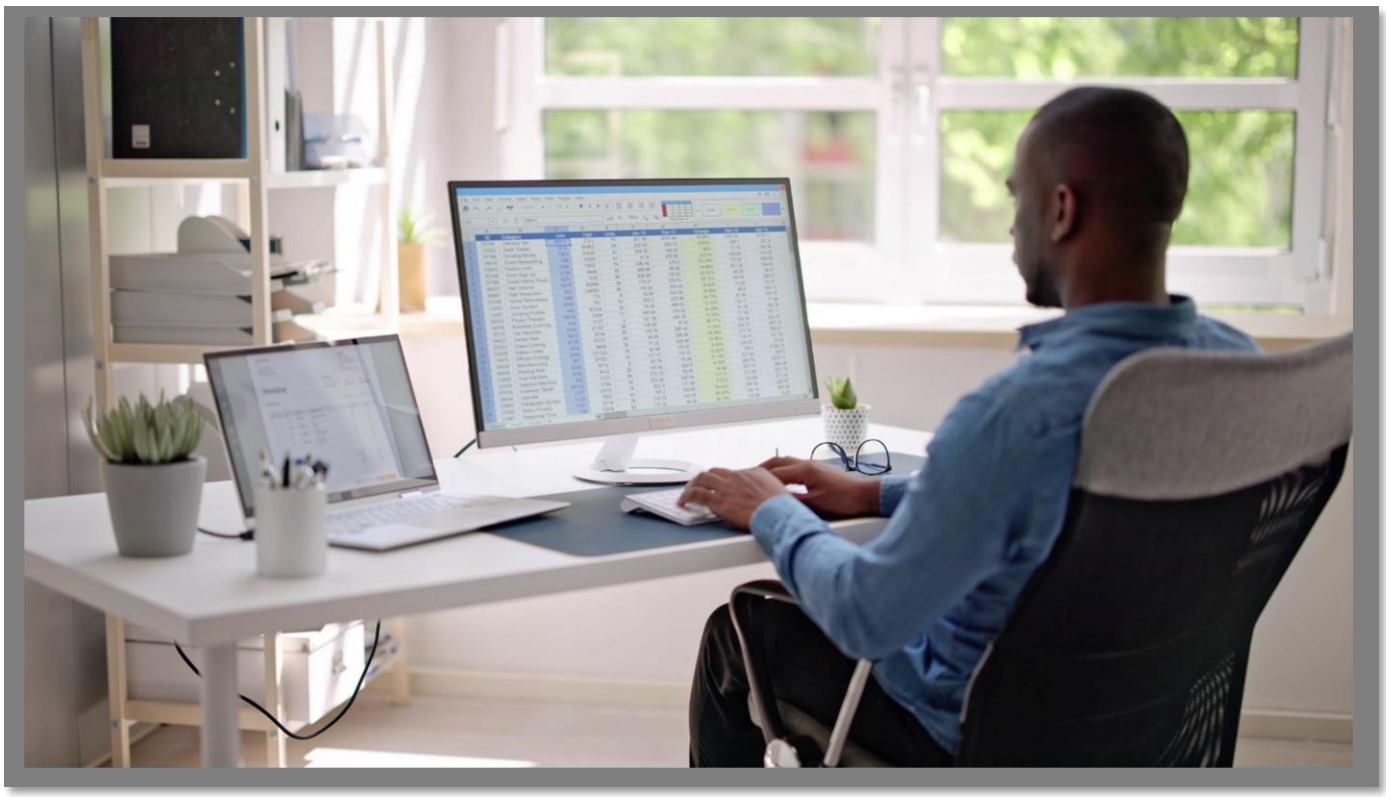
Arbutus Connectors

Amazon Redshift
CONFIGURATION GUIDE



ARBUTUS
Powerful Analytics Simplified

Contents

Arbutus Connector – Amazon Redshift

A. Introduction

The purpose of this Guide is to provide assistance with configuring the Arbutus Amazon Redshift Connector using the ODBC Data Source Administrator. The configuration process can involve several technical steps that require a good understanding of IT systems and database management.

To make the most of this guide, it's advisable to have a good understanding of database connectivity, driver installation, and system settings. The ODBC Data Source Administrator, which is used as part of the configuration process, allows for the setup and management of data sources, enabling applications to access data from various database systems.

Due to the complexity and potential impact of these configurations, it is recommended that only those individuals with IT or database expertise undertake this task. In addition, it should also be understood that each client's network environment is different. A one-size-fits-all approach is rarely effective, as what works well in one environment may not be suitable in another.

B. About Amazon Redshift

Amazon Redshift is a fully managed, cloud-based data warehouse service provided by Amazon Web Services (AWS). It is designed to handle large-scale data sets and perform complex queries and analytics efficiently. Amazon Redshift is ideal for businesses looking to perform high-performance data analysis and gain insights from their data. Amazon Redshift integrates with Amazon S3 and Amazon DynamoDB to facilitate seamless data transfer and analytics. However, each of them, including Amazon Redshift, serve distinct purposes within AWS.

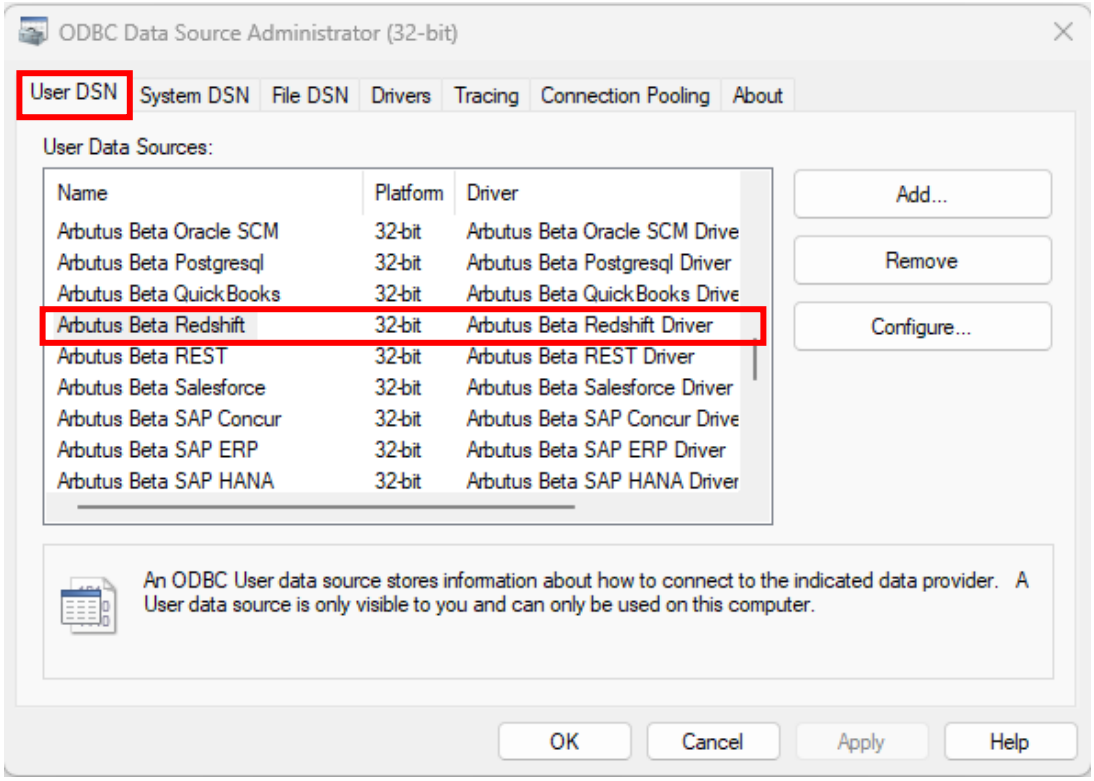
Unlike row-based databases, Redshift stores data column-wise, which improves query performance and compression. This makes aggregations, scans, and complex queries faster compared to row-based storage.

C. Determining if the Connector exists prior to configuring

Installation of the Arbutus Amazon Redshift Connector is done at the time of installing the Arbutus software. For more information on this, please see the **Overview Guide Document**.

Once the Connector has been installed, the next step is to configure it.

Prior to configuring it, you can check to see if the Connector has been installed by opening the **32-bit ODBC Data Source Administrator**, pictured below, and clicking the **User DSN** tab. Included below is information on how you can access the **ODBC Data Source Administrator**.



- If the Arbutus Amazon Redshift Connector appears in the list, it can be considered as installed.
- If it is not listed, it is likely that you did not select it during the installation or modification of the Arbutus software. In this case, it is recommended to reinstall the Arbutus software and choose the **Modify** option when prompted. For more details, please refer to the **Overview Guide Document**.

Below is the file path to access and run the ODBC Data Source Administrator application:

C:\Windows\SysWOW64\odbcad32.exe

Alternative, you can also try locating and opening the **ODBC Data Source Administrator** application by doing a search on your desktop application.

D. Configuring the Connector after it has been installed

Once you have verified that the Arbutus Connector has been installed, it is time to configure it.

This process is done using the **ODBC Data Source Administrator**. It can be described as "**editing the DSN configuration"**.

DSN, Drivers, and Data Sources
What is a DSN? DSN stands for Data Source Name, and is a unique name used to create a data connection to a database using open database connectivity (ODBC). A DSN is a data structure that contains the information required to connect to a database. It is essentially a string that identifies the source database, including the driver details, the database name, and often authentication credentials and other necessary connection parameters. DSNs facilitate a standardized method for applications to access databases without needing hard-coded connection details, enhancing flexibility and scalability in database management.

Arbutus Connectors

- ***Drivers*** are the components that process ODBC requests and return data to the application. If necessary, drivers modify an application's request into a form that is understood by the data source. The **Drivers** tab in the **ODBC Data Source Administrator** dialog box lists all drivers installed on your computer, including the name, version, company, file name, and file creation date of each driver.

- ***Data sources*** are the databases of files accessed by a driver and are identified by a data source name (DSN). You use the ODBC Data Source Administrator to add, configure, and delete data sources from your system.

All ODBC connections require that a DSN be configured to support the connection. When a client application wants to access an ODBC-compliant database, it references the database using the DSN.

The types of DSNs are:

- **User DSN** – User DSNs are local to a computer and can be used only by the current user. They are registered in the HKEY_Current_USER registry subtree.

- **System DSN** – System DSNs are local to a computer rather than dedicated to a user. The system or any user with privileges can use a data source set up with a system DSN. System DSNs are registered in the HKEY_LOCAL_MACHINE registry subtree.

- **File DSN** – File DSNs are file-based sources that can be shared among all users who have the same drivers installed and therefore have access to the database. These data sources need not be dedicated to a user nor be local to a computer. File data source names are identified by a file name with a .dsn extension.

User and system data sources are collectively known as *machine* data sources because they are local to a computer.

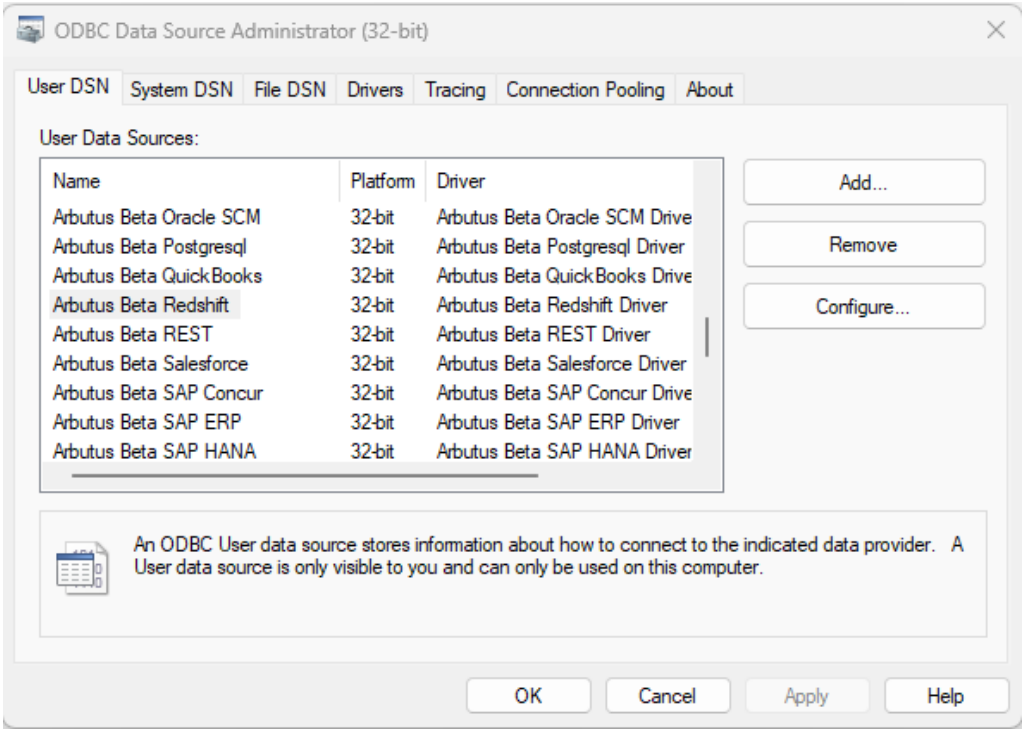
Each of these DSNs has a tab in the **ODBC Data Source Administrator** dialog.

The Arbutus ODBC Driver for Amazon Redshift enables real-time access to Amazon Redshift data, directly from any applications that support ODBC connectivity, the most widely supported interface for connecting applications with data.

Arbutus Connectors

Follow these steps to edit the DSN configuration and configure the Connector.

1. First open the **ODBC Data Source Administrator**.



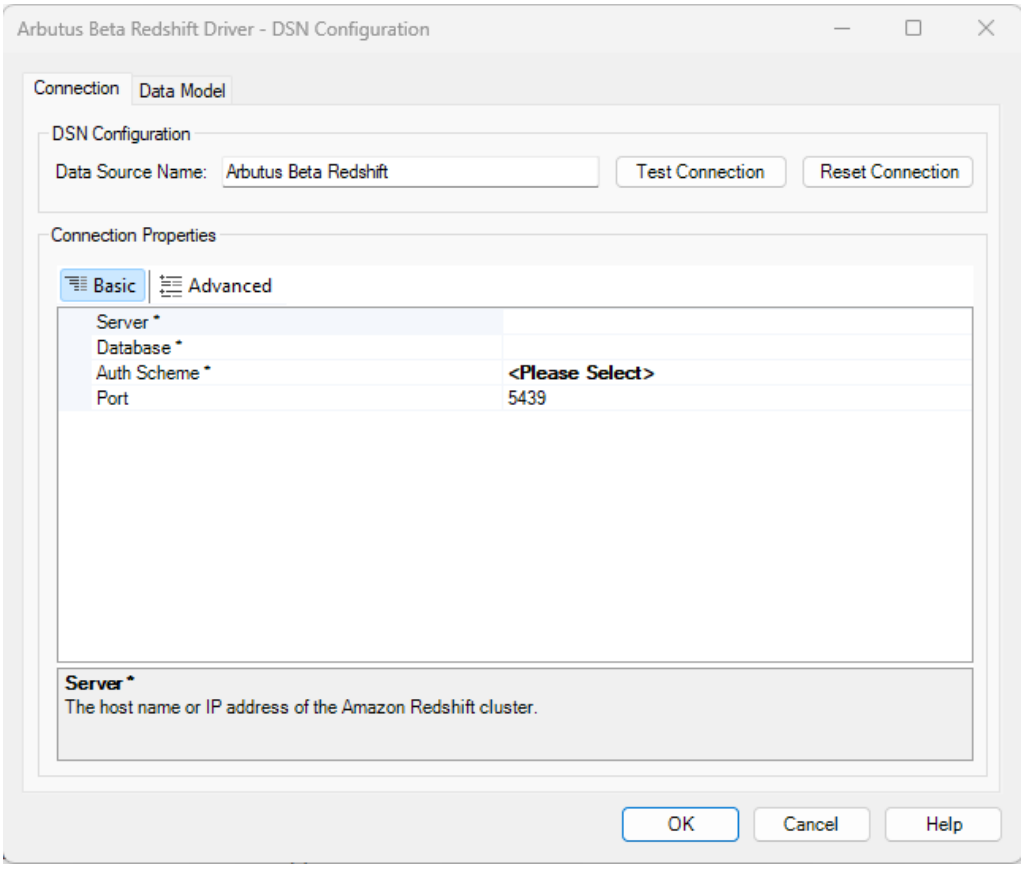
2. Click the **User DSN** tab.

   Selected data connectors are installed as **User DSN's** in Window's 32 Bit ODBC Data Source Administrator.

   Also, each of the data connector's names are prefaced with Arbutus, for example, **Arbutus Redshift.**

3. Select the Arbutus Connector, in this case it is **Arbutus Redshift**.
4. Click **Configure**.

This opens the **Arbutus Amazon Redshift Driver – DSN Configuration** dialog.



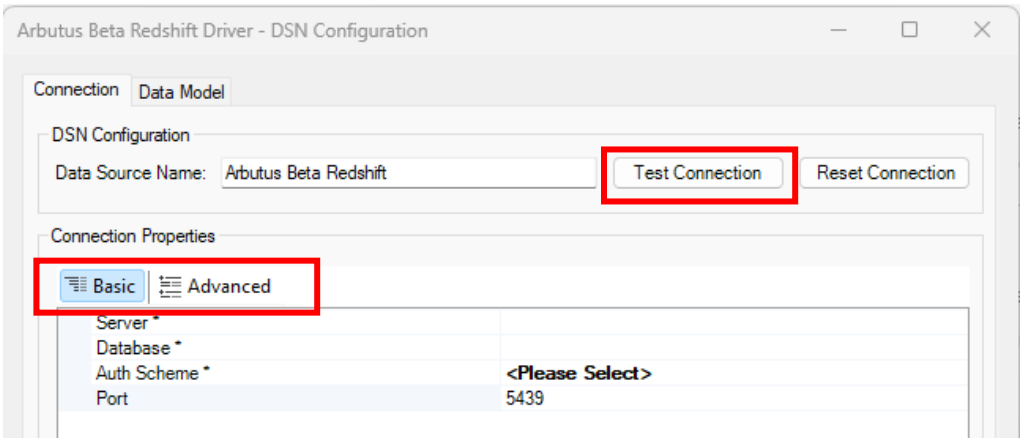
E. Editing the DSN properties – the Basic and Advanced tabs

With the DSN Configuration dialog open, the next step is to edit the DSN properties, where necessary, in the **Basic** and **Advanced** tabs. For example, editing the **Port** property (per screenshot below) to match the port number of the Amazon Redshift server.

E1. Editing the DSN properties in the Basic tab

The properties listed in the **Basic** tab are typically the ones that are most commonly used, and as such are designed to be more user-friendly and straightforward, allowing you to quickly make changes without needing in-depth technical knowledge.

Once you have completed editing the properties in the **Basic** tab, you can go ahead and try testing the connection to the Amazon Redshift system by clicking the **Test Connection** button, as highlighted in the screenshot below.



In the **Basic** tab, there are **four** main properties to review:

1. **Server** – this is the host name or IP address of the Amazon Redshift cluster.

2. **Database** – this is the name of the Amazon Redshift database.

3. **Auth Scheme** – click the dropdown to select from the list the appropriate scheme used for authentication. The options available for selection are as follows:
   - Basic – select this when you want to use standard authentication with your Amazon Redshift username and password.

     This setup allows the driver to authenticate using the credentials directly managed by Amazon Redshift, ensuring a simple and direct connection.

Selecting **Basic** requires you to specify the following:

- User – this is the Amazon Redshift user account used to authenticate. Together with **Password** (see below), this field is used to authenticate against the Amazon Redshift server.

- Password – this is the password used to authenticate the user. The **User** (see above) and **Password** are together used to authenticate with the server.

- ADFS – select this when you need to use Active Directory Federation Services (ADFS) for authentication.

  This setup allows the driver to authenticate using the credentials managed by ADFS, ensuring secure and streamlined access to Amazon Redshift.

  Selecting **ADFS** requires you to specify the following:

  - User – for more information, please see this same property listed and described above (under the **Basic** section).

  - Password - for more information, please see this same property listed and described above (under the **Basic** section).

  - SSO Login URL – this is the identity provider's login URL.

  - SSO Properties – additional properties required to connect to the identity provider, formatted in a semicolon-separated list. This is used with **SSO Login URL** (see above).

- IAMCredentials – select this when you want to use AWS Identity and Access Management (IAM) for authentication. IAM credentials provide a secure way to manage access to your

Amazon Redshift cluster, leveraging AWS's robust security features.

This setup allows the driver to authenticate using IAM credentials, ensuring secure and managed access to Amazon Redshift.

Selecting **IAMCredentials** requires you to specify the following:

- User - for more information, please see this same property listed and described above (under the **Basic** section).

- AWS Access Key – this is your AWS account access key or the access key for an authorized Identity Access Management (IAM) user.

  To authorize Amazon Redshift requests, provide the credentials for an administrator account or for an IAM user with custom permissions. Set this property along with **AWS Secret Key**.

  Although you can connect as the AWS account administrator, it is recommended to use IAM user credentials to access AWS services.

  To obtain the credentials for an IAM user, follow the steps below:

  a. Sign into the IAM console.
  b. In the navigation pane, select **Users**.
  c. To create or manage the access keys for a user, select the user and then select the **Security Credentials** tab.

  To obtain the credentials for your AWS root account, follow the steps below:

  a. Sign into the AWS Management console with the credentials for your root account.
  b. Select your account name or number and select **My Security Credentials** in the menu that is displayed.

c. Click **Continue to Security Credentials** and expand the Access Keys section to manage or create root account access keys.

- AWS Secret Key – this is your AWS account secret key or the secret key for an authorized Identity Access Management (IAM) user.

  See **AWS Access Key** (above) to obtain the secret key and access key.

- PingPederate – select this when you need to use PingFederate for authentication.

  Note:
  PingFederate is a highly regarded enterprise federation server that specializes in user authentication and providing standardized single sign-on (SSO) solutions.

  As an example, PingFederate is typically useful when you want to enable single sign-on (SSO) capabilities, allowing users to authenticate once and gain access to multiple systems without needing to log back in again.

  This setup allows the driver to authenticate using the credentials managed by PingFederate, ensuring secure and streamlined access to Amazon Redshift.

  Selecting **PingFederate** requires you to specify the following:
  - User – for more information, please see this same property listed and described above (under the **Basic** section).
  - Password – for more information, please see this same property listed and described above (under the **Basic** section).

- SSO Login URL – for more information, please see this same property listed and described above (under the **ADFS** section).

- SSO Properties – for more information, please see this same property listed and described above (under the **ADFS** section).

- SSO Exchange URL – this is the URL used for consuming the SAML response and exchanging it for service specific credentials.

  ODBC Driver for Amazon Redshift will use the URL specified here to consume a SAML response and exchange it for service specific credentials. The retrieved credentials are the final piece during the SSO connection that are used to communicate with Amazon Redshift.

- AWS Principal ARN – this is the ARN of the SAML Identity provider in your AWS account.

  ARN = ***A***mazon ***R***esource ***N***ame. This is a unique identifier used to specify AWS resources. It identifies the specific AWS entity (such as an IAM user, role, or service) that is allowed or denied access to a resource.

  SAML = ***S***ingle ***A***ssertion ***M***arkup ***L***anguage assertion – a SAML assertion is the message that tells a service provider that a user is signed in.

- AzureAD – select this when you need to use Microsoft Azure Active Directory (Azure AD) for authentication.

As an example, AzureAD is typically useful when you want to enable single sign-on (SSO) capabilities, allowing users to authenticate once and gain access to multiple systems without needing to log back in again.

This setup allows the driver to authenticate using the credentials managed by Azure AD, ensuring secure and streamlined access to Amazon Redshift.

Selecting **AzureAD** requires you to specify the following:

- User – for more information, please see this same property listed and described above (under the **Basic** section).

- Azure Tenant – this is the Microsoft Online tenant being used to access data. For instance, contoso.onmicrosoft.com. If not specified, your default tenant is used. Alternatively, specify the tenant Id. This value is the directory Id in the Azure Portal > Azure Active Directory > Properties

  Typically it is not necessary to specify the Tenant. This can be automatically determined by Microsoft when using the **OAuth Grant Type** set to **CODE** (default). However, it may fail in the case that the user belongs to multiple tenants. For instance, if an Admin of domain A invites a user of domain B to be a guest user. The user will now belong to both tenants. It is a good practice to specify the Tenant, although in general things should normally work without having to specify it.

  The **Azure Tenant** is required when setting **OAuth Grant Type** to CLIENT. When using client credentials, there is no user context. The credentials are taken from the context of the app itself. While Microsoft still allows client credentials to be obtained without specifying which Tenant, it has a much lower probability of picking the specific tenant you want to work with. For this reason, we require **Azure**

Tenant to be explicitly stated for all client credentials connections to ensure you get credentials that are applicable for the domain you intend to connect to.

- SSO Login URL – for more information, please see this same property listed and described above (under the **ADFS** section).

- OAuth Client Id – this is the client Id assigned when you register your application with an OAuth authorization server.

  As part of registering an OAuth application, you will receive the **OAuth Client Id** value, sometimes also called a consumer key, and a client secret, the **OAuth Client Secret** (see below)

- OAuth Client Secret – this is the client secret assigned when you register your application with an OAuth authorization server.

  As part of registering an OAuth application, you will receive the **OAuth Client Id** (see above), also called a consumer key. You will also receive a client secret, also called a consumer secret. Set the client secret in the **OAuth Client Secret** property.

- Callback URL – this is the OAuth callback URL to return to when authenticating. This value must match the callback URL you specify in your app settings.

  During the authentication process, the OAuth authorization server redirects the user to this URL. This value must match the callback URL you specify in your app settings.

- Scope – specify scope to obtain the initial access and refresh token.

- AzureADPKCE – select this when you need to use Microsoft Azure Active Directory (Azure AD) with Proof Key for Code Exchange (PKCE) for authentication.

  PKCE adds an extra layer of security to the OAuth 2.0 authorization code flow, preventing certain types of attacks, such as authorization code interception attacks.

  This setup allows the driver to authenticate using Azure AD with PKCE, ensuring secure and streamlined access to Amazon Redshift.

  Selecting **AzureADPKCE** requires you to specify the following:
  - OAuth Client Id – for more information, please see this same property listed and described above (under the **AzureAD** section).

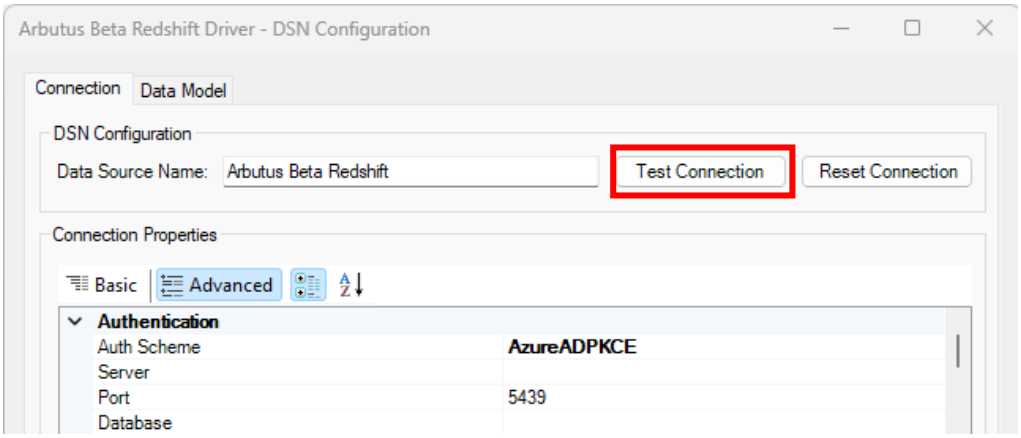
  - Scope – specify scope to obtain the initial access and refresh token.

4. **Port** – this is the port number of the Amazon Redshift server. If not specified, the default port number **5439** is used.

E2. Editing the DSN properties in the Advanced tab

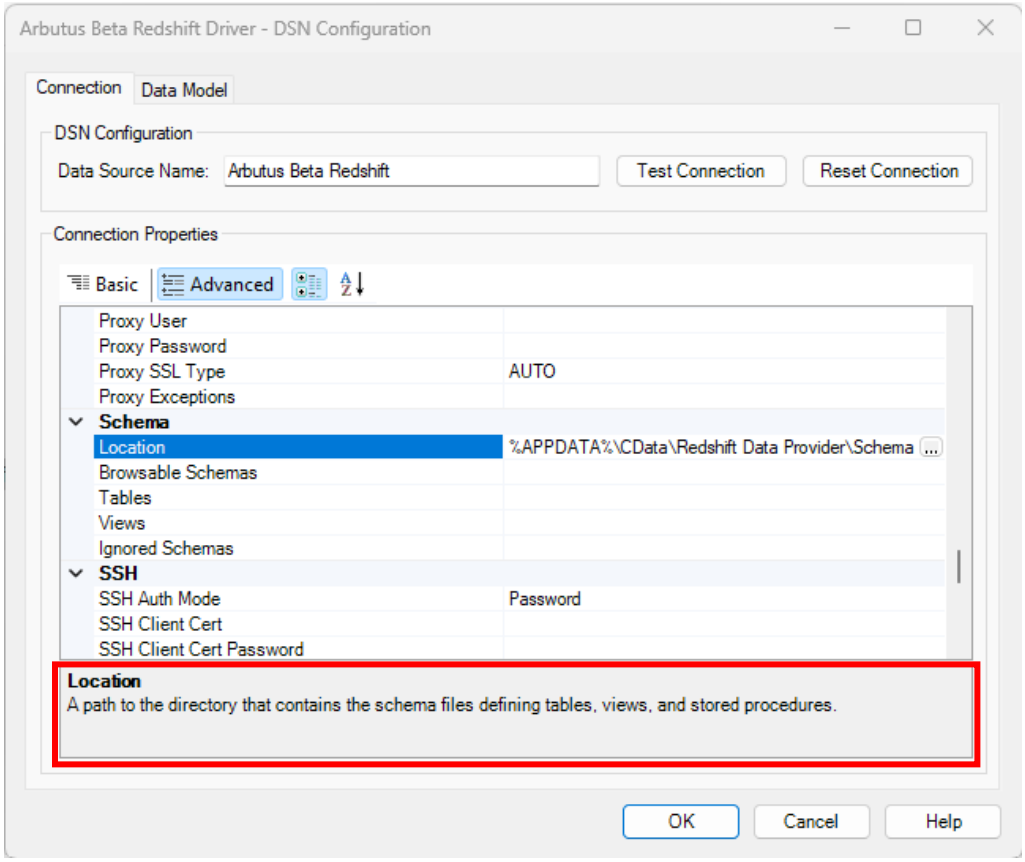
This tab includes more detailed and technical properties. It is intended for those users who need more control over the configuration and are comfortable with more complex options. The **Advanced** tab often includes properties that can fine-tune the behaviour of the system feature.

If you have already completed editing the properties in the **Basic** tab, as required, you do not necessarily need to also complete editing the properties in the **Advanced** tab. Instead, once you have completed editing the properties in the Basic tab, you may opt to proceed to testing the connection to the Amazon Redshift system by clicking the **Test Connection** button.



There are a lot more properties included for editing in the **Advanced** tab.

However, it is useful to know that each property does provide a short description of it and as such serves as a guide in terms of what to edit and/or enter. This short description can be seen at the bottom of the **DSN Configuration** dialog box, as seen in the screenshot below.



If it is deemed necessary to complete some/all the properties in the Advanced tab, it is recommended that you refer to the description shown for any of the properties being edited and/or entered.

If required, more information on the properties listed in the Advanced tab can also be provided.

F. Other questions and/or request for assistance

There may be times when you need to consult with the technical support team at Arbutus Software. If so, please send an email request to support@ArbutusSoftware.com.

For more information, please refer to the CONTACT US page on our website.