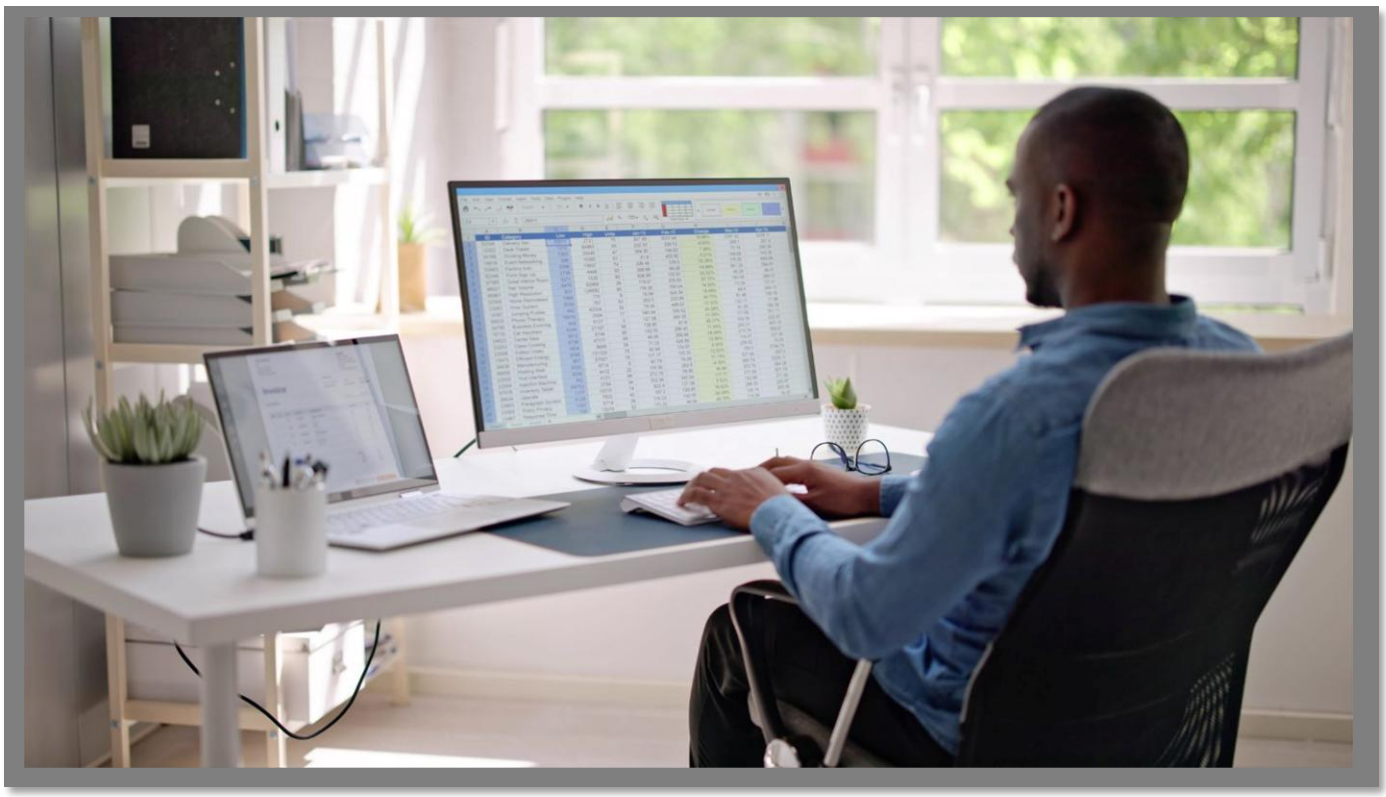
Arbutus Connectors

Cassandra

CONFIGURATION GUIDE



ARBUTUS

Powerful Analytics Simplified

Contents

Arbutus Connector – Cassandra

A. Introduction

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

Apache Cassandra is a highly scalable, distributed NoSQL database designed to handle large amounts of data across many commodity servers without a single point of failure. It offers high availability, fault tolerance, and excellent performance for read and write operations, making it ideal for big data applications. It uses a per-to-peer architecture, offers high availability, and supports flexible schema design.

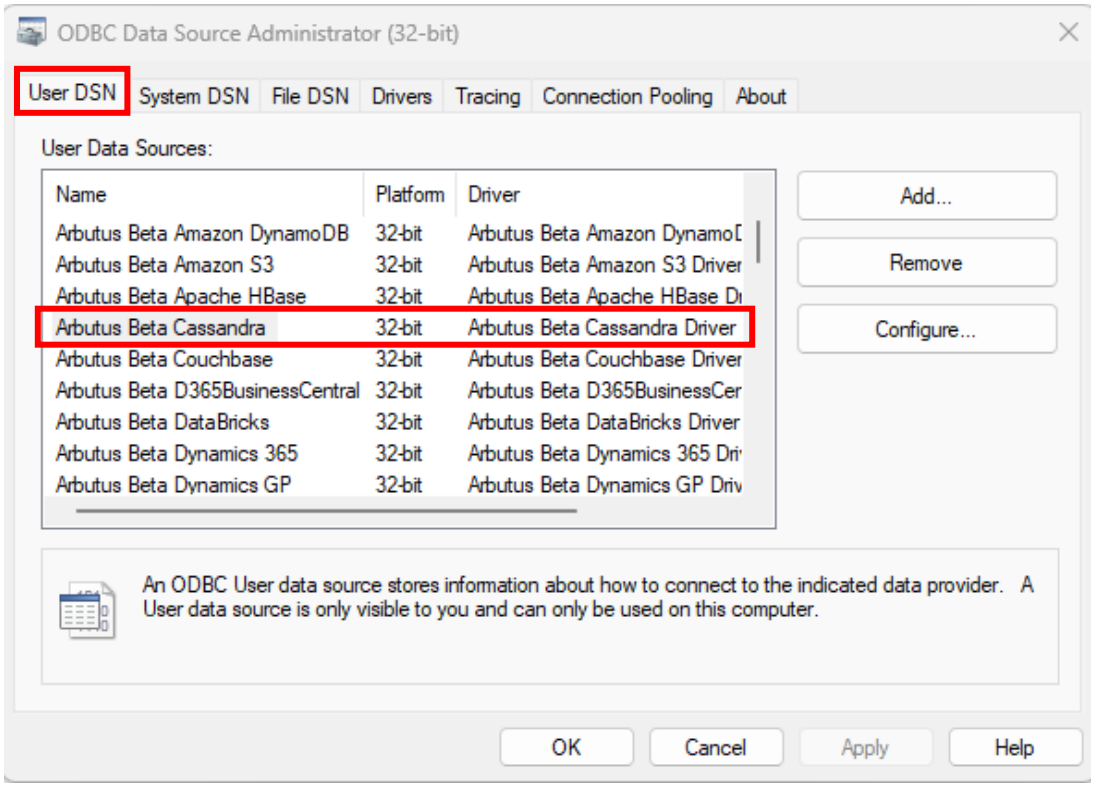
In Apache Cassandra, data is primarily stored on disk in **SSTables** (Sorted String Tables). These are immutable files (that cannot be modified after they are created) that hold the actual row data. With immutable files, a new file is created with the updated data, and the old file eventually gets replaced or compacted.

C. Determining if the Connector exists prior to configuring

Installation of the Arbutus Cassandra 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 Cassandra 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.

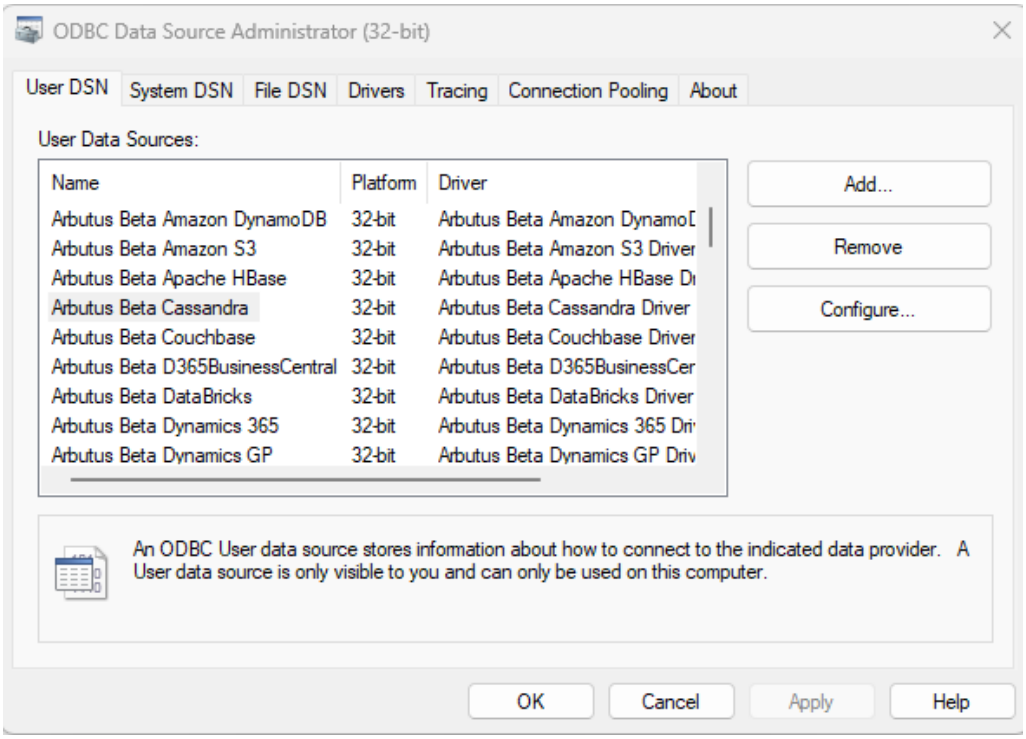
Each of these DSNs has a tab in the **ODBC Data Source Administrator** dialog.

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



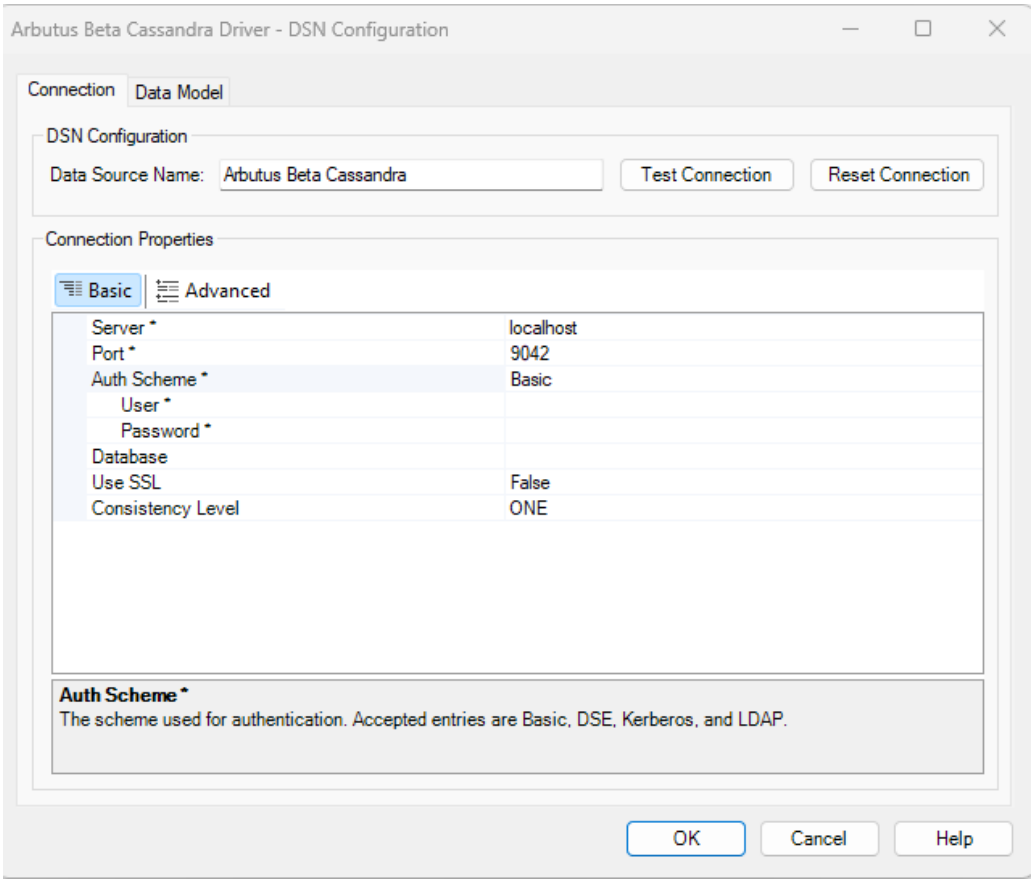
2. Click the **User DSN** tab.

   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 Cassandra.**

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

This opens the **Arbutus Cassandra Driver – DSN Configuration** dialog.



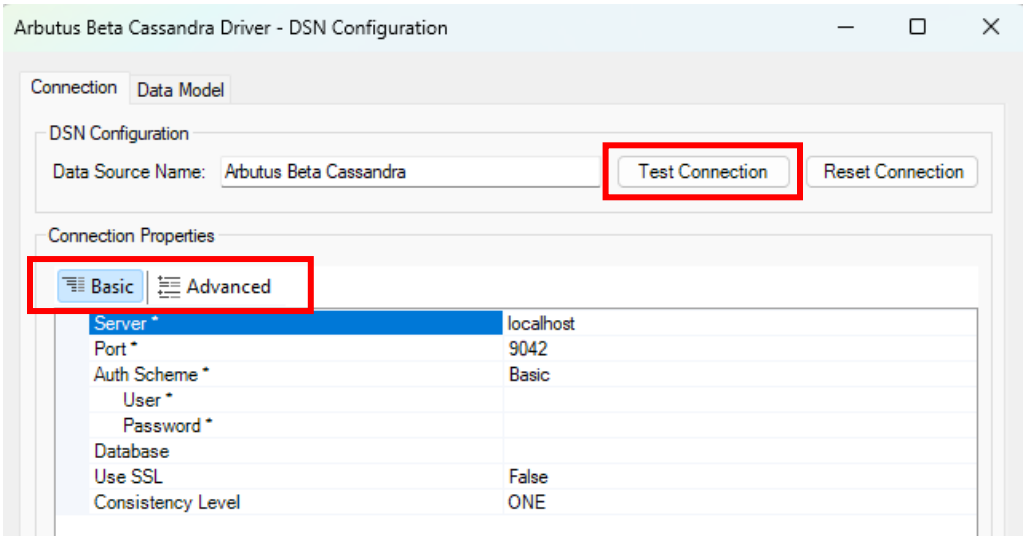
E. Editing the DSN properties – the Basic and Advanced tabs

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



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

1. **Server** – this is the host name or IP address of the server hosting the Cassandra database. To connect to a distributed system, you can set **Server** to a comma-separated list of servers and ports, separated by colons. You will also need to set **Consistency Level** (see below). Note that you must specify all of the servers required by your selected consistency level.

2. **Port** – this is the port for the Cassandra database.

   The default value is **9042**.

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 if you are using Basic Authentication. Set this to authenticate with login credentials and Cassandra's built-in authentication. This method involves using a username and password to authenticate your connection.

     However, it is important to note that Basic Authentication is less secure compared to other methods, as it involves sending credentials over the network. For production environments or integrations requiring higher security, more advanced authentication methods are generally recommended

     Selecting **Basic** requires you to specify the User and Password.
     - User – this is the Cassandra user account used to authenticate. The authenticating server requires both **User** and **Password** (see below) to validate the user's identity.

     - Password – this is the password used to authenticate the user. The authenticating server requires both **User** (see above) and **Password** to validate the user's identity.

     The default value is **Basic**.

   - DSE – select this if you are using DataStax Enterprise (DSE). Set this to authenticate with login credentials and the DSE Unified Authenticator. DSE provides enhanced security features and additional capabilities compared to the open-source version of Cassandra.

     DSE offers advanced security features, including role-based access control, encryption, and auditing. DSE authentication integrates seamlessly with enterprise security systems, such as LDAP and Kerberos, providing a unified security framework.

Selecting **DSE** requires you to specify the User and Password.

- User – this is the Cassandra user account used to authenticate. The authenticating server requires both **User** and **Password** (see below) to validate the user's identity.

- Password – this is the password used to authenticate the user. The authenticating server requires both **User** (see above) and **Password** to validate the user's identity.

- Kerberos – select this if you need strong authentication and single sign-on capabilities. Set this to use Kerberos to authenticate .

  Kerberos provides robust security features, including mutual authentication, which ensures that both the client and server verify each other's identity.

  Kerberos supports SSO, allowing users to authenticate once and gain access to multiple services without needing to re-enter credentials.

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

  - User – this is the Cassandra user account used to authenticate. The authenticating server requires both **User** and **Password** (see below) to validate the user's identity.

  - Password – this is the password used to authenticate the user. The authenticating server requires both **User** (see above) and **Password** to validate the user's identity.

- Kerberos KDC – this is the Kerberos Key Distribution Center (KDC) service used to authenticate the user.

  The Kerberos properties are used when using SPNEGO or Windows Authentication. The driver will request session tickets and temporary session keys from the Kerberos KDC service. The Kerberos KDC service is conventionally collocated with the domain controller.

  If Kerberos KDC is not specified, the driver will attempt to detect these properties automatically from the following locations:
  a. KRBS Config File (krb5.conf)
  b. Domain Name and Host

- Kerberos Realm – this is the Kerberos Realm used to authenticate the user.

- Kerberos SPN – this is the service principal name (SPN) for the Kerberos Domain Controller. If the SPN on the Kerberos Domain Controller is not the same as the URL that you are authenticating to, use this property to set the SPN.

- Kerberos Keytab File – this is the Keytab file containing your pairs of Kerberos principals and encrypted keys. The Keytab file containing your pairs of Kerberos principals and encrypted keys.

- Kerberos Ticket Cache – this is the full file path to an MIT Kerberos credential cache file. This property can be set if you wish to use a credential cache file that was created using the MIT Kerberos Ticket Manager or kinit command.

- LDAP – select this if you need to integrate with an **LDAP (Lightweight Directory Access Protocol) server** for authentication. LDAP provides secure authentication mechanisms and can enforce policies such as password complexity and expiration.

  Selecting **LDAP** requires you to specify the following:
  - LDAP Server – this is the host name or IP address of the LDAP server.
  - User – this is the Cassandra user account used to authenticate. The authenticating server requires both **User** and **Password** (see below) to validate the user's identity.
  - Password – this is the password used to authenticate the user. The authenticating server requires both **User** (see above) and **Password** to validate the user's identity.
  - LDAP Password – this is the password of the default LDAP user. It must be set if the LDAP server does not allow anonymous bind.
  - LDAP Port – this is the port of the LDAP server. The default value is **389**.
  - Default LDAP User – this is the default LDAP user used to connect to and communicate with the server. It must be set if the LDAP server does not allow anonymous bind.
  - Search Base – this is the search base for your LDAP server and is used to look up users.
  - Search Filter – this is the search filter for looking up usernames in LDAP. The default is **uid =**. When using Active Directory, set the filter to **sAMAccountName=**.

4. **Database** – this is the name of the Cassandra keyspace containing the tables.

5. **Use SSL** – this is a True/False selection to specify whether SSL is enabled. This field sets whether the driver will attempt to negotiate TLS/SSL connections to the server. By default, the driver checks the server's certificate against the system's trusted certificate store. To specify another certificate, set **SSL Server Cert** (see the **Advanced** tab of the ODBC Data Source Administrator DSN Configuration dialog).

   The default value is **False**.

6. **Consistency Level** – this is a dropdown selection containing different consistency levels to determine how many of the replicas of the data you are interacting with need to respond for the query to be considered a success. You need to specify the appropriate replicas in the **Server** (see above) property.

   The consistency levels available for selection are as follows:

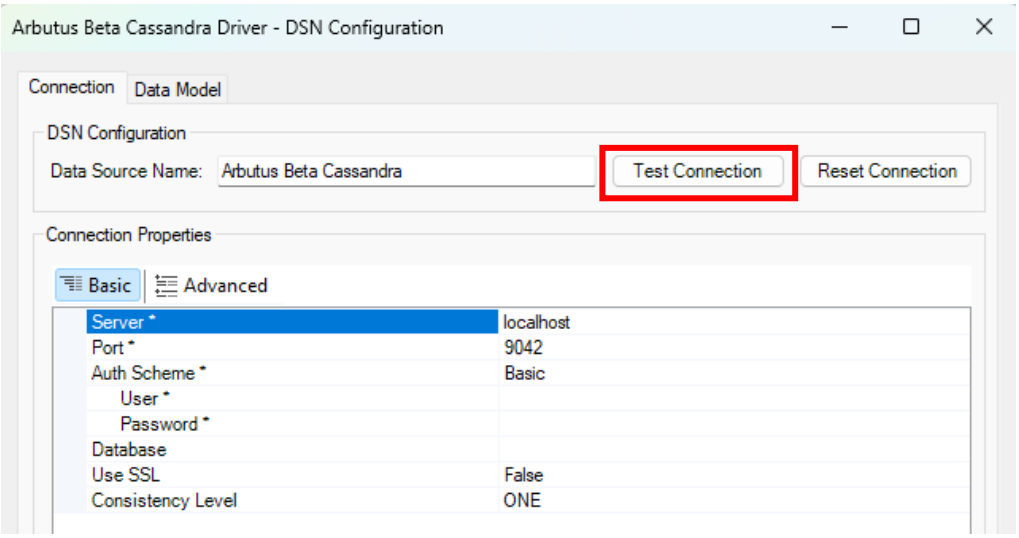
   ONE, TWO, THREE, QUORUM, ALL, LOCAL_QUORUM, EACH_QUORUM, SERIAL, LOCAL_SERIAL, LOCAL_ONE, ANY

   More information on each of the above consistency levels can be provided, if required.

E2. Editing the DSN properties in the Advanced tab

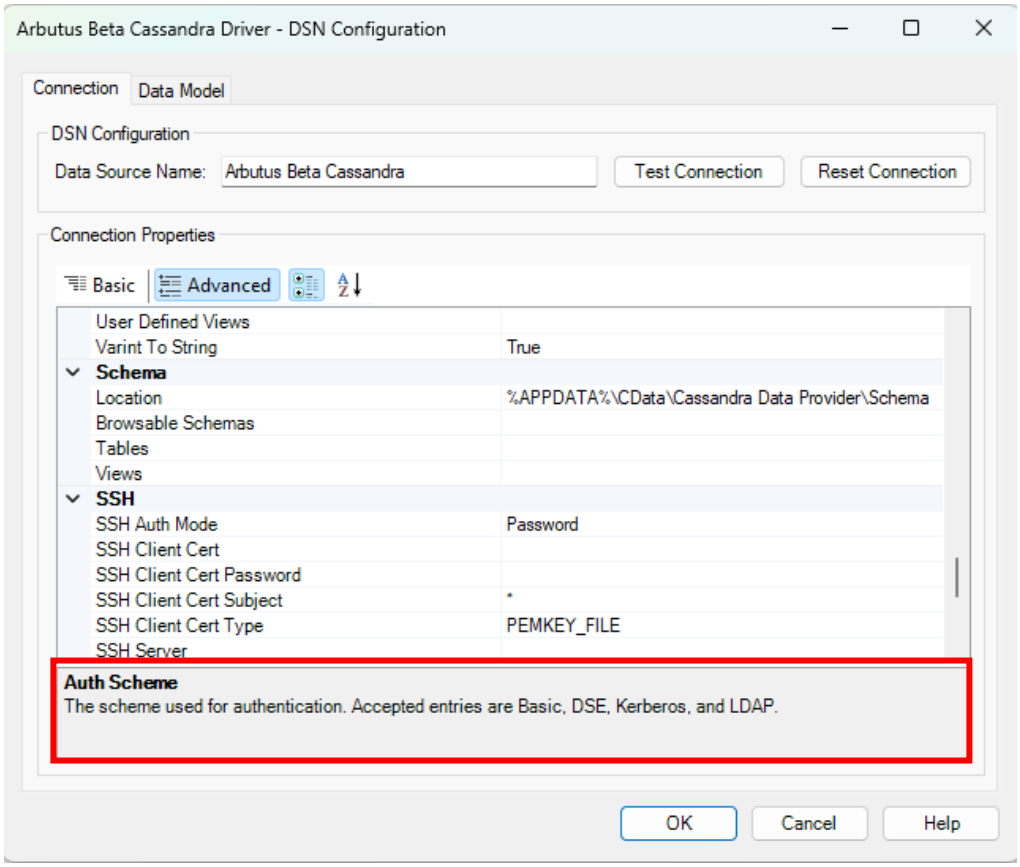
This tab includes more detailed and technical properties. It is intended for those users who need more control over the configuration and are comfortable with more complex options. The **Advanced** tab often includes properties that can fine-tune the behaviour of the system feature.

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 Cassandra 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.