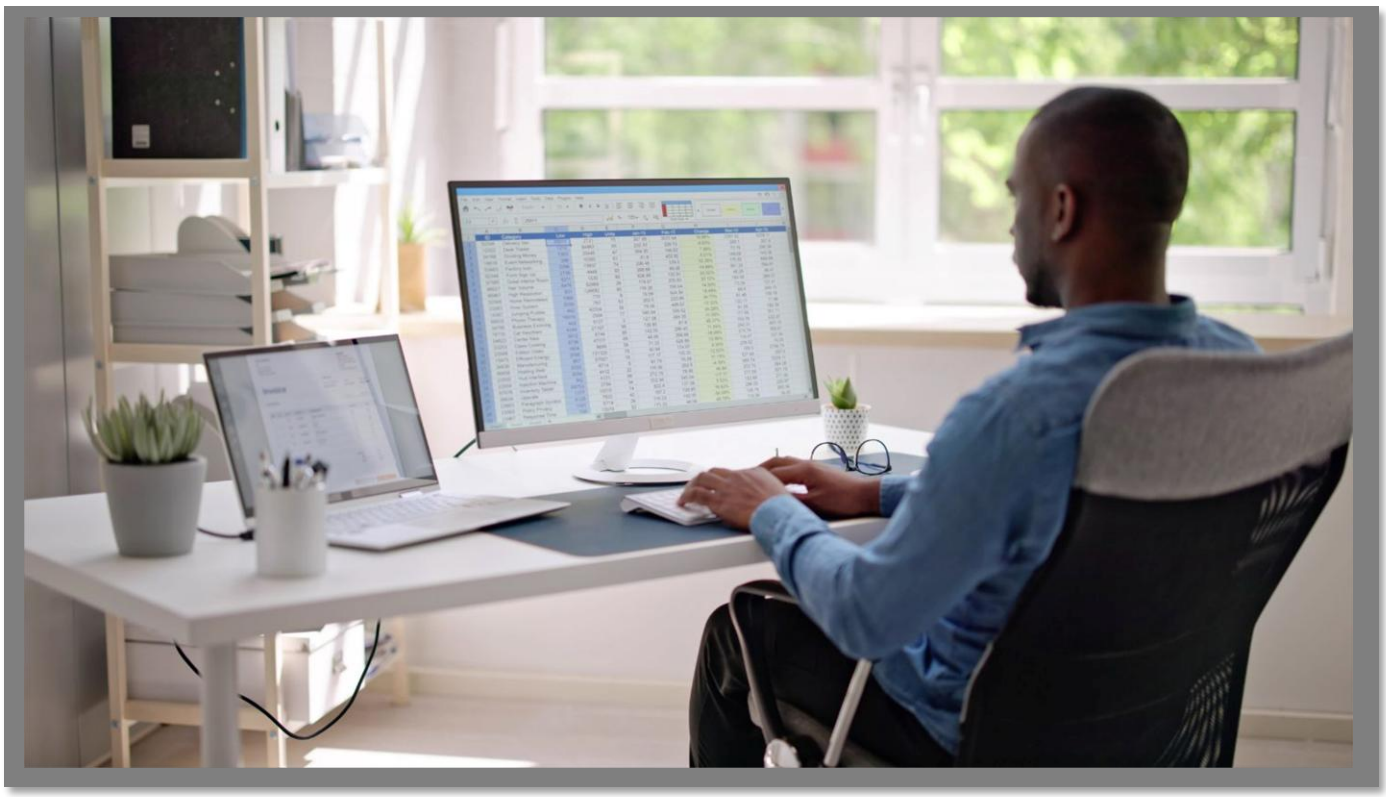
Arbutus Connectors

Apache Hbase
CONFIGURATION GUIDE



ARBUTUS
Powerful Analytics Simplified

Contents

Arbutus Connector – Apache Hbase

A. Introduction

The purpose of this Guide is to provide assistance with configuring the Arbutus Apache Hbase 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 Apache Hbase

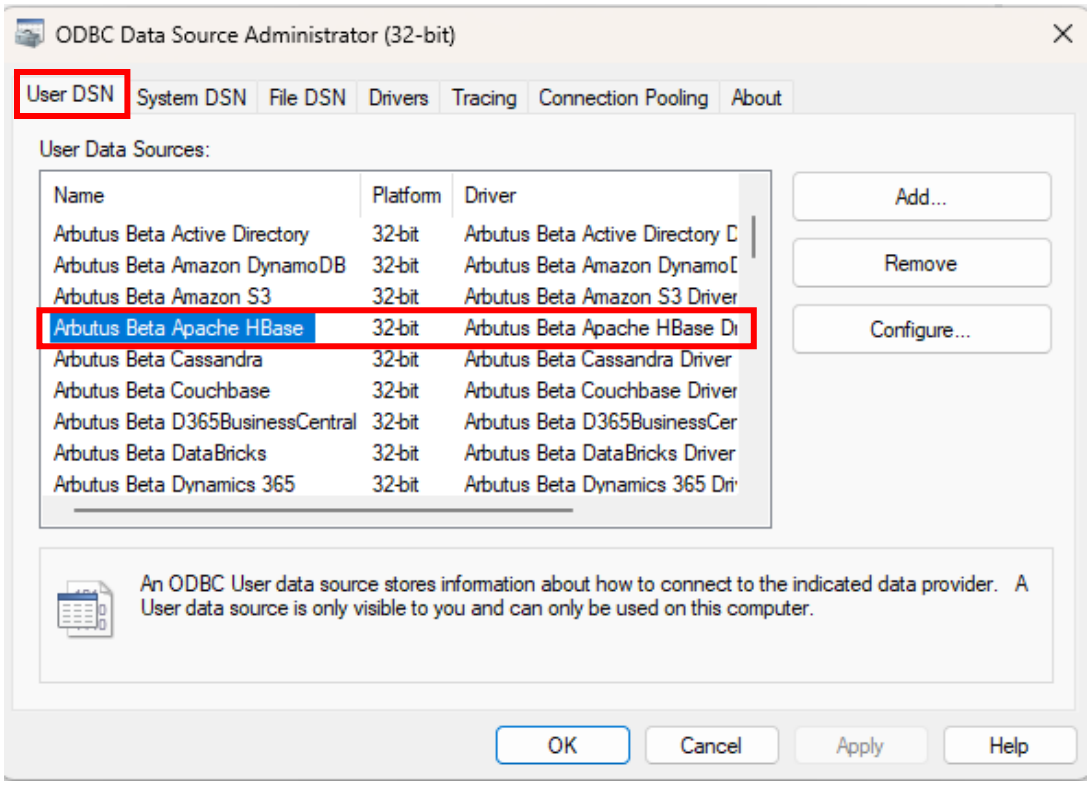
Apache Hbase is a distributed, scalable, NoSQL database built on Apache Hadoop. It provides real-time read/write access to large amounts of structured data and is designed to handle billions of rows and millions of columns. Hbase is modelled after Google Bigtable and runs on HDFS (Hadoop Distributed File System is the storage system used by Apache Hadoop), making it ideal for big data applications that require fast lookups and random access. It supports integration with Apache Spark and Apache Hive. Apache Hbase stores data in a distributed, column-oriented format on top of HDFS. Data is organized into tables, rows, and columns, similar to relational databases. Each row has a unique row key, which is used for fast lookups.

C. Determining if the Connector exists prior to configuring

Installation of the Arbutus Apache Hbase 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 Apache Hbase 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.

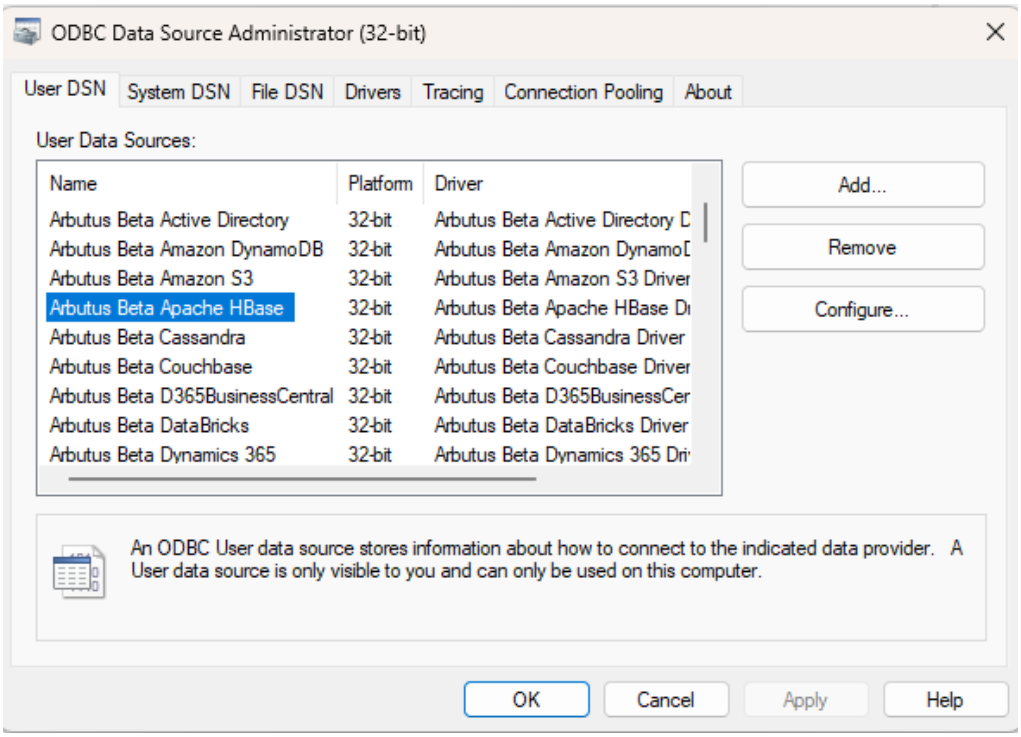
Each of these DSNs has a tab in the **ODBC Data Source Administrator** dialog.

The Arbutus ODBC Driver for Apache Hbase enables real-time access to Apache Hbase 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**.



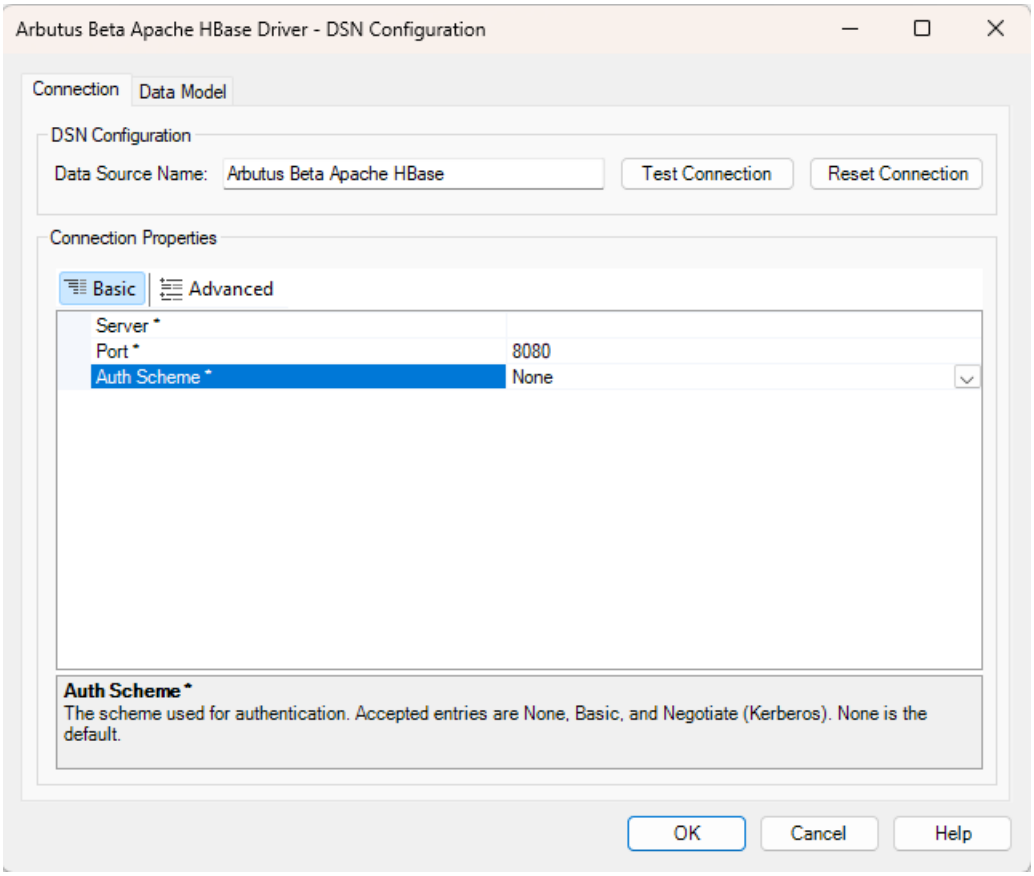
2. Click the **User DSN** tab.

   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 is prefaced with Arbutus, for example, **Arbutus Apache Hbase.**

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

This opens the **Arbutus Apache Hbase Driver – DSN Configuration** dialog.



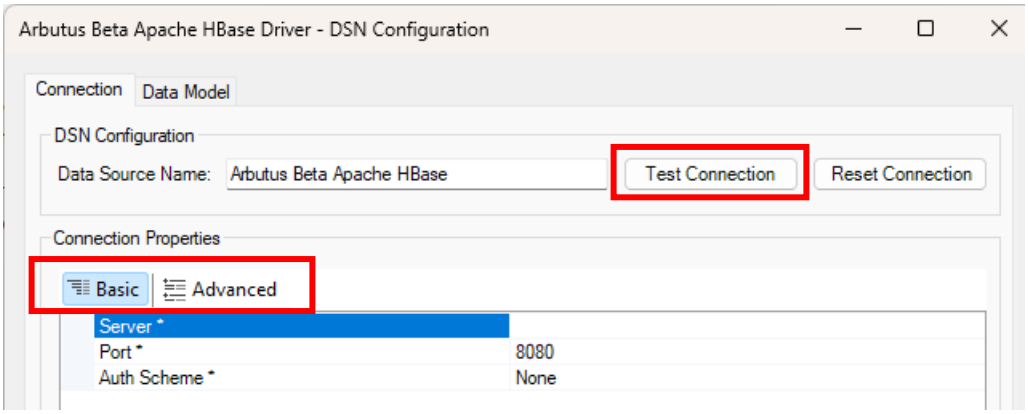
E. Editing the DSN properties – the Basic and Advanced tabs

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 **Auth Scheme properties** (per screenshot below) to ensure correct authentication to the server is applied.

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 Apache Hbase system by clicking the **Test Connection** button, as highlighted in the screenshot below.



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

1. **Server** – this is the host name, IP address, or URL of the Apache Hbase REST server. To use SSL, you can prefix the host name or IP address with *'https://'*.

2. **Port** – this is the port for the Apache Hbase REST (Stargate) server.

   The default value is **8080**.

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:
   - None – select this when you need a straightforward setup without any authentication requirements. For example, if you are in a development or testing environment where security is not a primary concern.

     It is important to note that using **None** means no authentication is performed, which can be a security risk in production environments.

     The default value is **None**.

   - Basic – select this when you need a straightforward method to authenticate users using a username and password.

     It is important to note that Basic authentication sends credentials in an unencrypted form, which can be a security risk in production environments. For better security, consider using more secure authentication methods like Kerberos or SSL/TLS.

     If your Hbase instance is not secured with Kerberos and does not require token-based authentication, or if your Hbase cluster is running without strict policies, **Basic** is a straightforward choice. It uses a username and password to authenticate the connection.

     However, if your organization is using Single Sign-On (SSO) or SPNEGO, then in such cases authentication should be handled dynamically using **Negotiate** (see below) instead of requiring manual credential entry.

- Negotiate – select this when you want to use Kerberos for secure authentication. The Negotiate scheme will automatically handle the negotiation process to determine the best authentication mechanism, typically Kerberos.

  Negotiate is ideal for environments where security and compliance are critical.

  Additionally:
  - If your Hbase cluster is secured with Kerberos, selecting **Negotiate** allows the driver to automatically determine and use Kerberos for authentication.
  - If your Hbase is configured to use SPNEGO, e.g., for web-based authentication via HTTP, **Negotiate** will facilitate proper authentication using available credentials.

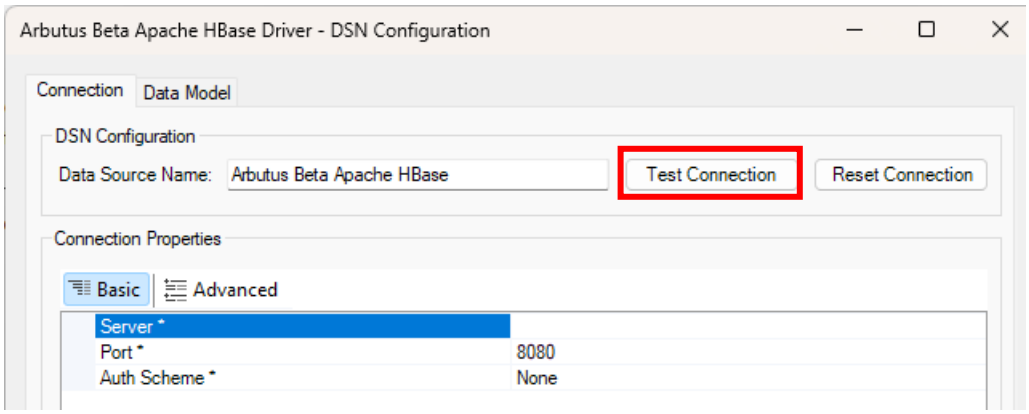
    Note:
    *SPNEGO (**S**imple and **P**rotected GSS-API **Nego**tiation Mechanism) in the context of Apache Hbase ODBC authentication is a protocol used for negotiating authentication mechanisms, typically when Kerberos authentication is involved. SPNEGO is often used to enable Single Sign-On (SSO) for web-based or API-based services. When an ODBC driver is configured with **Negotiate** authentication, it may use SPNEGO to determine whether to use Kerberos or other authentication methods automatically.*

  - If your organization used both Kerberos and other authentication mechanisms, **Negotiate** allows the driver to choose the most appropriate one dynamically.
  - If users log in to their workstations using domain credentials (Active Directory), selecting **Negotiate** may enable seamless authentication without requiring manual credential entry.

E2. Editing the DSN properties in the Advanced tab

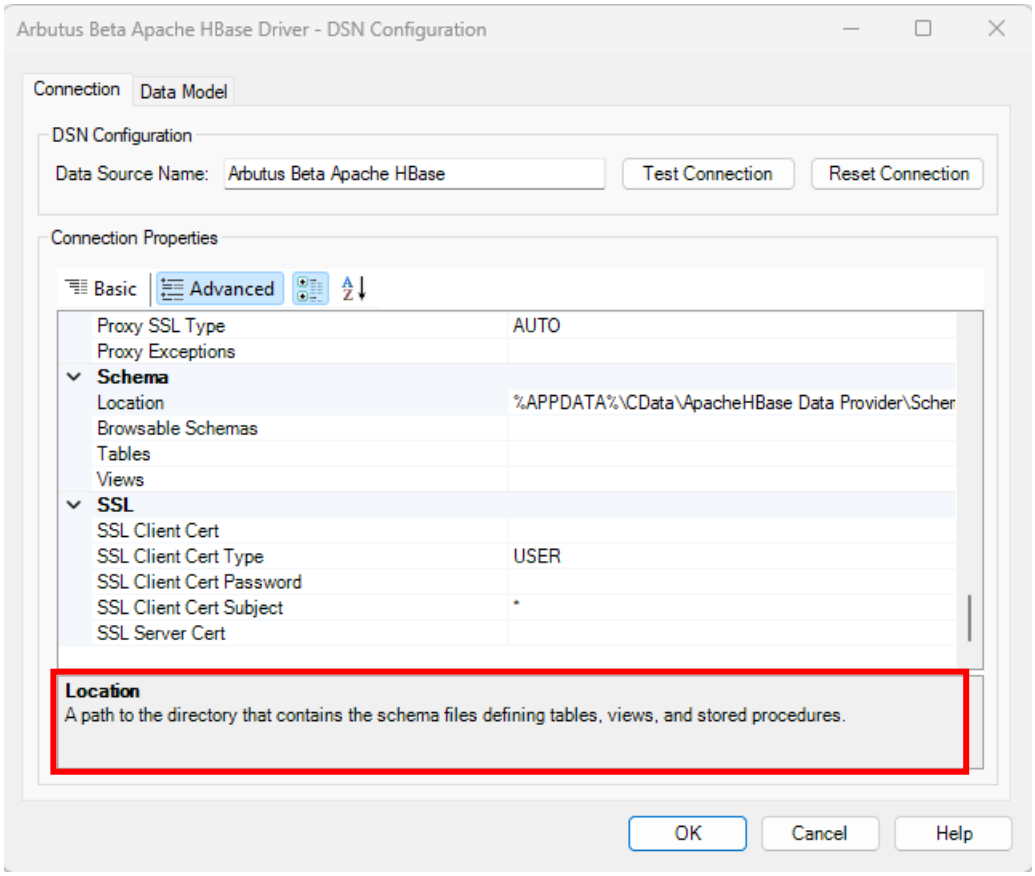
This tab includes more detailed and technical properties. It is intended for those users who need more control over the configuration and are comfortable with more complex options. The **Advanced** tab often includes properties that can fine-tune the behaviour of the system feature.

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 Apache Hbase 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.