

SAS[®] Data Explorer: User's Guide

2020.1 - 2021.2.1

This document might apply to additional versions of the software. Open this document in [SAS Help Center](#) and click on the version in the banner to see all available versions.

Overview of SAS Data Explorer	2
Understanding SAS Data Explorer	2
Understanding the Available, Data Sources, and Import Tabs	3
Using SAS Data Explorer as a Progressive Web Application	6
Making Data Available to CAS	7
Connecting to Databases	7
Connecting to Remote File Systems	9
User-Defined Connections for Data Sources	12
Importing Data from SAS Folders	18
Importing Local Files	21
Importing Esri Data for Geoenrichment and Geocoding	28
Working with Data in CAS	36
Working with Caslibs and Tables	36
Managing Tags for Tables and Columns	39
Profiling Data	42
Importing Data from the Available Tab or Data Sources Tab	47
Import a Snapshot of an ESP Project	55
Importing Delimited Files from a Caslib Directory	55
Importing Documents from a Caslib Directory	58
Importing Images from a Caslib Directory	62
Importing ORC Files from a Caslib Directory	65
Maintaining SAS Data Explorer	68
Administrative Tasks for SAS Data Explorer and the Data Selection Windows	68
SAS Data Explorer Options	69
General Usage Notes	71
Using the SAS Data Explorer CLI	75

Overview	75
Examples	75

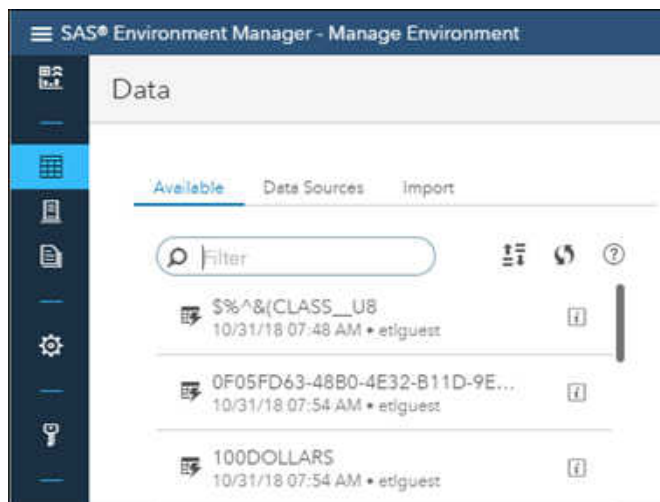
Overview of SAS Data Explorer

Understanding SAS Data Explorer

Data Selection Windows and SAS Data Explorer

SAS Viya applications work with in-memory data on a SAS Cloud Analytic Services (CAS) server. Some of these applications have a data selection window that enables you to upload and manage data. For example, if you click the **Data** area in SAS Environment Manager, a data selection window is displayed.

Figure 1 Data Selection Window in SAS Environment Manager



SAS Data Studio, SAS Visual Analytics, Model Studio, and SAS Intelligent Decisioning have similar data selection windows.


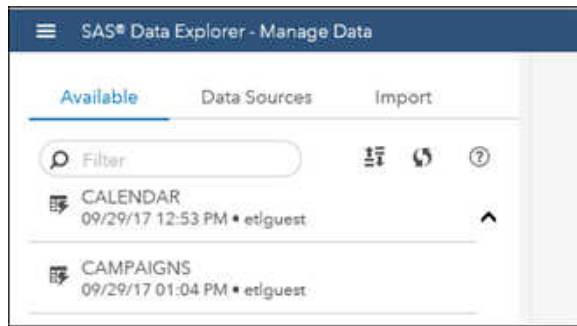
If you license SAS Data Preparation, you have access to some additional data management features. One of those features is the SAS Data Explorer web application. To display SAS Data Explorer, click  at the upper left of SAS Viya applications and then select **Manage Data**. SAS Data Explorer is displayed.

Figure 2 SAS Data Explorer Web Application



SAS Data Explorer and the embedded data selection windows are closely related. They share the tabs that are described in [“Understanding the Available, Data Sources, and Import Tabs” on page 3](#).

If you license SAS Data Preparation, the tabs in both SAS Data Explorer and the embedded data selection windows have the same advanced options. For more information about advanced data profiling options, see [“Profiling Data” on page 42](#).

Some administrative options affect both SAS Data Explorer and the embedded data selection windows. For more information, see [“Administrative Tasks for SAS Data Explorer and the Data Selection Windows” on page 68](#).

Understanding the Available, Data Sources, and Import Tabs

Overview of the Tabs

SAS Data Explorer and the embedded data selection windows have the same tabs: the **Available** tab, the **Data Sources** tab, and the **Import** tab.

Available Tab: Work with In-Memory Tables and Files

If present, the **Available** tab displays all tables and files that have been loaded to memory from any CAS server to which you have access. You can access only those tables or files that are permitted by your login. If you right-click a table or file, you can perform the tasks that are described in [“Working with Tables” on page 37](#).

Data Sources Tab: Connect to External Data Sources

The **Data Sources** tab enables you to create a connection to external data sources. This connection is a caslib. If the connection is successful, tables and files that you are authorized to access are

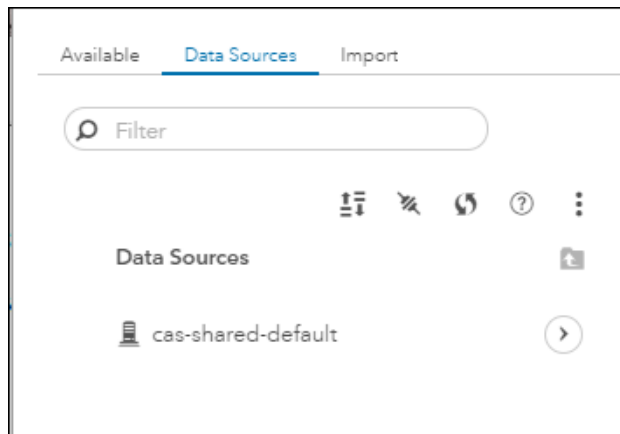
available on the **Data Sources** tab. For more information, see [“Making Data Available to CAS” on page 7](#).

The **Data Sources** tab can also be used to work with caslibs and tables on a CAS server, as described in the following topics:

- [“Working with Caslibs and Tables” on page 36](#)
- [“Managing Tags for Tables and Columns” on page 39](#)
- [“Profiling Data” on page 42](#)

Here is an overview of how the **Data Sources** tab displays CAS servers, caslibs, and tables.

Figure 3 Data Selection Window with the Data Sources Tab Selected





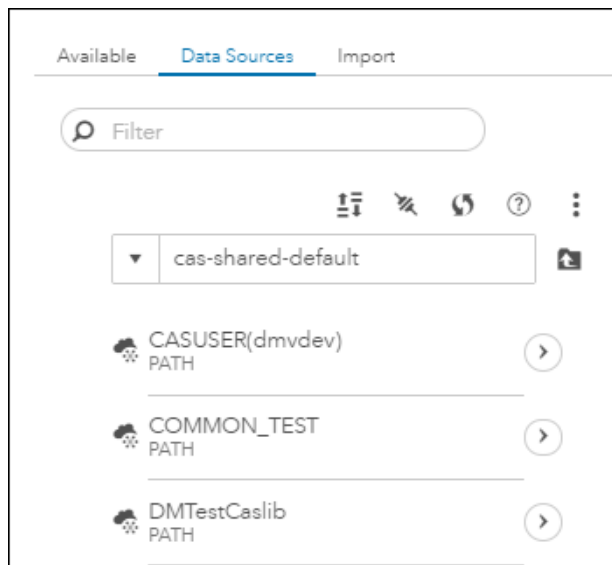


Items with  beside them are the CAS servers that your login enables you to access. Click  to access the caslibs on a given server, as shown in the next figure.

Figure 4 Caslibs on a Selected Server

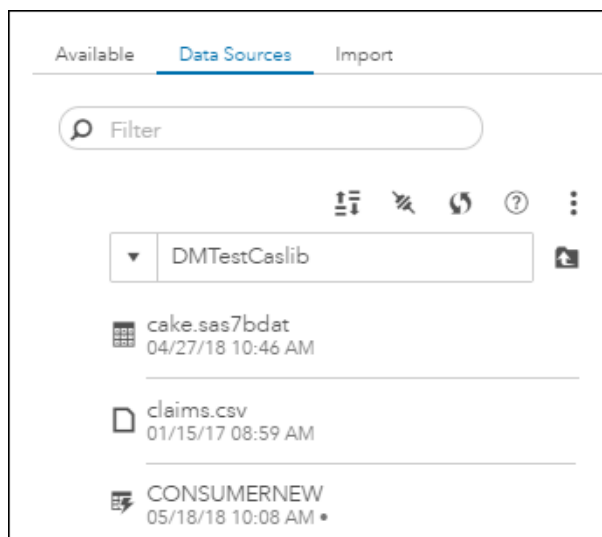




Items with  beside them are global caslibs on the selected CAS server. Global caslibs with names such as CASUSER(**user-ID**) or CASUSERHDFS(**user-ID**) are personal caslibs. Only your user ID can access the data in your personal caslib.

Items with  beside them (and with the **session** label) are session-based caslibs on the selected CAS server. For more information about global and session-based caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#). If you select a caslib, properties for that caslib are displayed on the right. These properties were specified when the caslib was added, such as the name, the CAS server, and the source type. If you right-click a caslib, you can perform the tasks that are described in [“Working with Caslibs” on page 36](#).

Click  to access the tables in a selected caslib, as shown in the next figure.

Figure 5 Tables in a Selected Caslib



Items with  beside them are physical tables that are accessible to the CAS server that have not been loaded to memory. The items with  beside them are tables that have been loaded to memory. If you select a table, three tabs of information are displayed on the right side: **Details**, **Sample Data**, and **Profile**. If you right-click a table, you can select various options, such as **Load** and **Profile**. For more information about these tabs and options, see [“Working with Tables” on page 37](#).

Import Tab: Copy Data from a Caslib

The **Import** tab enables you to copy data that is accessed with a caslib. For more information, see the following topics:

- [“Importing Data from the Available Tab or Data Sources Tab” on page 47](#)
- [“Import a Snapshot of an ESP Project” on page 55](#)
- [“Importing Delimited Files from a Caslib Directory” on page 55](#)
- [“Importing Documents from a Caslib Directory” on page 58](#)
- [“Importing Images from a Caslib Directory” on page 62](#)
- [“Importing ORC Files from a Caslib Directory” on page 65](#)

Administrators can change the default rules and deny access to the **Import** tab or to individual social media feeds on this tab. For more information, see [“Limit Access to the Import Tab” on page 68](#).

Using SAS Data Explorer as a Progressive Web Application

As of 2021.1.2, SAS Data Explorer can run as a Progressive Web Application (PWA). The benefits of using SAS Data Explorer as a PWA include the following:

- Application persistence – By default, your session will never time out, so you can restart your work more quickly. Your administrator can control the timing with a configuration property in SAS Environment Manager.
- Performance – When installed as a PWA, SAS Data Explorer typically starts faster than when accessed in the browser.
- Desktop experience – Installing SAS Data Explorer as a PWA confers all the benefits of a traditional installation. You can launch SAS Data Explorer from the **Start** menu or taskbar, and you do not need to sort through many tabs to find the correct instance of SAS Data Explorer.
- Renaming – You can rename each PWA instance to quickly access different environments, such as development, test, or production servers.

To install the PWA:

- 1 Open any other SAS web app (such as SAS Drive) in a Chromium-based web browser (such as Chrome).
- 2 Open the web browser's **More** menu and select **Install SAS**.

Note: The **Install SAS** option might be located under another menu option. For example, within the Microsoft Edge browser, it is located under **Apps**.

The web app is now installed as a desktop program named SAS. After one web application has been installed as a PWA, all other SAS applications are accessible in the PWA.

To uninstall the PWA:

- 1 Open the PWA.
- 2 In the PWA menu, select **Uninstall SAS**.

Making Data Available to CAS

Connecting to Databases

Overview of Database Connections

The **Data Sources** tab enables you to add a new caslib that connects to a database server. If the connection is successful, tables that you are authorized to access in the database are available on the **Data Sources** tab. The tables are not automatically loaded to memory.

Gather the following information to add a new caslib that connects to a database server:

- Identify a name for the new caslib and the CAS server to be used in the connection.
- The Connection Settings dialog box includes a check box: **Persist this connection beyond the current session**. Select this check box to add a global caslib for this connection. Deselect this check box to add a session-based caslib for this connection. For more information about this option, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).

- Identify the data source that you want to access.

You can create a connection between a caslib and any supported database. Ask the database administrator for the credentials and other information that you will need to connect to the database. For information about the versions of third-party databases that are supported, see [“Data Source Requirements” in System Requirements for SAS Viya](#).

You can add a connection for a database other than the databases listed above. See [“User-Defined Connections for Data Sources” on page 12](#).

- Obtain any access credentials, physical pathnames, and connection options that are required to make the connection to the desired database server.
- The Connection Settings window includes an **Advanced** tab where you can specify advanced connection settings. See the documentation for any special connection options that you want to use.

Connection options for most databases are described in [Data Connectors](#).

SAS LASR Analytic Server connection options are described in [Data Source Arguments for SAS LASR Analytic Server](#).

Create a Database Connection

Use information that you previously gathered to add a new caslib that connects to a database server:

- 1 Display the window that contains the **Data Sources** tab. See [“Data Selection Windows and SAS Data Explorer” on page 2](#).
- 2 Click the **Data Sources** tab.

- 3 Click  on the **Data Sources** tab.

The Connection Settings dialog box is displayed.

- 4 Enter a name for the caslib in the **Name** field.

Follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). The target operating system and data source might have additional constraints on these names.

- 5 Accept the default CAS server or select another CAS server in the **Server** field.

- 6 Select **Database** in the connection **Type** field.

- 7 Select the database that you want to access in the **Source type** field. Click ⓘ for details about the data connector options for the selected database.


- 8 Select the **Persist this connection beyond the current session** check box to add a global caslib for this connection. Deselect this check box to add a session-based caslib for this connection. For more information about this option, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).

- 9 Using information that you gathered about this connection, enter access credentials, physical pathnames, and other connection details into the fields on the **Settings** tab.



- 10 Click the **Advanced** tab to specify additional options for the current data source. Click ⓘ beside the **Source type** field for details about options for the current data source.

- 11 Click **Test Connection** to test your connection.

- 12 When ready, click **Save** to save your connection.

If the connection succeeds, tables that you are authorized to access in the database are available from the caslib that you specified in Step 4. Information about caslibs and tables on the **Available** and **Data Sources** tabs is stored in the cache for your web browser. If you think this information does not reflect the current state of your system, click  in the nearest toolbar.

If the connection fails, see [“General Usage Notes” on page 71](#).

- 13 If the target caslib is not visible in your current view, click  or scroll (such as with ) to find the caslib on the specified CAS server.

- 14 Check to see what menu options are available for the tables or files in the new caslib.

If the saved connection uses a global caslib, the following options are available when you right-click a table in the caslib: **Load**, **Delete**, **Add to import**, and **Run profile**. For information about these options, see [“Working with Tables” on page 37](#).

If the saved connection uses a session-based caslib, the following options are valid when you right-click a table in the caslib: **Delete** and **Add to import**. You do not have access to the **Load** option, which is used to load a table or file to memory on a CAS server. The **Run profile** is available, but it is not valid for tables in a session-based caslib.

However, if you can access an appropriate global caslib, you can use the **Add to import** option to copy the table to that global caslib. From there, you have access to the **Load** option and the **Run profile** option. For information about the **Add to import** option, see [“Importing Data from the Available Tab or Data Sources Tab” on page 47](#).

Usage Notes for Database Connections

Add a New Database Connection If the Password Changes

To update the password for a database connection, delete the old connection and add a new one with the new password. You cannot edit a database connection and replace the password.

Hadoop Hive Connections: Default Length for String Columns

Hadoop Hive connections include a **Default character length** field. This field determines the default length of string columns in Hadoop Hive tables. The default value is 32767, which is a large value and could cause memory allocation problems and decreased performance. In most cases, you should specify a value that is more accurate for the actual length of the string rather than using the default.

Managing Table Names for SAS LASR Analytic Server

The LIBNAME statement for a SAS LASR Analytic Server data source includes a server tag value. If no value is specified for the server tag, the value of WORK is used as the server tag value.

If a caslib points to a LASR data source without specifying a server tag value, the CAS server queries the SAS LASR Analytic Server for the loaded tables. It returns the table names in the form `serverTag.LASRtableName`. If a caslib points to a LASR data source and does specify the server tag value, the CAS server queries the SAS LASR Analytic Server for the loaded tables. It returns the table names in the form `LASRtableName`.

If you do not want the server tag value prepended to the names of LASR tables, specify the server tag on the **Advanced** tab of the caslib for the LASR data source. Identify the server tag value for the LASR data source. Use the **Additional Parameters** control to add a name/value pair that specifies the server tag value. The name is `tag` and the value is the value for server tag from the SAS LASR Analytic Server library definition in SAS Management Console or the `tag=` attribute in the LIBNAME statement. Example parameters: **Name:** `tag`, **Value:** `cftb1a1`.

Connecting to Remote File Systems

Overview of Remote File System Connections

The **Data Sources** tab enables you to add a new caslib that connects to one of the following remote file systems:

- DNFS (distributed network file system). Examples include MapR-FS and EMC Isilon, which are alternatives to Hadoop.
- PATH (a server-side directory that a CAS controller can access).

You can also add a connection for a remote file system other than the systems listed here. See [“User-Defined Connections for Data Sources” on page 12](#).

If the connection is successful, the tables that you are authorized to access are available on the **Data Sources** tab. The tables are not automatically loaded to memory.

Gather the following information to add a new caslib that connects to a remote file system:

- Identify a name for the new caslib and the CAS server to be used in the connection.
- Identify the type of remote file system that you want to access: a DNFS or a server-side directory (PATH).
- The Connection Settings window includes a check box: **Persist this connection beyond the current session**. Select this check box to add a global caslib for this connection. Deselect this check box to add a session-based caslib for this connection. For more information about this option, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).
- Ask the file system administrator for the physical path and other information that is necessary to access an appropriate directory on that system. If the new caslib is a session-based caslib, you must specify an existing directory, or the connection fails. If the new caslib is global in scope, you can specify a directory that does not exist on the target file system, and that directory are created. To create a new directory, you must have Write access to the target location. The new directory must be located no more than one level below an existing directory.
- The **Connection Settings** window has an **Include subdirectories** check box. Select this check box only if you are creating a path-based caslib for a feature that requires directory support. Features that require a path-based caslib with directory support include tables in Parquet format, tables in ORC format, and the **Directory** data sources on the **Import** tab. In most cases, do not select this check box. Each **File System** connection to a directory usually provides access to data in that directory only, not to data in any subdirectories. For example, the **Data Sources** tab shows tables and files in a directory but not subdirectories in a directory.
- The **Connection Settings** window includes an **Advanced** tab where you can specify advanced connection settings. See the documentation for any special connection options that you want to use.


[DNFS Data Source](#)

[Path-Based Data](#)

Note: When you use a path-based caslib to import data from a remote file system, the CAS server does not maintain pre-existing sort orders. You must re-sort the data after you import it.

Connect to a Remote File System

Use information that you previously gathered to add a new caslib that connects to a remote file system:

- 1 Display the window that contains the **Data Sources** tab. See [“Data Selection Windows and SAS Data Explorer” on page 2](#).
- 2 Click the **Data Sources** tab.
- 3 Click  on the **Data Sources** tab. The Connection Settings dialog box is displayed.
- 4 Enter a name for the caslib in the **Name** field. If you change the name, follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). The target operating system and data source might have additional constraints on these names.
- 5 Accept the default CAS server or select another CAS server in the **Server** field.
- 6 Select **File System** in the connection **Type** field.

- 7 Select the remote file system that you want to access in the **Select source type** field: **DNFS** or **PATH**.

- 8 Select the **Persist this connection beyond the current session** check box to add a global caslib for this connection. Deselect this check box to add a session-based caslib for this connection. For more information about this option, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).

- 9 Enter the physical path to the remote file system in the **Path** field.

If the new caslib is global in scope, you can specify a directory that does not exist on the target file system, and that directory are created. If the new caslib is a session-based caslib, you must specify an existing directory, or the connection fails. For more information about global and session-based caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).

- 10 Enter a **Description** for the connection, if desired.

- 11 Select the **Include subdirectories** check box only if you are creating a path-based caslib for a feature that requires directory support. Features that require a path-based caslib with directory support include tables in Parquet format, tables in ORC format, and the **Directory** data sources on the **Import** tab. In most cases, do not select this check box. Each **File System** connection to a directory usually provides access to data in that directory only, not to data in any subdirectories. For example, the **Data Sources** tab shows tables and files in a directory but not subdirectories in a directory.

- 12 Click the **Advanced** tab to specify additional options for the current data source. Click ⓘ beside the **Source type** field for details about options for the current data source.

- 13 Click **Test Connection** to test your connection.

- 14 When ready, click **Save** to save your connection.

If the connection succeeds, tables that you are authorized to access in the remote directory are available from the caslib that you specified in Step 4. Information about caslibs and tables on the **Available** and **Data Sources** tabs is stored in the cache for your web browser. If you think this information does not reflect the current state of your system, click ↺ in the nearest toolbar.

If the connection fails, see [“General Usage Notes” on page 71](#).

- 15 If the target caslib is not visible from your current view, click 🗂 or scroll (such as with ➤) to find the caslib on the specified CAS server.

- 16 Check to see what options are available for the tables or files in the new caslib.

If the saved connection uses a global caslib, the following options are available when you right-click a table in the caslib: **Load**, **Delete**, **Add to import**, and **Run profile**. For information about these options, see [“Working with Tables” on page 37](#).

If the saved connection uses a session-based caslib, the following options are valid when you right-click a table in the caslib: **Delete** and **Add to import**. You do not have access to the **Load** option, which is used to load a table or file to memory on a CAS server. The **Run profile** is available, but it is not valid for tables in a session-based caslib.

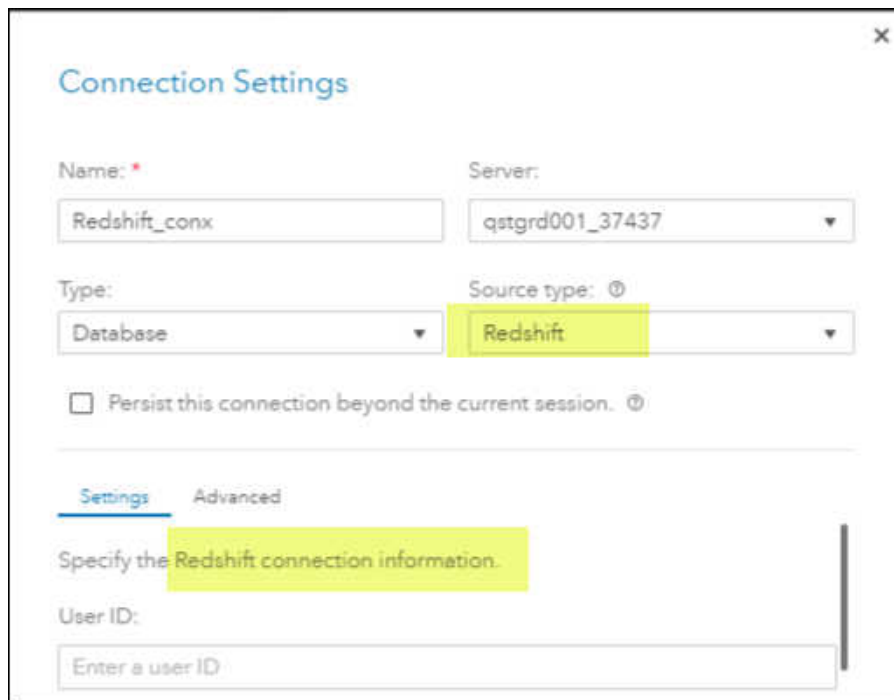
However, if you can access an appropriate global caslib, you can use the **Add to import** option to copy the table to that global caslib. From there, you have access to the **Load** option and the **Run profile** option. For information about the **Add to import** option, see [“Importing Data from the Available Tab or Data Sources Tab” on page 47](#).

User-Defined Connections for Data Sources

Overview of User-Defined Connections

When you add a connection in the Connection Settings window, you typically select a connection type of **Database** or **File System**. Then you select the source type, such as **Redshift** or **HDFS**. The Connection Settings window is populated with settings for the selected **Source type**.

Figure 6 Redshift Source Type Selected on the Connection Settings Window




The screenshot shows the 'Connection Settings' window. The 'Name' field contains 'Redshift_conx' and the 'Server' dropdown shows 'qstgrd001_37437'. The 'Type' dropdown is set to 'Database', and the 'Source type' dropdown is set to 'Redshift', which is highlighted in yellow. Below these fields is a checkbox labeled 'Persist this connection beyond the current session.' which is unchecked. At the bottom, there are two tabs: 'Settings' (selected) and 'Advanced'. Under the 'Settings' tab, there is a yellow-highlighted box containing the text 'Specify the Redshift connection information.' and a 'User ID' field with the placeholder text 'Enter a user ID'.

To connect to a database or remote file system that is not listed in the **Source type** field, you can select **Other** in the connection **Type** field. The selection of **Other** enables you to add a user-defined connection for a database or remote file system that is not available for selection in the **Source type** field. Use the following topics to help you plan the connection settings that you want to specify.


General Steps for Adding a User-Defined Connection


Use information that you previously gathered to add a new caslib with user-defined connection properties for a data source.

- 1 Display the window that contains the **Data Sources** tab. See [“Data Selection Windows and SAS Data Explorer” on page 2](#).
- 2 Click the **Data Sources** tab.



- 3 Click  on the **Data Sources** tab. The Connection Settings window is displayed.
- 4 Enter a name for the caslib in the **Name** field. Follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). The target operating system and data source might have additional constraints on these names.
- 5 Accept the default CAS server or select another CAS server in the **Server** field.
- 6 Select **Other** in the connection **Type** field.
- 7 Enter the source type for your connection in the **Source type** field. See the example values for **Source type** in the examples below.
- 8 Select the **Persist this connection beyond the current session** check box to add a global caslib for this connection. Deselect this check box to add a session-based caslib for this connection. For more information about this option, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).
- 9 In the **Settings** tab, enter a description for this connection.
- 10 Click the **Advanced** tab to specify connection options for the current data source. Specify **Data transfer mode**, **Number of read nodes**, and **Number of write nodes** according to the information that you previously gathered.

If your SAS administrator has provided an Authentication Domain for this data source, you can specify the domain name instead of specifying your credentials under **Additional Parameters**. For more information about SAS Authentication Domains, see [“Authentication Domain” in SAS Viya Administration: External Credentials](#).

- 11 Enter name/value pairs to specify **Additional Parameters** for this connection.
 In the **Name** field, enter the name of the option that you want to specify, as described in the SAS documentation for the target data source. In the **Value** field, enter the value that you want to specify for that option. For example, you could specify the option `server` and specify a value of `VTserver`. You can also click  beside the **Source type** field for details about options for user-defined data sources. If you have not specified an Authentication Domain above, you can specify a server name, user credentials, and other attributes as name/value pairs in this area.
- 12 After you have entered the name/value pair for a parameter, click **Add Parameter** to add it to the caslib.
- 13 When ready, click **Save** to save your connection.

If the connection succeeds, tables that you are authorized to access in the remote directory are available from the caslib that you specified in Step 4. Information about caslibs and tables on the **Available** and **Data Sources** tabs is stored in the cache for your web browser. If you think this information does not reflect the current state of your system, click  in the nearest toolbar.

If the connection fails, see [“General Usage Notes” on page 71](#).

- 14 If the target caslib is not visible in your current view, click  or scroll (such as with ) to find the caslib on the specified CAS server.
- 15 Check the menu options that are available for the tables or files in the new caslib.
 If the saved connection uses a global caslib, the following options are available when you right-click a table in the caslib: **Load**, **Delete**, **Add to import**, and **Run profile**. For information about these options, see [“Working with Tables” on page 37](#).
 If the saved connection uses a session-based caslib, the following options are valid when you right-click a table in the caslib: **Delete** and **Add to import**. You do not have access to the **Load**

option, which is used to load a table or file to memory on a CAS server. The **Run profile** option is available, but it is not valid for tables in a session-based caslib.

However, if you can access an appropriate global caslib, you can use the **Add to import** option to copy the table to that global caslib. From there, you have access to the **Load** option and the **Run profile** option. For information about the **Add to import** option, see [“Importing Data from the Available Tab or Data Sources Tab” on page 47](#).

Example: User-Defined Connection to a Database

Ask the database administrator for the information required to connect to the target database. Refer to the SAS data connector syntax for the target database for connection options. See [Quick Reference for Data Connector Syntax](#).

You will need user credentials, for example. You can specify a user name and password as name/value pairs in the **Additional Parameters** section of the **Advanced** tab, as shown below. Alternatively, a SAS administrator can create an Authentication Domain for a connection, and you can specify `authenticationdomain` and the domain name as name/value pair under **Additional Parameters** on the **Advanced** tab. For more information about SAS Authentication Domains, see [“Authentication Domain” in SAS Viya Administration: External Credentials](#).

Follow the steps described in [“General Steps for Adding a User-Defined Connection” on page 12](#). Here is an example of a user-defined connection to a database.

■ Basic connection settings

- ☐ **Name:** Vertica Connection
- ☐ **Server:** cas_server02
- ☐ **Type:** Other
- ☐ **Source type:** vertica
- ☐ **Persist this connection beyond the current session:** selected

■ Settings Tab

- ☐ **Description:** Access Vertica tables

■ Advanced Tab

- ☐ **Authentication domain:** If your SAS administrator has provided an Authentication Domain for this data source, you can specify the domain name instead of specifying your credentials under **Additional Parameters**.
- ☐ **Data transfer mode:** Default
- ☐ **Number of read nodes:** 3
- ☐ **Number of write nodes:** 3
- ☐ **Additional Parameters** If you have not specified an Authentication Domain above, you can specify a server name, user credentials, and other attributes as name/value pairs in this field. Here are some example name/value pairs:
 - **Name:** server-name **Value:** VTserver
 - **Name:** user-name **Value:** user1
 - **Name:** password **Value:** myPwd
 - **Name:** schema-name **Value:** mySchema

Example: User-Defined Connection to a Remote File System

Ask the file system administrator for the information required to connect to the target directory on the remote file system. Refer to SAS data connector syntax or CASLIB statement syntax for the target file system for connection options. See [Quick Reference for Data Connector Syntax](#) or [Required Arguments for the CASLIB Statement](#).

You will need user credentials, for example. You can specify a user name and password as name/value pairs in the **Additional Parameters** section of the **Advanced** tab, as shown below. Alternatively, a SAS administrator can create an Authentication Domain for a connection, and you can specify `authenticationdomain` and the domain name as name/value pair under **Additional Parameters** on the **Advanced** tab. For more information about SAS Authentication Domains, see [“Authentication Domain” in SAS Viya Administration: External Credentials](#).

Follow the steps described in [“General Steps for Adding a User-Defined Connection” on page 12](#). Here is an example of a user-defined connection to a remote file system.

■ Basic connection settings

- ☐ **Name:** `Spark_Connection`
- ☐ **Server:** `cas_server02`
- ☐ **Type:** `Other`
- ☐ **Source type:** `spark`
- ☐ **Persist this connection beyond the current session:** selected

■ Settings Tab

- ☐ **Description:** `Access Hadoop Hive tables with Spark`

■ Advanced Tab

- ☐ **Authentication domain:** If your SAS administrator has provided an Authentication Domain for this data source, you can specify the domain instead of specifying your credentials under **Additional Parameters**.
- ☐ **Data transfer mode:** `Parallel`
- ☐ **Number of read nodes:** `3`
- ☐ **Number of write nodes:** `3`
- ☐ **Additional Parameters** If you have not specified an Authentication Domain above, you can specify a server name, user credentials, and other attributes as name/value pairs in this field. Here are some example name/value pairs:
 - **Name:** `hadoopjarpath` **Value:** `/hadoop/jars:hadoop/spark/jars`
 - **Name:** `hadoopConfigDir` **Value:** `/hadoop/conf`
 - **Name:** `server` **Value:** `thriftserver`
 - **Name:** `schema` **Value:** `mySchema`
 - **Name:** `username` **Value:** `user1`
 - **Name:** `password` **Value:** `myPwd`
 - **Name:** `dbmaxText` **Value:** `2048`

Example: User-Defined Connection to an ESP Project

SAS Event Stream Processing enables you to capture and analyze streaming data from sensors, IoT devices, and other data sources. You can add a user-defined connection to a project on an ESP server. The "windows" created by that project are displayed as unloaded tables on the **Data Sources** tab. If you import one of these tables and load it to memory, you have a snapshot of the data that was streamed through that window. You can then run a data profile and perform other tasks with this snapshot of streamed data.

Ask the ESP system administrator for the information that is required to connect to the desired server and project. For more information about connection options for an ESP server, "[SAS Event Stream Processing Data Source](#)." Follow the steps described in "[General Steps for Adding a User-Defined Connection](#)" on page 12. Here is an example of a user-defined connection to an ESP project.

- **Basic connection settings**
 - ☐ **Name:** `ESP_Project_Connection`
 - ☐ **Server:** `cas_server02`
 - ☐ **Type:** Other
 - ☐ **Source type:** ESP
 - ☐ **Persist this connection beyond the current session:** selected
- **Settings Tab** Specify a path and a description for the ESP project.
 - ☐ **Path:** `esp_project_path`
 - ☐ **Description:** Access ESP Project
- **Advanced Tab**
 - ☐ **Data transfer mode:** Default
 - ☐ **Additional Parameters** Specify the port and name of the ESP server.
 - **Name:** `port` **Value:** 1111
 - **Name:** `server` **Value:** `esp_server_name`

After you are connected to a project on an ESP server, see "[Import a Snapshot of an ESP Project](#)" on page 55.

Example: User-Defined Connection to ADLS

ADLS (Azure Data Lake Storage, Generation 2) is a cloud storage system from Microsoft. Ask the ADLS administrator for the information that is required to connect to the target directory on the remote file system. For more information about securing Azure connections, see "[Azure Data Lake Storage Data Source](#)" in *SAS Cloud Analytic Services: User's Guide*.

Follow the steps described in "[General Steps for Adding a User-Defined Connection](#)" on page 12. Here is an example of a user-defined connection to ADLS.

- **Basic connection settings**
 - ☐ **Name:** `Azure_Connection`
 - ☐ **Server:** `cas_server02`

- ☐ **Type:** Other
- ☐ **Source type:** ADLS
- ☐ **Persist this connection beyond the current session:** selected
- **Settings Tab**
 - ☐ **Path:** /prod/dataprep/
 - ☐ **Description:** Access Prod Tables in Azure
 - ☐ **Include subdirectories** selected
- **Advanced Tab**
 - ☐ **Additional Parameters** Specify authentication information and other options. Here are some example name/value pairs:
 - **Name:** *accountName* **Value:** *adls_account_name*
 - **Name:** *applicationID* **Value:** *application_id*
 - **Name:** *timeout* **Value:** 50000
 - **Name:** *filesystem* **Value:** *hdfplib*

Example: User-Defined Connection to S3

S3 (Simple Storage Service of Amazon Web Services) is a cloud storage system. Ask the S3 administrator for the information that is required to connect to the target directory on the remote file system. For more information about S3 connections, see [“S3 Data Source” in SAS Cloud Analytic Services: User’s Guide](#).

Follow the steps described in [“General Steps for Adding a User-Defined Connection”](#) on page 12. Here is an example of a user-defined connection to S3.

- **Basic connection settings**
 - ☐ **Name:** *S3_Connection*
 - ☐ **Server:** *cas_server02*
 - ☐ **Type:** Other
 - ☐ **Source type:** s3
 - ☐ **Persist this connection beyond the current session:** selected
- **Settings Tab**
 - ☐ **Path:** /orgdata/
 - ☐ **Description:** Access Prod Tables in S3
 - ☐ **Include subdirectories** selected
- **Advanced Tab**
 - ☐ **Additional Parameters** Specify authentication information and other options. Here are some example name/value pairs:
 - **Name:** *accessKeyId* **Value:** *s3_access_id*
 - **Name:** *useSSL* **Value:** *false*
 - **Name:** *secretAccessKey* **Value:** *secret_id*

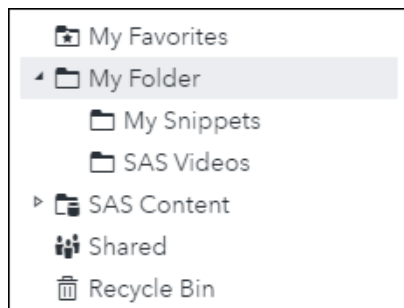
- **Name:** *bucket* **Value:** *orgdata*
- **Name:** *region* **Value:** *US_East*
- **Name:** *useIAMRole* **Value:** *false*

Importing Data from SAS Folders

Overview of SAS Folders

When you install SAS Viya, a set of default SAS folders is created. The folders are arranged in a structure that separates system information from business information, provides personal folders for individual users, and provides an area for shared data. The following figure shows a set of folders for a user.

Figure 7 SAS Folders



SAS Viya applications save their content to SAS folders. In SAS Drive, you can select **New** ⇒ **Upload** ⇒ **Files to My Folder** to upload external files to a SAS folder.

Import Files from SAS Folders

Use the **Folders** option on the **Import** tab to import files that were saved to a SAS folder. The file must be in one of the following formats:

- Text files delimited with a comma or some other character.
- SAS data sets (SASHDAT or SAS7BDAT). SAS data set views (SAS7BVEW) cannot be loaded into CAS tables.
- Microsoft Excel workbook (XLSX) files and Excel 97-2003 workbook (XLS) files. You cannot import XLST, XLSB, XLSM, or other Excel file types. You cannot import pivot tables.

1 Display the window that contains the **Import** tab.

See [“Data Selection Windows and SAS Data Explorer” on page 2](#).

2 Click the **Import** tab.

3 Click **Folders** in the import queue.

A selection window for SAS folders is displayed.

- 4 Navigate to the SAS folder that contains the file to be imported. Select the file and click **OK**.

The file is added to the queue on the **Import** tab. Import properties for the selected file appear on the right, as shown in the next figure.

Figure 8 Import a File from a SAS Folder

The screenshot shows the SAS Import wizard interface. On the left, under the 'Import (1)' tab, the file 'employeeExcel.xls' is listed with its location '/My Folder'. On the right, the configuration for the target table 'employeeExcel' is shown. The 'Target table name' field contains 'employeeExcel'. The 'Target location' is set to 'cas-shared-default/Public' with a 'Find' button. There are checkboxes for 'Save as an in-memory table only' (unchecked), 'If target table name exists:' with radio buttons for 'Cancel import' (selected) and 'Replace file' (unchecked), a 'Label' field with 'Enter label', and a 'Format' dropdown set to 'sashdat'. At the bottom, under 'File Specifications', there are checkboxes for 'Specify a worksheet to import:' (unchecked), 'First row contains column names' (checked), and 'Limit the range of imported rows and columns' (unchecked).

- 5 The name of the source was copied into the **Target table name** field on the right. Accept or change this name as appropriate for the copy.

If you change the name, follow the SAS conventions for table names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). The target operating system and data source might have additional constraints on these names. For example, the target table name can be a maximum of 247 characters long. However, a shorter file name might be required due to the data source settings for the caslib that is specified in the **Target location** field. Contact your data administrator for more information about the permitted length for target table names. See also [“Table Names, Column Names, and Special Characters for Imported Tables” on page 28](#).

- 6 A default caslib is provided in the **Target location** field. To override the default, select an existing global caslib or add a new one.

If you have the privilege that is required to add a new global caslib, follow the name conventions described in [“Names for Caslibs, Tables, and Columns” on page 72](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).

- 7 Specify what action the import operation should take if the target file name exists in the caslib that was specified in the **Target location** field. The options are to cancel the import or to replace the existing file that has the same name.
- 8 Specify a label that helps identify the copied table, if desired. The caslib for the target table must support labels. For example, caslibs of type PATH, DNFS, and HDFS support labels. A caslib for a DBMS does not support labels.
- 9 Use the **Format** selector to select the desired format for the target table. Only SASHDAT and CSV formats are supported.

- 10** Use the **File Specifications** tab to specify any options that are needed to accurately copy the contents of the imported file.

Microsoft Excel spreadsheet options:

The following options are available for Microsoft Excel spreadsheets:

Specify sheet to import

For spreadsheets that contain multiple worksheets, indicate the name of the worksheet that you want to import. If you do not specify a worksheet name, then only the first worksheet in the spreadsheet is imported.

First row contains column names

Select this check box if the first row in the file contains column names.

Limit range of imported rows and columns

Select this check box to limit the import operation to a range of cells in the spreadsheet. To use this field, you must also indicate the name of the worksheet in the **Specify sheet to import** field, even if the spreadsheet contains only one worksheet.

Note: If you select both **Limit range of imported rows and columns** and **First row contains column names**, the first row in the specified range of cells is assumed to contain column names.

Delimited text file options:

The following options are available for delimited text files:

Input file delimiter

Select the delimiter that is used in the file that you want to import. The default value is a comma. If you select **Custom**, you can specify a single character to use as a user-defined delimiter.

Scanned rows

Select the number of rows to scan in order to determine column data types and lengths. A smaller value causes the import to complete quickly, but you increase the possibility of obtaining a value that is too small to accommodate character columns. A larger value reduces the possibility of truncating character columns, but it increases processing time.

Locale

Enter the locale code for the file. An example is **fr-FR**. Specifying the locale code is important so that special characters (for example, commas) in the file are interpreted correctly based on locale. If the field does not contain a code value, the default is **en-US**.

For examples of locale code values, see [Table of Language Culture Names, Codes, and ISO Values](#).

Encoding

Enter the encoding of the file. When importing UTF-8 or UTF-16 data, make sure that SAS Web Application Server is a Unicode server or that the file contents can be transcoded to the encoding of SAS Web Application Server. Examples of valid values include **utf8**, **utf16le**, and **euc-cn**. The default value is **utf8**.

First row contains column names

Select this check box if the first row in the file contains column names.

Convert character columns to variable size

Select this check box if the file includes columns that contain varying-length character strings. When the file is imported, this option assigns the varying-length character string (VARCHAR) format to all columns that contain character data.

CAUTION

If you select this check box, the VARCHAR format is applied to all columns that contain character data, rather than to only those with varying-length character strings.

SAS data set, SAS7BDAT options:

The following options are available for SAS7BDAT files:

Password

If the file is password-protected, enter the password for the file. If the file is password-protected and you do not enter the password in this field, then the file fails to import.

Encryption key

If the file is encrypted, enter the encryption key for the file. Otherwise, the file fails to import.

Character multiplier

Modify this option to increase the number of characters that can fit in each cell so that character data truncation does not occur. The lengths for character variables can be increased by multiplying the current length by the value that you specify. You can specify a multiplier value from 1 to 5. The default value is 1. If your data contains international characters, and the data is not importing correctly, you might want to specify a higher value.

SAS data set, SASHDAT options:

The following options are available for SASHDAT files:

Encryption key

If the file is encrypted, enter the encryption key for the file. Otherwise, the file fails to import.

- 11** You can use the **Unique ID** control on the **Advanced** tab to add a column to the target table. The new column contains a unique value for each row. A tag is added to the new column. You can view the new column and tag on the **Details** tab for the target table.

For delimited text files, you can use the **Additional Input Options** control on the **Advanced** tab to specify any additional options that might be needed to accurately copy the contents of the imported file. To add options, select **Create additional input options**. Click ⓘ for a description of the options for local delimited text files. Options are entered as name/value pairs, such as **Name:** *containsHeaderRow*, **Value:** *TRUE*. Click **Add Option** to add the option to the import record.

- 12** When ready, you can right-click the file to be copied and select **Import Item**.

If the import succeeds, a copy of the file is loaded to memory on the CAS server that is specified in the caslib. The copy of the file can be selected from the **Available** tab or the **Data Sources** tab.

If the import fails, see [“Usage Notes for Importing Local Files” on page 27](#). See also [“General Usage Notes” on page 71](#).

[“Usage Notes for Importing Local Files” on page 27](#) also apply to files imported from SAS folders.

Importing Local Files

Overview of Importing Local Files

Local files are local to the browser where your SAS Viya web application is running. The c:\ drive on a Windows machine is one example of a local file system. Network file systems and shared folders are also included such as \\nas\spreadsheets.

The **Local files** option on the **Import** tab supports two methods for copying a local file and loading it to memory on a CAS server:

- Use the **Local files** ⇒ **Microsoft Excel (multiple worksheets)** option to import Excel files with multiple worksheets. Each worksheet is loaded to memory as a separate table.
- Use the **Local files** ⇒ **Local file** option to import other local files.

Use the Local File Option

Use the **Local files** ⇒ **Local file** option to import files in the following formats.

- Text files delimited with a comma or some other character.
- SAS data sets (SASHDAT or SAS7BDAT). SAS data set views (SAS7BVIEW) cannot be loaded into CAS tables.
- Microsoft Excel workbook (XLSX) files and Excel 97-2003 workbook (XLS) files. You cannot import XLST, XLSB, XLSM, or other Excel file types. You cannot import pivot tables.

1 Display the window that contains the **Import** tab.

See “[Data Selection Windows and SAS Data Explorer](#)” on page 2.

2 Click the **Import** tab.

3 Drag and drop a local file onto the **Import** tab. Alternatively, click **Local files** on the **Import** tab and select **Local file**. Use the file selection window to navigate to a local file and select it.

The file is added to the queue on the **Import** tab. Import properties for the selected file appear on the right, as shown in the next figure,

Figure 9 Import a Local File

The screenshot shows the SAS Data Explorer interface with the 'Import' tab selected. On the left, under 'Import (1)', a file named 'employeeExcel.xls' is listed as a 'Local file'. On the right, the configuration panel for 'employeeExcel.xls' is displayed. It includes fields for 'Target table name' (set to 'employeeExcel') and 'Target location' (set to 'cas-shared-default/Public'). There are checkboxes for 'Save as an in-memory table only' and 'If target table name exists' (with 'Cancel import' selected). A 'Label' field is set to 'Enter label', and the 'Format' is set to 'sashdat'. At the bottom, under 'File Specifications', there are checkboxes for 'Specify a worksheet to import', 'First row contains column names' (checked), and 'Limit the range of imported rows and columns'.

4 The name of the source was copied into the **Target table name** field on the right. Accept or change this name as appropriate for the copy.

If you change the name, follow the SAS conventions for table names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). The target operating system and data source might have additional constraints on these names. For example, the target table name can be a maximum of 247 characters long. However, the name might have to be shorter, due to the data source settings for the caslib that is specified in the **Target location** field. Contact your data administrator for more information about the permitted length for target table names. See also [“Table Names, Column Names, and Special Characters for Imported Tables” on page 28](#).

- 5 A default caslib is provided in the **Target location** field. If this is not where you want to store the file, you can select an existing global caslib or add a new one.

If you have the privilege that is required to add a new global caslib, follow the name conventions described in [“Names for Caslibs, Tables, and Columns” on page 72](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).
- 6 Specify what action the import operation should take if the target filename exists in the caslib that was specified in the **Target location** field. The options are to cancel the import or to replace the existing file that has the same name.
- 7 Specify a label that will help identify the copied table, if desired. The caslib for the target table must support labels. Caslibs of type PATH, DNFS, and HDFS support labels, for example. A caslib for a DBMS does not.
- 8 Use the **Format** selector to select the desired format for the target table. Only sasbdat and CSV formats are supported.
- 9 Use the **File Specifications** tab to specify any options that are needed to accurately copy the contents of the imported file.

Microsoft Excel spreadsheet options:

The following options are available for Microsoft Excel spreadsheets:

Specify sheet to import

For spreadsheets that contain multiple worksheets, indicate the name of the worksheet that you want to import. If you do not specify a worksheet name, then only the first worksheet in the spreadsheet is imported.

First row contains column names

Select this check box when the first row in the file contains column names.

Limit range of imported rows and columns

Select this check box to limit the import operation to a range of cells in the spreadsheet. To use this field, you must also indicate the name of the worksheet in the **Specify sheet to import** field, even if the spreadsheet contains only one worksheet.

Note: If you select both **Limit range of imported rows and columns** and **First row contains column names**, the first row is the first row in the specified range.

Delimited text file options:

The following options are available for delimited text files:

Input file delimiter

Select the delimiter that is used in the file that you want to import. The default value is a comma. If you select **Custom**, you can specify a single character to use as a user-defined delimiter.

Scanned rows

Select the number of rows to scan in order to determine column data types and lengths. A smaller value causes the import to complete quickly, but you increase the possibility of

obtaining a value that is too small to accommodate character columns. A larger value reduces the possibility of truncating character columns, but it increases processing time.

Locale

Enter the locale code for the file. An example is **fr-FR**. Specifying the locale code is important so that special characters (for example, commas) in the file are interpreted correctly based on locale. If you leave this field blank, it defaults to **en-US**.

For examples of locale code values, see [Table of Language Culture Names, Codes, and ISO Values](#).

Encoding

Enter the encoding of the file. When importing UTF-8 or UTF-16 data, make sure that SAS Web Application Server is a Unicode server or that the file contents can be transcoded to the encoding of SAS Web Application Server. Examples of valid values include **utf8**, **utf16le**, and **euc-cn**. The default value is **utf8**.

First row contains column names

Select this check box when the first row in the file contains column names.

Convert character columns to variable size

Select this check box if the file includes columns that contain varying-length character strings. When the file is imported, this option assigns the varying-length character string (VARCHAR) format to all columns that contain character data.

CAUTION

If you select this check box, the VARCHAR format is applied to all columns that contain character data, rather than to only those with varying-length character strings.

SAS data set, SAS7BDAT options:

The following options are available for SAS7BDAT files:

Password

If the file is password-protected, enter the password for the file. If the file is password-protected and you do not enter the password in this field, then the file fails to import.

Encryption key

If the file is encrypted, enter the encryption key for the file. Otherwise, the file fails to import.

Character multiplier

Modify this option to increase the number of characters that can fit in each cell so that character data truncation does not occur. The lengths for character variables can be increased by multiplying the current length by the value that you specify. You can specify a multiplier value from 1 to 5. The default value is 1. If your data contains international characters, and the data is not importing correctly, you might want to specify a higher value.

SAS data set, SASHDAT options:

The following options are available for SASHDAT files:

Encryption key

If the file is encrypted, enter the encryption key for the file. Otherwise, the file fails to import.

- 10** You can use the **Unique ID** control on the **Advanced** tab to add a column to the target table. The new column contains a unique value for each row. A tag is added to the new column. You can view the new column and tag on the **Details** tab for the target table.

For delimited text files, you can use the **Additional Input Options** control on the **Advanced** tab to specify any additional options that might be needed to accurately copy the contents of the imported file. To add options, select **Create additional input options**. Click ⓘ for a description of the options for local delimited text files. Options are entered as name/value pairs, such as **Name:** *containsHeaderRow*, **Value:** **TRUE**. Click **Add Option** to add the option to the import record.

- 11 When ready, you can right-click the file to be copied and select **Import Item**.

If the import succeeds, a copy of the file is loaded to memory on the CAS server that is specified in the caslib. The copy of the file can be selected from the **Available** tab or the **Data Sources** tab.

If the import fails, see [“Usage Notes for Importing Local Files” on page 27](#). See also [“General Usage Notes” on page 71](#).

Use the Excel Spreadsheet (Multiple Worksheets) Option

Use the **Local files** ⇨ **Microsoft Excel (multiple worksheets)** option to import Excel files with multiple worksheets. Each worksheet is loaded to memory as a separate table.

- 1 Display the window that contains the **Import** tab.
See [“Data Selection Windows and SAS Data Explorer” on page 2](#).
- 2 Click the **Import** tab.
- 3 Click **Local files** on the **Import** tab and select **Microsoft Excel (multiple worksheets)**. You are prompted to specify which worksheets to import. You can also change some default import settings.
- 4 When you are ready to import the worksheets, click **Add**.

A job for each worksheet is added to the queue on the **Import** tab. Import properties for the selected worksheet appear on the right, as shown in the following figure.

Figure 10 Import an Excel File with Multiple Worksheets

The screenshot shows the SAS Data Explorer interface with the 'Import (3)' tab active. On the left, three Excel files named 'Color_wksheets (1).xlsx' are listed under the 'Import (3)' section. The right pane shows the configuration for the selected file 'Color_wksheets (1).xlsx'. The 'Target table name' is 'Color_wksheets (1)_Green'. The 'Target location' is 'cas-shared-default/Public'. The 'Save as an in-memory table only' checkbox is unchecked. Under 'If target table name exists', the 'Cancel import' radio button is selected. The 'Label' field is 'Enter label' and the 'Format' is 'sas7dat'. At the bottom, under 'File Specifications', the 'Specify a worksheet to import' checkbox is checked with 'Green' selected, and the 'First row contains column names' checkbox is also checked. The 'Limit the range of imported rows and columns' checkbox is unchecked.

The previous figure shows three import jobs: one for each worksheet in the Excel file `Color_worksheets.xlsx`. Note the path (`/My Folder`) for the Excel file. During the import, the Excel file is staged in your default SAS Folder.

Note: You can delete these staged Excel files from your SAS folder in order to preserve space on SAS Drive.

- 5 The name of the source was copied to the **Target table name** field on the right. Accept or change this name for the copy as appropriate.

If you change the name, follow the SAS conventions for table names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). The target operating system and the data source might have additional constraints on these names. For example, the target table name can be a maximum of 247 characters long. However, a shorter name might be required due to the data source settings for the caslib that is specified in the **Target location** field. Contact your data administrator for more information about the permitted length for target table names. See also [“Table Names, Column Names, and Special Characters for Imported Tables” on page 28](#).

- 6 A default caslib is provided in the **Target location** field. To override the default, select an existing global caslib or add a new one.

If you have the privilege that is required to add a new global caslib, follow the name conventions described in [“Names for Caslibs, Tables, and Columns” on page 72](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).

- 7 Specify what action the import operation should take if the target filename exists in the caslib that was specified in the **Target location** field. The options are to cancel the import or to replace the existing file that has the same name.
- 8 Specify a label that helps identify the copied table, if desired. The caslib for the target table must support labels. For example, caslibs of type PATH, DNFS, and HDFS support labels. A caslib for a DBMS does not support labels.
- 9 Use the **Format** selector to select the desired format for the target table. Only SASHDAT and CSV formats are supported.
- 10 Use the **File Specifications** tab to specify any options that are needed to accurately copy the contents of the imported file.
- 11 You can use the **Unique ID** control on the **Advanced** tab to add a column to the target table. The new column contains a unique value for each row. A tag is added to the new column. You can view the new column and tag on the **Details** tab for the target table.

- 12 When ready, you can right-click one worksheet to be copied and select **Import Item**.

If the import succeeds, a copy of the worksheet is loaded to memory on the CAS server that is specified in the caslib. The copy of the worksheet can be selected from the **Available** tab or the **Data Sources** tab.

If the import fails, see [“Usage Notes for Importing Local Files” on page 27](#). See also [“General Usage Notes” on page 71](#).

Alternatively, you can import all worksheets at once by clicking **Import All** at the top right of the **Import** tab.

Usage Notes for Importing Local Files

All Local Data Files

You might experience long wait-times when you import large local files through a web browser. Accordingly, the CAS Management service option `maxFileUploadSize` is set to 4 Gb by default. If the import of a local file fails because the file size is too large, it might be because you have exceeded the `maxFileUploadSize` value for the CAS server. Your administrator can change the value of this option as appropriate for your site.

To update CAS Management service options, administrators enter *CAS Management* in the **Filter** window for the **Configuration** page of SAS Environment Manager. They open the CAS Management service and find the `maxFileUploadSize` option. For more information, see [“Configuration Page.”](#)

Importing an Excel File with Multiple Worksheets

You might experience long wait times when using the **Local files** ⇒ **Microsoft Excel (multiple worksheets)** option to import large local files through a web browser. Accordingly, the File services option `maxFileSize` is set to 104.8576 megabytes by default. If the import of a local file fails because the file size is too large, it might be because you have exceeded the `maxFileSize` value for the CAS server.

One option is to add the Excel file to the **Import** tab multiple times, and import each worksheet separately.

Alternatively, your administrator can change the value of the `maxFileSize` option as appropriate for your site. To update File services options, administrators enter *File Services* in the **Filter** window for the **Configuration** page of SAS Environment Manager. They open Files services and find the `maxFileSize` option. For more information, see [“Configuration Page.”](#)

Importing Text Files

Here are some key points about importing text files:

- When you import a text file, any data that is enclosed in quotation marks is identified as the VARCHAR type.
- To represent a missing value in the table, you must leave the field empty instead of entering a period (.). Fields that contain a period (.) are imported as a character type.
- If you import a comma-delimited text file with embedded carriage returns and line feeds (CRLFs), set the **readSequential** option to TRUE on the **Advanced** tab of the import properties window. Otherwise, the import might fail.

Importing SAS Data Sets

Here are some key points about importing SAS data sets:

- Importing indexed SAS data sets is not supported.
- Before importing a SAS data set with user-defined formats, ensure that the custom format catalog is available to the CAS server. For more information, see [Manage User-Defined Formats](#).
- If your character data contains international characters, and the data is truncated on import, you might want to specify a higher value in the **Character multiplier** field. The default value of this

field is 1. You can specify a multiplier up to 5 to increase the number of characters that can fit in each cell.

Note: Some formats, such as `char$50`, specify the number of characters that can fit in each cell. If you increase the multiplier in the **Character multiplier** field, you might have to update formats that specify the number of characters that can fit in each cell.

Table Names, Column Names, and Special Characters for Imported Tables

In general, you can import files that use blanks and special characters in the file names and column names. The following list identifies how table names are handled:

- For text files or SAS data sets, the table name is initially set from the file name.
- When you import a spreadsheet, table names are handled as follows:
 - If the spreadsheet contains a single worksheet, then the output table name is initially set as the file name.
 - If the spreadsheet contains multiple worksheets, then the output table name is initially set as the file name for the first worksheet. To indicate a different worksheet, enter the worksheet name in the **Sheet name** field in the Options window for the file.
 - Some special characters can be used, including spaces. Unsupported special characters include / \ * ? " < > | : - , and period (.). If you include unsupported special characters in the output name for a table, the import fails.

If you clear the **Include column names** check box or the **First row contains column names** check box, then the column names are generated as follows:

- **Spreadsheets** Column names are assigned A, B, C, and so on.
- **Text files** Column names are assigned VAR1, VAR2, and so on.

Importing Esri Data for Geoenrichment and Geocoding

Geoenrichment

Overview of Esri GeoEnrichment

The GeoEnrichment Service from Environmental Systems Research Institute (Esri) provides a large collection of data sets, including population, income, housing, consumer behavior, and the natural environment. You can use this service to combine Esri enrichment data with data from a table that you select in your CAS environment. The resulting output table contains new columns of Esri data that are associated with geographic location codes in the source table. For example, you could combine Esri demographic information for the ZIP codes listed in a source table. The resulting output table could then be analyzed in SAS Visual Analytics or other SAS Viya applications.

If your login has appropriate privilege, you can select **Esri** ⇒ **Geo Enrichment** on the **Import** tab. You are prompted to select an in-memory source table in your CAS environment. After you select the source table, a **Geo Enrichment Import** item is added to the queue on the left, as shown in the next figure.

Figure 11 Geo Enrichment Import Item in the Queue on the Import Tab

The screenshot displays the 'Geo Enrichment Import' configuration interface. On the left, the 'Import' tab shows a queue with one item: 'Geo Enrichment Import' (Esri feed). The right pane is titled 'Geo Enrichment Import' and contains the following fields and sections:

- Target table name:** A text field containing 'SAMPLE_ADDRESSES_GEO_ENRICHMENT'.
- Target destination:** A text field containing 'cas-shared-default/Public' with a 'Find' button.
- If target table name exists:** Two radio buttons: 'Cancel import' (selected) and 'Replace file'.
- Label:** A text field containing 'Enter label'.
- Esri Settings:**
 - Source table:** A dropdown menu showing 'SAMPLE_ADDRESSES'.
 - Source column:** A dropdown menu showing 'Postal Code'.
 - Country:** A dropdown menu showing 'United States'.
 - Level:** A dropdown menu showing 'ZIP Codes'.
 - Category:** A dropdown menu showing 'All categories'.
- Available items (7309 of 7309):** A list of categories with expandable arrows: Age, At Risk, Behaviors, Business, and Education.
- Selected items (0):** A section showing 'No items' with a list of items on the right.

Import properties for the **Geo Enrichment** item appear on the right. You specify various options to combine Esri enrichment data with data from the selected source table. Output from the merge is written to the table specified in the **Target table name** field and loaded to memory. The target table is in the caslib that is specified the **Target location** field.

You can use the options under **Esri Settings** to select the type of enrichment data that you want to add to the information from the source table.

When ready, you can right-click the **Esri Import** item in the import queue and select **Import Item**. The specified Esri data is combined with data from the selected source table and is loaded to memory on the CAS server that is specified in the **Target destination** field.

Alternatively, you can create multiple import items and import them all at once by clicking **Import All** at the top right of the **Import** tab. The target table names in all import items must be unique in order for the **Import All** option to work.

Note: Any items in the queue on the **Import** tab are dropped from the queue when you log off. The target tables that you load to memory persist beyond the current session. The tables are associated with a global caslib in the import properties window. For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).

Prerequisites for Geoenrichment

Perform the following tasks before using the **Geoenrichment** option:

- 1 Ask an administrator to add your login to the custom group **Esri Users** in SAS Environment Manager. For information about custom groups, administrators can refer to [“Users Page.”](#)
- 2 Obtain the user name and password for the Esri GeoEnrichment Service. Verify these credentials by logging in from the **Sign In** link on the Esri website [esri.com](#).
- 3 Access the Settings window for your application by selecting **Settings** at the top right of the application banner.
- 4 Expand the **Data Explorer** section.
- 5 Click **Geographic Mapping**.
If you have been added to the **Esri Users** group in SAS Environment Manager, two check boxes are displayed: **Accept Esri ArcGIS Online Services terms and conditions** and **Enable Esri premium services**.
- 6 Select the **Accept Esri ArcGIS Online Services terms and conditions** check box.
- 7 Select the **Enable Esri premium services** check box.
- 8 Specify the user name and password for the ArcGIS World Geocoding Service.
- 9 Identify the caslib and name of the source table to which you will add Esri data. The source table for a **Geoenrichment Import** item must be an in-memory table.
- 10 Identify the caslib and name of the target table that will combine Esri data with data from the selected source table.
- 11 Identify the geographic level that you want to analyze such as by state or ZIP code.
- 12 Identify the Esri enrichment data that will support the analysis that you want to perform on the output table. For example, you might want to analyze demographic data, business data, or landscape data by geographic location. For more information about Esri enrichment data, see [The Esri GeoEnrichment Service](#).

Import Esri Data for Geoenrichment

After you have done the prerequisite tasks, perform the following steps to combine Esri enrichment data with data from an in-memory table on a CAS server:

- 1 On the **Import** tab, expand the **Esri** folder and select **Geoenrichment**.
You are prompted to select a source table in your CAS environment. Only in-memory tables are available for selection.
- 2 Select a source table.

After you select the source table, a **Geo Enrichment Import** item is added to the queue on the left. Import properties appear on the right.

- 3 The default **Target table name** is the source table name with **_GEO_ENRICHMENT** added to the end. This is the name of the output table that will combine Esri enrichment data with data from the selected source table. Accept or change this name as appropriate.

If you change the name, follow the conventions for table names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). The target operating system and data source might have additional constraints on these names. For example, the target table name can be a maximum of 247 characters long. However, the name might have to be shorter, due to the data source settings for the caslib that is specified in the **Target location** field. Contact your data administrator for more information about the permitted length for target table names.

- 4 A default caslib for the target table is provided in the **Target location** field. If this is not where you want to store the output table, you can select an existing global caslib or add a new one.

If you have the privilege that is required to add a new global caslib, follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).

- 5 Specify what action the import operation should take if the target filename exists in the caslib specified in the **Target location** field. The options are to cancel the import or to replace the existing item that has the same name.
- 6 Specify a label that will help identify the output table, if desired. The caslib for the target table must support labels. Caslibs of type PATH, DNFS, and HDFS support labels, for example. A caslib for a DBMS does not.
- 7 Use the options under **Esri Settings** to select Esri enrichment data that will be combined with the data in the source table.

Source table

The in-memory table whose data is combined with Esri enrichment data.

Source column

Select the column in the source table that specifies geographic location codes, such as a column for ZIP codes.

Country

Select the country where the geographic location codes in the **Source column** are valid.

Level

Select the Esri geographic level that corresponds to the geographic location codes in the **Source column** above. Levels include **States**, **Census Tracts**, and **ZIP Codes**.

For example, suppose that the source table has a column named ZIP_Code, and you want to add Esri demographic information for the ZIP codes listed in the source table. You could select **ZIP_Code** in the **Source column** field and select the Esri level that is named **ZIP Codes** in the **Level** field.

Category

Select a category of Esri enrichment data. Categories include **Business**, **Households**, and **Population**.

For example, suppose that you want to add Esri demographic information for ZIP codes, as specified in the **Level** field. You could select **Population** in the **Category** field.

Available items

Select items within the category selected in the **Category** field. For example, the **Population** category includes items for **Gender** and **Language**. Click the arrow to move an available item to the **Selected items** panel.

Selected items

Enrichment items that will be combined with information from the source table.

For example, if the **Gender** item appears in this panel, Esri gender information is combined with information from the source table. Gender information would be provided by the geographic level that is specified in the **Level** field, such as **ZIP Codes**. To save the credits that are consumed for the data, no more than 10 items are selected for a single import operation.

- 8 If you want to add more than one Esri category of information, go back to the **Category** field and add another category and items for that category.

- 9 When ready, you can right-click the **Esri Import** item in the queue and select **Import Item**.

A confirmation dialog box displays an estimated number of Esri credits that this import costs.

- 10 Click **Yes** to continue or **No** to cancel the import.

Esri enrichment data is combined with the data in the source table and written to the output table. The output table is loaded to memory on the CAS server that is specified in the **Target location** field. The output table can be selected from the **Available** tab or the **Data Sources** tab.

If the import fails, see [“Usage Notes for Importing Local Files” on page 27](#).

Geocoding

Overview of Esri Geocoding

The **Geocode** option on the **Import** tab enables you to add geographic coordinates that match entities in your data. To use this option, you select one or more columns of address information from a table. The selected addresses are provided to the ArcGIS World Geocoding Service from Environmental Systems Research Institute (Esri). The resulting output table contains latitude and longitude coordinates for each row in your data so that it can be plotted on a map.

If your login has appropriate privilege, you can select **Esri** ⇒ **Geocode** on the **Import** tab. You are prompted to select an in-memory source table in your CAS environment. After you select the source table, a **Geocode Import** item is added to the queue on the left, as shown in the next figure.

Figure 12 Geocode Import Item in the Queue on the Import Tab

The screenshot displays the 'Geocode Import' interface. On the left, the 'Import' tab is active, showing a queue with one item: 'Geocode Import cas-shared-default/Public'. On the right, the configuration panel for this item is shown. It includes fields for 'Target table name' (SAMPLE_ADDRESSES_GEO_CODE) and 'Target destination' (cas-shared-default/Public). Below these, there are radio buttons for 'If target table name exists' (Cancel import is selected). A 'Label' field is also present. The 'Geocode Settings' section shows the 'Source table' as 'SAMPLE_ADDRESSES'. It lists 'Available columns (0 of 5)' (City, State/Province, Postal Code, Country) and 'Selected columns (5)' (Postal Code, Street Address, City, State/Province, Country). A 'Sample address' field contains '24253 5270 Lectus. St. Evansville Florida United States', and below it, the geocoded coordinates and score are shown: 'x=-87.56 y=37.98 score=79'. A 'Test' button is located at the bottom of the configuration panel.

Import properties for the **Geocode** item appear on the right. You specify various options to combine Esri geocode data with data from the selected source table. Output from the merge is written to the table specified in the **Target table name** field and loaded to memory. The target table is in the caslib that is specified the **Target location** field.

You can use the options under **Geocode Settings** to select one or more columns in the source table that contain address information. After you select the address columns, you can click **Test** to generate latitude and longitude coordinates for a sample record in the source table. These coordinates are based on address information that you provided.

When ready, you can right-click the **Geocode Import** item in the import queue and select **Import Item**. The geocode data is combined with data from the selected source table and is loaded to memory on the CAS server that is specified in the **Target destination** field.

Alternatively, you can create multiple import items and import them all at once by clicking **Import All** at the top right of the **Import** tab. The target table names in all import items must be unique in order for the **Import All** option to work.

Note: Any items in the queue on the **Import** tab are dropped from the queue when you log off. The target tables that you load to memory persist beyond the current session. The tables are associated with a global caslib in the import properties window. For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).

Prerequisites for Geocoding

Perform the following tasks before using the **Geocode** option:

- 1 Ask an administrator to add your login to the custom group **Esri Users** in SAS Environment Manager. For information about custom groups, administrators can refer to [“Users Page.”](#)
- 2 Obtain the user name and password for the ArcGIS World Geocoding Service. Verify these credentials by logging in from the **Sign In** link on the Esri website [esri.com](#).
- 3 Access the Settings window for your application: select **Settings** at the top right of the application banner.
- 4 Expand the **Data Explorer** section.
- 5 Click **Geographic Mapping**.
If you have been added to the **Esri Users** group in SAS Environment Manager, two check boxes are displayed: **Accept Esri ArcGIS Online Services terms and conditions** and **Enable Esri premium services**.
- 6 Select the **Accept Esri ArcGIS Online Services terms and conditions** check box.
- 7 Select the **Enable Esri premium services** check box.
- 8 Specify the user name and password for the ArcGIS World Geocoding Service.
- 9 Identify the caslib and name of the source table to which you will add Esri data. The source table for an **Geocode Import** item must be an in-memory table.
- 10 Identify the caslib and name of the target table that will combine Esri data with data from the selected source table.

For more information about the ArcGIS World Geocoding Service, see [Geocoding with ArcGIS](#).

Import Esri Data for Geocoding

After you have performed the prerequisite tasks above, perform the following steps to combine Esri geocoding data with data from an in-memory table on a CAS server:

- 1 On the **Import** tab, expand the **Esri** folder and select **Geocode**.
You are prompted to select a source table in your CAS environment. Only in-memory tables are available for selection.
- 2 Select a source table.
After you select the source table, a **Geocode Import** item is added to the queue on the left. Import properties appear on the right.
- 3 The default **Target table name** is the source table name with **_GEO_CODE** added to the end. This is the name of the output table that will combine Esri geocode data with data from the selected source table. Accept or change this name as appropriate.

If you change the name, follow the conventions for table names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). The target operating system and data source might have additional constraints on these names. For example, the target table name can be a maximum of 247 characters long. However, the name might have to be shorter due to the data source settings for the caslib that is specified in the **Target location** field. Contact your data administrator for more information about the permitted length for target table names.

- 4 A default caslib for the target table is provided in the **Target location** field. To store the output table elsewhere, you can select an existing global caslib or add a new one.

If you have the privilege that is required to add a new global caslib, follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).

- 5 Specify what action the import operation should take if the target filename exists in the caslib that is specified in the **Target location** field. The options are to cancel the import or to replace the existing item that has the same name.
- 6 (Optional) Specify a label that will help identify the output table. The caslib for the target table must support labels. Caslibs of type PATH, DNFS, and HDFS support labels, for example. A caslib for a DBMS does not support labels.
- 7 On the **Geocode Settings** tab, the **Available columns** panel lists the columns that are available in the source table. Select one or more columns in the source table that contain address information.
- 8 Use the arrow controls to move selected columns from the **Available columns** panel to the **Selected columns** panel on the right.
- 9 (Optional) You can click **Test** to generate latitude and longitude coordinates for a sample record in the source table. These coordinates are based on address information that you provided.
- 10 When ready, you can right-click the **Geocode Import** item in the queue and select **Import Item**.
A confirmation dialog box displays an estimated number of Esri credits that this import costs.
- 11 Click **Yes** to continue or **No** to cancel the import.

Esri geocoding data is combined with the data in the source table and written to the output table. The output table is loaded to memory on the CAS server that is specified in the **Target location** field. The output table can be selected from the **Available** tab or the **Data Sources** tab.

If the import fails, see [“Usage Notes for Importing Local Files” on page 27](#).

Working with Data in CAS

Working with Caslibs and Tables

Working with Caslibs

When you select a caslib on the **Data Sources** tab, properties for that caslib are displayed on the right. These properties were specified when the caslib was added, such as the CAS server, the source type, and any advanced options. When you right-click a caslib, the following options are available.

Table 1 Pop-Up Menu Options for a Selected Caslib

Name	Action
Set as default target location	Makes the selected caslib the default location for various operations, including operations on the Import tab. This option is available only for global caslibs.
Import from library ⇌ Delimited files...	Imports multiple delimited files, documents, or images from a directory in the selected caslib. This option is available only for PATH caslibs. For more information about these options, see the following topics: <ul style="list-style-type: none"> ■ “Importing Delimited Files from a Caslib Directory” on page 55 ■ “Importing Documents from a Caslib Directory” on page 58 ■ “Importing Images from a Caslib Directory” on page 62
View authorization	If available, displays the permissions that each user or group has for the selected caslib. See “CAS Authorization Window” in SAS Viya: CAS Authorization Window .
Edit authorization	If available, enables you to change the permissions that each user or group has for the selected caslib. The ManageAccess permission is required to modify access controls. See “CAS Authorization Window” in SAS Viya: CAS Authorization Window .
Edit	If available, enables you to change the physical path and description for PATH, HDFS, and DNFS caslibs. This option is available if you have AlterCaslib permission for the selected caslib.

Name	Action
Remove	Removes the selected caslib from memory on a CAS server.

Working with Tables

Details Tab, Sample Data Tab, and Profile Tab

When you select a table in the **Available** tab or the **Data Sources** tab, three tabs of information about this table are displayed on the right side:


- **Details.** Displays basic information about the table, such as the number of columns and rows and the data type of columns. You can also use this tab to manage tags for tables and columns. For more information, see [“Manage Tags from the Details Tab” on page 39](#).
- **Sample Data.** Displays sample data from the selected table. If the sample data includes images, click an image to preview it. Some image formats cannot be displayed by some browsers unless they are decompressed. Other formats might not be displayable unless the browser is configured to launch an external viewer. The overflow menu beside **Sample rows** enables you to select **Show labels** or **Show names** for columns in the table.
- **Profile.** Displays a report that enables you to recognize data patterns, to identify scarcity in the data, and to review basic statistics for the selected table. Data profiles can be generated only for tables in a global caslib. For more information, see [“Profiling Data” on page 42](#).

Pop-up Menu Options

If you right-click a table in the **Available** tab or the **Data Sources** tab, some of the options in the next table will be available, as appropriate for the context.

Table 2 Pop-up Menu Options for a Selected Table on the Available Tab or Data Sources Tab

Name	Action
Actions ⇒ Prepare Data ...	Opens the selected table in another application such as SAS Data Studio or SAS Visual Analytics. For more information, see the documentation for that application.
Load	Loads the selected table to memory on a CAS server. The Load option is available only for tables in a global caslib. For more information about global caslibs, see “Caslibs on the Data Sources Tab and Import Tab” on page 71 .
Unload	Unloads the selected table from memory on a CAS server. The table must be loaded again if you want to work with it in CAS.
View authorization	If available, displays the permissions that each user or group has for the selected table. See “CAS Authorization Window” in SAS Viya: CAS Authorization Window .

Name	Action
Edit authorization	<p>If available, enables you to change the permissions that each user or group has for the selected table. The ManageAccess permission is required to modify access controls. See “CAS Authorization Window” in SAS Viya: CAS Authorization Window.</p>
Delete	<p>For in-memory tables: Delete unloads the table from memory and removes its associated physical table.</p> <p>For physical tables: Delete removes the physical table.</p> <p>Refresh the view to verify your changes. Select  from the toolbar.</p> <p>Note: SAS Data Explorer might fail to delete a SAS data set (.sas7bdat file) if the file was created on a different platform than the one that the CAS server runs on. For example, your attempt to delete a Windows file from a CAS server that runs on Linux might fail. In this case, you can use the DATASETS procedure to delete the data set.</p>
Add to import	<p>Adds the selected table to the queue that is managed on the Import tab. You might want to do this to create a copy of a table on the Available tab or the Data Sources tab. For more information, see “Add to Import from the Data Sources Tab or Available Tab” on page 50.</p>
Run profile	<p>Generates a data profile for the selected table and displays the report on the Profile tab. The report enables you to identify anomalies or inconsistencies in the data. Data profiles can be generated only for tables in a global caslib. For more information, see “Profiling Data” on page 42.</p>
Run profile and save	<p>Generates a data profile for the selected table and saves the results to another table. A programmer can generate a report from the results table. This option is available only when your site licenses the SAS Data Preparation offering. For more information, see “Run Profile and Save” on page 46.</p>
Download table	<p>If available, downloads the selected in-memory table to your computer and saves it as a delimited file. You can select CSV or TSV format. Administrators can limit access to this option. See “Limit Access to the Download Table Option” on page 68.</p> <p>Note: The selected table must contain data that can be saved to CSV or TSV format. For example, you cannot download a binary image file to a CSV file and get useful data.</p>

If you right-click a table or file in the queue on the **Import** tab, and the **Create job** option is available, you can use this option to create an import job. The job can be scheduled or executed in SAS Environment Manager. For more information, see [“Create Import Job Requests” on page 54](#).

Managing Tags for Tables and Columns

Overview of Tags

SAS Viya applications enable you to add tags that identify the type of information in a table or column. For example, you could add a **Street Address** tag to a column of street addresses. Other applications can use these tags to analyze the data in tables.

You can add tags to tables and columns in the following ways:

- Use the **Details** tab for a table. See [“Manage Tags from the Details Tab” on page 39](#).
- Use the **Unique ID** control for local files that are imported on the **Import** tab. This control enables you to add a column to the target table that contains a unique value for each row. A tag is added to the new column. See [“Importing Local Files” on page 21](#).
- Select the auto-tagging option for data profiles. This option enables column content analysis during data profiling. If the analysis can determine what type of content is in the column, an appropriate tag is added to the column. See [“Enable Automatic Content Tagging for Columns in a Table” on page 43](#).
- Use the Unique ID transform or the Partition transform in SAS Data Studio. These transforms add a special column and add a tag to that column. They add the same tag to the table. See [“Generating a Unique Identifier” in SAS Data Studio: User’s Guide](#) and [“Creating a Partition Column” in SAS Data Studio: User’s Guide](#).

Manage Tags from the Details Tab

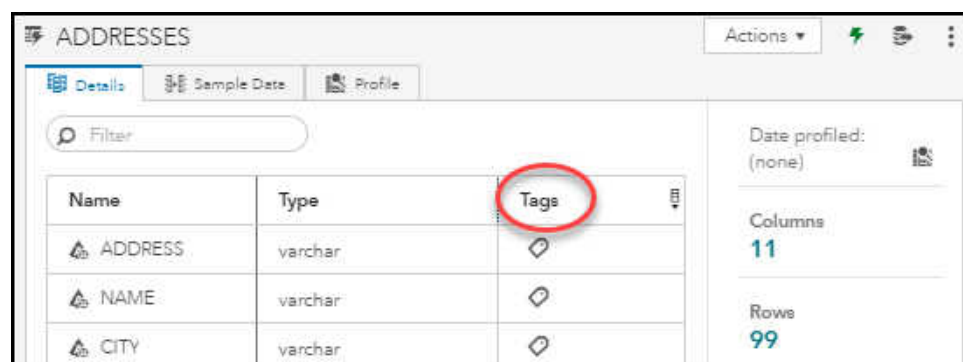
Manage Tags for a Column

Perform the following steps to manage tags for a column:

- 1 Select a table on the **Available** tab or the **Data Sources** tab.

The **Details** tab displays information about the table, including a **Tags** column on the far right. The **Tags** column displays a tag icon (◇) for each column in the table.

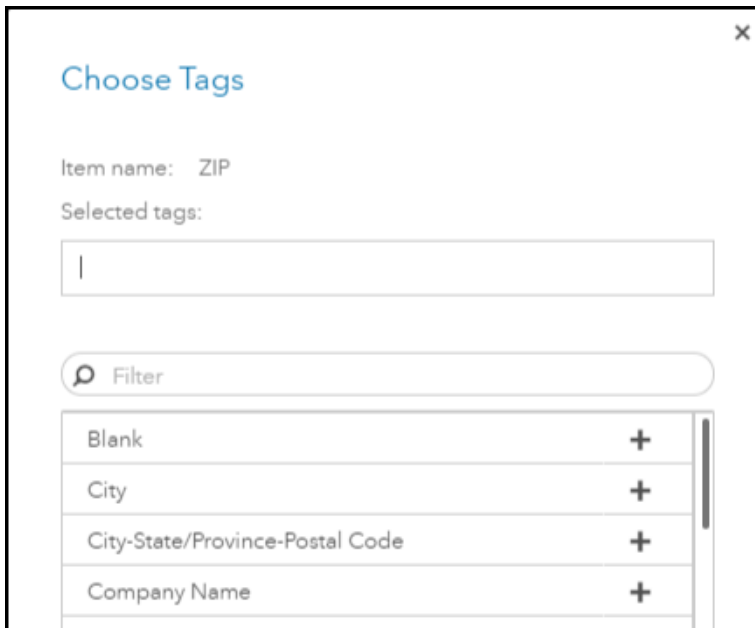
Figure 13 Tags Column on the Details Tab



- 2 Click  next to a column to manage tags for that column.

The **Choose Tags** window is displayed. The next figure shows the **Choose Tags** window for a ZIP code column.

Figure 14 Choose Tags Window for a ZIP Code Column



The name of the selected column appears in the **Item name** field. Any tags that have been added to this column appear in the **Selected tags** field. A number of predefined tags, such as **Blank** and **City**, are provided with the software and are available under the **Filter** window.

- 3 To add an existing tag to the column, click **+** beside the tag. Click **OK** to save your changes.

To remove a tag from a column, click the **X** beside the tag. Click **OK** to save your changes.

To filter the list of available tags, enter letters or words into the **Filter** window.

To add a new tag, enter the name of the tag into the **Selected tags** field and press **Enter**. Click **OK** to save your changes.

Administrators with appropriate privilege can use the [SAS Viya REST API for Annotations](#) to delete or rename a tag in the list of available tags in the **Choose Tags** window.

Manage Tags for a Table

Perform the following steps to manage tags for a table:


- 1 Select a table on the **Available** tab or the **Data Sources** tab.

The **Details** tab displays information about the table, including a **Tags** section at the lower right of the tab.

Figure 15 Tags Section at Lower Right on the Details Tab

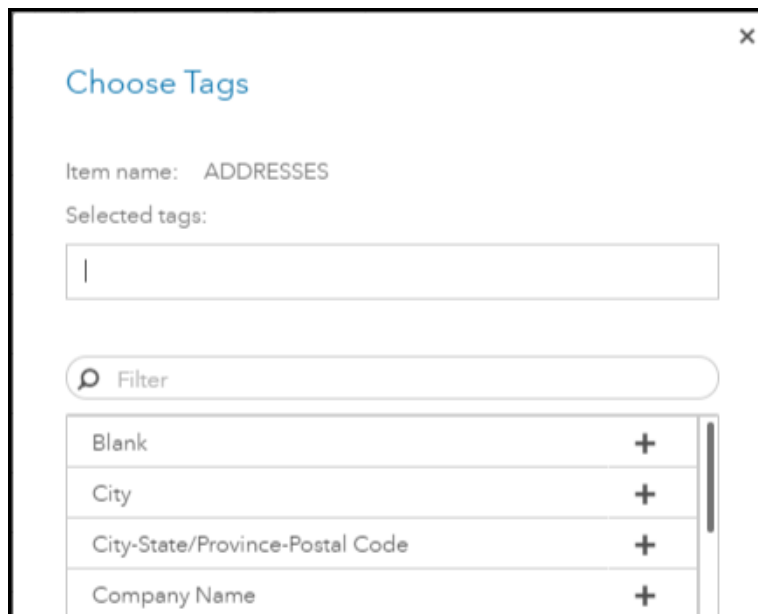


This section displays any tags that have been added to the selected table. It does not display tags that have been added to columns in the table.

- 2 Click  in the **Tags** section to manage tags for the table.

The **Choose Tags** window is displayed. The next figure shows the **Choose Tags** window for a table named ADDRESSES.

Figure 16 Choose Tags Window for the ADDRESSES Table



The name of the selected table appears in the **Item name** field. Any tags that have been added to this table appear in the **Selected tags** field.

A number of predefined tags, such as **Blank** and **City**, are provided with the software and are available under the **Filter** window.

- 3 To add an existing tag to the table, click the plus sign beside the tag. Click **OK** to save your changes.

To remove a tag from a table, click the **X** beside the tag. Click **OK** to save your changes.

To filter the list of available tags, enter letters or words into the **Filter** window.

To add a new tag, enter the name of the tag into the **Selected tags** field and press **Enter**. Click **OK** to save your changes.

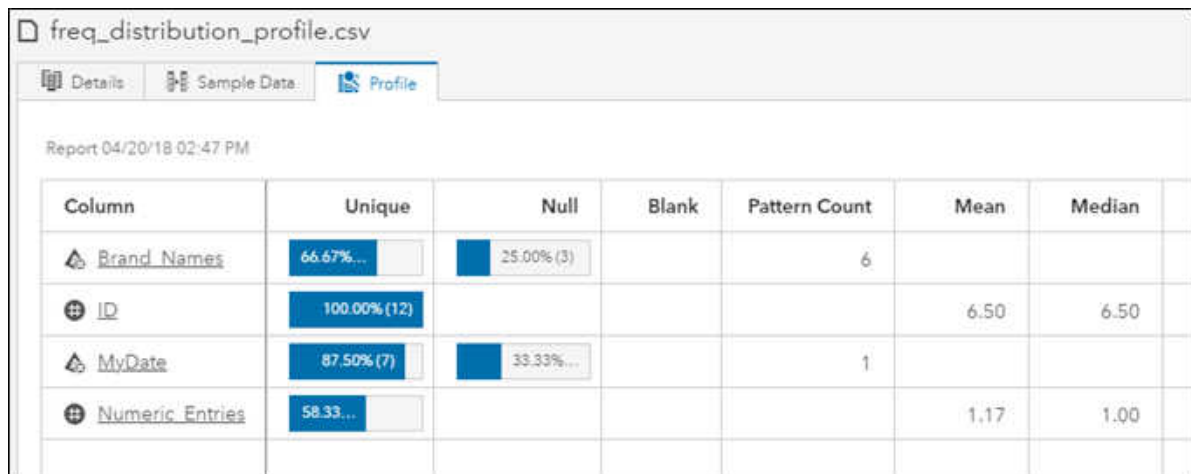
Administrators with appropriate privilege can use the [SAS Viya REST API for Annotations](#) to delete or rename a tag in the list of available tags in the **Choose Tags** window.

Profiling Data

Overview of Data Profiles

A data profile report enables you to identify anomalies or inconsistencies in the data for a selected table. For example, the profile in the next figure shows that some columns have null values, and all of the values in one column are unique.

Figure 17 Data Profile Report



If you right-click a table that is associated with a global caslib, you can choose between two profiling options:

- **Run profile** loads the selected table into memory if it is not already loaded, generates a data profile for the table, and displays the report on the **Profile** tab. Unloaded tables are returned to an unloaded state after the profile completes.
- **Run profile and save** generates a data profile for the selected table and saves the results to another table. A programmer can generate a report from the results table. This option is available only when your site licenses the SAS Data Preparation offering.

You can select a global setting that applies formats to the output data for the **Run profile** option. Some data is more meaningful when it is formatted. If formatting is not enabled, profiling could fail for certain data, such as hexadecimal data. For more information about the apply formats option for profiles, see [“Modify SAS Data Explorer Settings” on page 69](#).

If your site has licensed the SAS Data Preparation offering, your profile reports include advanced metrics. These metrics are described in [“Understanding Data Profile Reports” on page 44](#). You can also select a global setting that triggers column content analysis during profiling, as described in the next section.

Enable Automatic Content Tagging for Columns in a Table

You can enable column content analysis during data profiling. If the analysis can determine what type of content is in the column, an appropriate tag is added to the column. For example, a column that contains street addresses might get a **Street Address** tag. Other applications can use these tags to analyze the data in tables. Any tags that are added can be viewed on the **Details** tab for the profiled table. For more information, see [“Manage Tags from the Details Tab” on page 39](#).

To enable column content analysis during data profiling, the following prerequisites must be met:

- The SAS Data Preparation offering must be licensed.
- SAS Data Quality 3.4 or later must be installed. This software is included with SAS Data Preparation 2.2 or later.
- SAS Quality Knowledge Base (QKB) for Contact Information 29 must be installed. If your site has licensed SAS Data Preparation, a QKB is typically installed with the rest of the software. A QKB can also be installed later, as described in see [“Set the Default QKB and the Default Locale” in SAS Viya Administration: QKB Management](#).
- You must set the **Default locale for Quality Knowledge Base** option and the **Analyze column contents while running profile** option. These settings are described in [“Modify SAS Data Explorer Settings” on page 69](#).

Note the following about automatic column content tagging:

- The table being profiled must be an in-memory table.
- Columns with a data type of CHAR or VARCHAR are analyzed for content. Other columns are ignored.
- Columns with an **ID Analysis** value of UNKNOWN do not get a content tag. For more information about this value, see [“Profile Report for a Column” on page 45](#).

Run Profile

Perform the following steps to profile a selected table and display the report on the **Profile** tab.

- 1 Select a table in a global caslib on the **Available** tab or the **Data Sources** tab.
Three tabs of information about this table are displayed: **Details**, **Sample Data**, and **Profile**.
- 2 Click the **Profile** tab.

If a profile was previously generated for the table, the most recent report is displayed. If a profile is not available, click **Run Profile**. The next figure shows a data profile.

Figure 18 Data Profile for a Selected Table

Column	Unique	Null	Blank	Pattern Count
Failure Date	47%	0	0	4
Institution Name	91% (3185)	0	0	1,524

Click a column to see statistics for that column. You can expand three reports for the column: **Descriptive Metrics**, **Pattern Distribution**, and **Frequency Distribution**. For details about the statistics in a profile report, see [“Understanding Data Profile Reports” on page 44](#).

- Notice the buttons at the top right corner of the profile report: (Options) and (Version).
 - By default, columns for all statistics are included in the profile. To remove some statistics from the profile, click and select **Manage columns**. Use the arrows in the window to remove some statistics from the profile. Click **OK** to display the revised profile.
 - Each time you profile a table, a version of the report is saved. Click to view versions of the profile report. Click to remove outdated profile reports.

Understanding Data Profile Reports

Profile Report for a Table

The data profile report for a table includes the following statistics by default:

Column

Name of the column. This is a fixed-width character value. SAS and CAS treat all blank fixed-width character columns as missing (null), so this field always reports zero blank values.

Unique

Percentage of unique values. The count of unique values appears in parenthesis.

Mean

Mean value of your numeric data content. This value is calculated by dividing the sum of all the numbers by the total count of numbers.

Standard Deviation

Standard deviation of your numeric content. The standard deviation measures the spread of the data about the mean value. It is useful in comparing sets of data that might have the same mean but a different range.

Standard Error

Mean standard error of your numeric data. The standard error is the standard deviation of the sampling distribution of a statistic. Thus, the standard error of the mean is the standard deviation of the sampling distribution of the mean.

Minimum

Minimum value in your data.

Maximum

Maximum value in your data.

Data Type

Type of your data as it was created in the design of your table. This is a variable-width character value. SAS and CAS treat all blank variable-width character columns as missing (null). However, in addition to counting missing columns as null, nonzero length blank values are also counted as blanks.

Data Length

Length of your data as it was created in the design of your table. This is retrieved from table metadata.

If your site has licensed the SAS Data Preparation offering, additional statistics are generated. If your site has not licensed the SAS Data Preparation offering, your profile results might have blank values for these additional statistics:

Actual Type

Actual type of a variable from its SQL type. For example, VARCHAR is seen as a string and CURRENCY is seen as numeric.

Median

Median value out of your numeric content. The median is the middle of a distribution: half the scores are above the median and half are below the median.

Mode

Mode value of your numeric content. The number of times the value occurs, or -1 if multiple values have equal number of occurrences.

Null

Percentage of null values.

Blank

Percentage of blank values.

Frequency Distribution

Unique values for a column, including information about how frequently a value occurs in the table.

Pattern Count

Count of unique word or character patterns in your data.

Minimum Length

The value of the minimum length. This applies only to string types.

Maximum Length

The value of the maximum length. This applies only to string types.

Ordinal Position

Actual location of a field within the table. This is retrieved from table metadata.

Primary Key Candidate

If the value is nonzero, the column is a primary key candidate.

Non-Null Count

Number of values that are not null.


Profile Report for a Column

Click a column in the **Profile** tab to see statistics for that column. You can expand three reports for the column: **Descriptive Metrics**, **Pattern Distribution**, and **Frequency Distribution**. See the previous section for details about these statistics. The **ID Analysis** score for a column is the score for

the specified value based on the frequency distribution of the identified types for this column. A score of 200 is the top score. Multiple scores for a single column should add up to 200.

Run Profile and Save

Perform the following steps to profile a selected table and save the results to another table. A programmer can generate a report from the results table. This option is available only when your site licenses the SAS Data Preparation offering.

- 1 Right-click a table in a global caslib on the **Available** tab or the **Data Sources** tab and select **Run profile and save**. The Run Profile and Save Results dialog box is displayed.
- 2 Specify the name of a target table for the profile results.
- 3 Select a caslib for the target table. Only global caslibs on the CAS server where the profiled table is stored are available.
- 4 Click **Run**. The source table is profiled, and the results are written to the target table.
- 5 Navigate to the caslib for the target table.
- 6 Click  to display the target table.

The target table is an in-memory table. It persists until you unload it or the CAS server is restarted. The **Run profile and save** option does not create a physical table.

A programmer can generate a report from the results table. For a description of the table layout, see [Output Data Descriptions](#).

Usage Notes for Data Profiles

Data Profiles and Global Caslibs

Data profiling options are available only for tables in global caslibs. For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).

Performance Tips

When the table exceeds 200,000 records, run the profile on a sample of the records.

Activate frequency distribution statistics only on those columns where these statistics are important.

Frequency Distribution and Pattern Metrics Might Be Truncated

When you profile a table, the length of values for the frequency distribution and pattern metrics are truncated to 1000 characters if the original value exceeds 1000 characters.

Profiling Multiple Files with the Same Name but Different Extensions

If one directory contains multiple files with the same name but different extensions, the CAS Management Service decides which one is preferred. The SASHDAT version might be preferred over the CSV version, for example. In this case, only the file with the preferred extension can be profiled. If

you want to profile multiple files with the same name but different extensions, put the files with different extensions in different directories.

Rerun Profile Reports after an Upgrade-in-Place

If you perform an upgrade-in-place of SAS Data Quality 3.2 or 3.3 to SAS Data Quality 3.4, profile percentage metrics including nulls, uniqueness, and patterns have blank values. Tables must be profiled again to get these metrics.

Data Profile Reports Are Not Transferred during Migration from One SAS Viya Environment to Another

Data profile reports are not transferred during a migration from one SAS Viya environment to another. Data sets on the target system might differ from data sets on the source system. Accordingly, data profile reports generated on the source system might be incorrect for the target system. If you want to view data profile reports for a data set on your target system, profile the data set on the target system.

Microsoft Windows and SAS Data Quality

Some features in SAS Data Explorer depend on SAS Data Quality software. SAS Data Quality software is not available when SAS Viya is running on Microsoft Windows. Accordingly, the features that depend on QKBs or other SAS Data Quality software are not available when SAS Viya is running on Microsoft Windows. This includes advanced data profiling metrics and column content analysis.

Delay in Execution of a Data Quality Operation

The first data quality operation that is performed after CAS starts might take longer than normal to execute. If this delay becomes an issue at your site, ask your CAS administrator to consider the following remedy.

Data quality operations require access to the CAS table that contains the QKB on all workers in the analytics cluster. Loading this table takes longer for the first data quality operation after CAS starts because the table has not been loaded before. (Subsequent loads take less time.) CAS automatically loads this table when it starts. However, CAS tables for all non-default QKBs are not loaded automatically.

If CAS sessions use non-default QKBs, the CAS administrator can load these non-default QKBs during CAS start-up. In the CAS configuration file, `casstartup_usermods.lua`, the administrator can add a `qkb.loadQKB()` action for each non-default QKB.

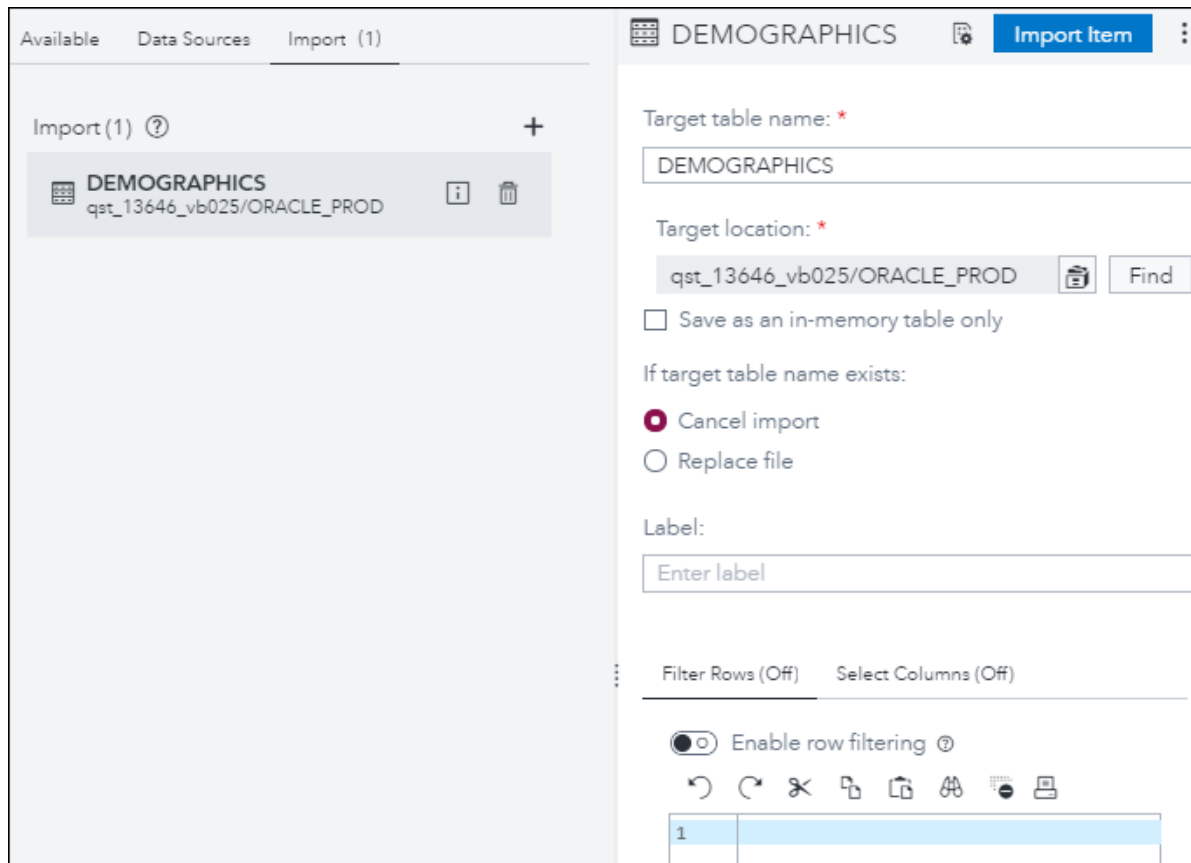
Importing Data from the Available Tab or Data Sources Tab

Overview of the Add to Import Option

The **Add to import** option enables you to copy a table or a file on the **Available** tab or the **Data Sources** tab to a global caslib. You can copy a table or a file from any caslib to any global caslib to which you have access. You can use caslibs to copy a table from one DBMS to another, for example.

When you right-click a table or file that you want to copy, and select **Add to import**, the table or file is added to the queue on the **Import** tab. For example, in the next figure, a physical table has been added to the import queue on the left side.

Figure 19 Physical Table in the Queue on the Import Tab



Note the **Target table name** field and the other import properties on the right. These properties apply to the copy of the table or file. The main properties are described below.

The **Target table name** field is usually set to the source table name. Accept or change this name as appropriate for the copy.

If the source caslib is global in scope, its name is copied into the **Target location** field. You can accept or change this caslib as needed.

The **Save as in-memory table only** check box enables you to load the imported table to memory without saving a physical copy of the table to the target destination. Only users who have the Promote permission on the target caslib can successfully perform an import using this option. Use this option to work with a table without the costs in time and storage space that are required to write a physical copy. This option offers a significant performance gain for small tables.

A **Format** selector for the target table is available when certain types of caslibs are selected in the **Target location** field.

Figure 20 Format Selector for a Target Table

The **Format** selector enables you to copy the source table to a target table in a different format. For more information, see [“Copying Tables in One Format to Another Format” on page 51](#).

Note the **Filter Rows** tab and **Select Columns** tab in the figure [Figure 19 on page 48](#). These tabs are available for physical tables or files only, not for in-memory tables or files. You can use these tabs to create a copy that contains a subset of columns or rows from the original.

Some files and tables have an **Advanced** tab with an **Additional Input Options** control. This control enables you to specify additional options that might be needed to accurately copy the contents of the imported file. Click ⓘ for a description of the valid options. For more information about import options, see [Common Parameter: importOptions](#).

When ready, you can right-click the table or file in the import queue and select **Import Item**. A copy of the table or file is loaded to memory on the CAS server that is specified in the target caslib. A physical copy of the data is also created in the target caslib.

The **Import Item** option copies the table or file in the current session. For large items, you might want to schedule the import after business hours to minimize the impact on production systems. If you right-click the item in the import queue, and the **Create job** option is available, you can use this option to create an import job. The import job can be scheduled or executed in SAS Environment Manager. For more information about this option, see [“Create Import Job Requests” on page 54](#).

Note: Any items in the queue on the **Import** tab are dropped from the queue when you log off. The target tables or files that you load to memory persist beyond the current session. The targets are associated with a global caslib in the import properties. For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).

Prerequisites for the Add to Import Option

Gather the following information to use the **Add to import** option:

- 1 Identify the source table or file on the **Available** tab or the **Data Sources** tab that you want to copy. Note the size of the item that you want to import. Importing a large table or file can be time-consuming.
- 2 If the source caslib is global in scope, its name is copied into the **Target location** field. If this is not where you want to store the copy, you can select a different global caslib or add a new one. If you want to copy sensitive data to a different caslib, make sure the target caslib and file have appropriate protections. See [“Copy Data from One Caslib to Another” on page 53](#).

- 3 The name of the source table or file is copied into the **Target table name** field. Change the target table name, if desired. The target table name cannot be identical to the source table name unless you change the caslib or select the **Replace file** option.
- 4 If you want to copy a source table to a target table in another format, identify an appropriate format and target caslib.
- 5 For physical tables or files only, you can create a copy that has a subset of columns or rows from the original. Plan which rows or columns would be appropriate for the copy. For more information, see [“Filter Rows from an Imported Table or File” on page 53](#) or [“Select Columns from an Imported Table or File” on page 53](#).

Add to Import from the Data Sources Tab or Available Tab

Use information that you previously gathered to copy a table or a file on the **Available** tab or the **Data Sources** tab:

- 1 Display the window that contains the **Available** tab or the **Data Sources** tab. See [“Data Selection Windows and SAS Data Explorer” on page 2](#).
- 2 Click the **Available** tab or the **Data Sources** tab.
- 3 Right-click a table or file that you want to copy and select **Add to import**. The selected table or file is added to the queue on the **Import** tab.
- 4 Click the **Import** tab. The table or file that you want to copy appears on the left. Import properties for the copy appear on the right.
- 5 The **Target table name** field is usually set to the source table name. Accept or change this name as appropriate for the copy.

If you change the table name, follow the conventions for table names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). The target operating system and data source might have additional constraints on these names. For example, the target table name can be a maximum of 247 characters long. However, the name might have to be shorter, due to the data source settings for the caslib that is specified in the **Target location** field. Contact your data administrator for more information about the permitted length for target table names.


Note: The target table name cannot be identical to the source table name unless you change the caslib or select the **Replace file** option.

- 6 If the source caslib is global in scope, the name of the source caslib was copied into the **Target location** field. To store the copy elsewhere, you can select a different global caslib or add a new one. To copy sensitive data to a different caslib, make sure that the target caslib and file have appropriate protections. See [“Copy Data from One Caslib to Another” on page 53](#).

If you have the privilege that is required to add a new global caslib, follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).

- 7 Review the setting for the **Save as in-memory table only** check box. Select this check box to load the imported table to memory without saving a physical copy of the table to the target

destination. Only users who have the Promote permission on the target caslib can successfully perform an import using this option.

- 8 Specify what action the import operation should take if the target table name exists in the caslib that is specified in the **Target location** field. The options are to cancel the import or to replace the existing table that has the same name.
- 9 Specify a label that will help identify the copy, if desired. The caslib for the target table must support labels. Caslibs of type PATH, DNFS, and HDFS support labels. A caslib for a DBMS does not support labels.
- 10 A **Format** selector for the target table is available when certain types of caslibs are selected in the **Target location** field. The selector enables you to copy the source table to a target table in a different format. Select a format that is appropriate for your goals. For more information, see [“Copying Tables in One Format to Another Format” on page 51](#).
- 11 If the item to be copied is a physical item (a table or file with  beside it), the **Filter Rows** tab and **Select Columns** tab are displayed on the right. You can use these tabs to subset columns or rows from the original. For more information about these tabs, see [“Filter Rows from an Imported Table or File” on page 53](#) and [“Select Columns from an Imported Table or File” on page 53](#).

Some files and tables have an **Advanced** tab with an **Additional Input Options** control. This control enables you to specify additional options that might be needed to accurately copy the contents of the imported file. To add options, select **Create additional input options**. Click ⓘ for a description of the valid options. For more information about import options, see [Common Parameter: importOptions](#). Options are entered as name/value pairs, such as **Name:** *useNarrowCharacterTypes*, **Value:** `TRUE`. Click **Add Option** to add the option to the import record.

- 12 When ready, you can run the import operation. Do one of the following:
 - You can right-click the table or file to be copied and select **Import Item**. If the import succeeds, a copy of the table or file is loaded to memory on the CAS server that is specified in the caslib. The copy of the table or file can be selected from the **Available** tab or the **Data Sources** tab. If the import fails, see [“General Usage Notes” on page 71](#).

The **Import Item** option is used to run a job that imports the selected data. You can review the status of import jobs on the **Monitoring** tab of the Jobs page in SAS Environment Manager. For more information about using this tab, see [“Jobs and Flows Page.”](#)
 - If you right-click the table or file in the import queue, and the **Create job** option is available, you can use this option to create a job request on the **Scheduling** tab of the Jobs window in SAS Environment Manager. You can then use the **Scheduling** tab to run or schedule the job at an appropriate time. For more information, see [“Create Import Job Requests” on page 54](#). After the import job request is run, a copy of the table or file is loaded to memory on the CAS server that is specified in the caslib. The copy can be selected from the **Available** tab or the **Data Sources** tab.

Copying Tables in One Format to Another Format

Overview of the Format Selector

Some SAS Viya applications enable you to copy data to a caslib. If the target caslib type is PATH, DNFS, HDFS, ADLS, or S3, a **Format** selector for the target table might be available, as shown in the next figure.

Figure 21 Format Selector for a Target Table

Target location: *

qst_13646_vb025/COMMON_PROD Find

☐ Save as an in-memory table only

If target table name exists:

☒ Cancel import

☐ Replace file

Label: Enter label

Format: sashdat

The **Format** selector enables you to copy a source table in a specific format to a target table in a different format. Only those formats that are supported by the current task and the target caslib are available for selection. Here is a list of the supported formats for tables in each type of caslib.

PATH

sashdat, sas7bdat, csv, txt, dta, sav, xls, xlsx, jmp, parquet, orc

DNFS

sashdat, csv, txt, parquet

HDFS

sashdat, csv, txt

ADLS

orc, csv

S3

sashdat, csv, txt, parquet

Other caslibs

The **Format** selector is hidden.

Usage Notes for the Format Selector

Some formats that are available for the target caslib might lack support for data types that are present in the source data. For example, importing a directory of image files and saving to a CSV file might be supported by the target caslib. However, the binary image data cannot be properly stored in a CSV file, and the resulting file is unusable. In such cases, choosing a format that has more extensive support for data types, such as sashdat, is preferable.

Verify that the target table name is appropriate for your goals:

- Do not add a file format extension, such as .csv, to the target table name. Adding an extension to the target table name can cause the Save operation to fail.
- Very long table names can cause an import to Parquet format to fail. If this happens, specify a shorter name for the target table.

Verify that the target caslib is appropriate for your goals:

- The target caslib must be a PATH, DNFS, HDFS, ADLS, or S3 caslib type.
- The target caslib should be appropriate for the work that you want to do with the reformatted table. For example, it is possible to save a Parquet file to a PATH caslib, but it might be more useful to save it to a DNFS caslib or an S3 caslib.

Copy Data from One Caslib to Another


You can use the **Add to import** option to copy data from one caslib to another. For example, you can use caslibs to copy data from one DBMS to another. To do this, follow the steps in [“Add to Import from the Data Sources Tab or Available Tab” on page 50](#) and specify a caslib in the **Target location** field that is different from the source caslib. This creates a physical copy of the data in the target caslib as well as an in-memory copy.

To copy sensitive data to a different caslib, make sure that the target caslib and file have appropriate protections. See [“Considerations for Cross-Caslib Data” in SAS Viya: CAS Authorization](#).

Filter Rows from an Imported Table or File

As described in [“Add to Import from the Data Sources Tab or Available Tab” on page 50](#), you can right-click a physical table or file and select **Add to import**. The table or file is added to the queue on the **Import** tab. The **Filter Rows** tab appears on the right side.

You can use the **Filter Rows** tab to enter the argument for a SAS WHERE expression that selects and copies rows from the selected table or file. For a syntax reference, see [Syntax of WHERE Expression](#). The SAS expression is translated into native database syntax and is pushed down to the database, as appropriate.

The data source must be a physical item (a table or file with  beside it). The table or file must support a WHERE clause. Supported databases include DB2, Hadoop Hive, Impala, SAS LASR Analytic Server, ODBC, Oracle, PostgreSQL, Redshift, and Teradata. Other supported data sources include SAS data sets, SASHDAT files, SAS LASR files, Microsoft Excel files, and delimited text files. Social media files and Esri data are not supported.

To select and copy rows from a physical table or file in the queue on the **Import** tab:


- 1 If you have not done so already, perform steps 1–9 as described in [“Add to Import from the Data Sources Tab or Available Tab” on page 50](#).
- 2 Select **Enable row filtering** on the **Filter Rows** tab.
- 3 Enter the argument for the SAS WHERE expression that selects and copies rows from the table or file to be copied. Omit the `WHERE=` part of the expression. Use double quotation marks to surround column names with special characters, as shown in the following example:

```
("Sal" between 1000 and 3000) and ("comm" is NULL)
```
- 4 When ready to test your code, you can right-click the table or file to be copied and select **Import Item**. If the import succeeds, a copy of the table or file is loaded to memory on the CAS server that is specified in the **Target destination** field. The copy of the table or file can be selected from the **Available** tab or the **Data Sources** tab.

If the import fails, see [“Usage Notes for Importing Local Files” on page 27](#).

Select Columns from an Imported Table or File

As described in [“Add to Import from the Data Sources Tab or Available Tab” on page 50](#), you can right-click a physical table or file and select **Add to import**. The table or file is added to the queue on the **Import** tab. The **Select Columns** tab appears on the right.

By default, all columns from the table or file are copied to the target table. You can use the **Select Columns** tab to select a subset of columns from the selected table or file. The data source must be a physical item (a table or file with  beside it).

To select a subset of columns from a physical table or file in the queue on the **Import** tab:

- 1 If you have not done so already, perform steps 1–9 as described in [“Add to Import from the Data Sources Tab or Available Tab” on page 50](#).
- 2 Select **Enable column selection** on the **Select Columns** tab. By default, all columns from the table or file are listed in the **Selected columns** panel.
- 3 To select a subset of columns, use the arrows to move columns from the **Selected columns** panel to the **Available columns** panel. You can use the up and down arrows to rearrange the columns in the **Selected columns** panel.
- 4 When you are ready to copy the table or file, select **Import Item**.

If the import is successful, a copy of the table or file is loaded to memory on the CAS server that is specified in the **Target location** field. A message that indicates a successful operation is displayed above the import properties on the right. The copy of the table or file can be selected from the **Available** tab or the **Data Sources** tab.

Create Import Job Requests


If you right-click a table or file in the queue on the **Import** tab, then select **Import item**, the import process begins immediately. This process can take a long time if the item is large. If the item is in a global caslib, you have an alternative: the **Create job** option. You can use this option to create a job request on the **Scheduling** tab of the Jobs window in SAS Environment Manager. You can then use the **Scheduling** tab to run or schedule the job at an appropriate time. A job request ensures that you can run large jobs at an appropriate time, to reduce the load on your system.

The **Create job** option is not valid for local files, social media feeds, or Esri data. It is valid only for tables or files that have been copied from the **Available** tab or the **Data Sources** tab, using the **Add to import** option. These tables or files must be associated with a global caslib. For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).

To select a subset of columns from a physical table or file in the queue on the **Import** tab:

- 1 If you have not done so already, perform steps 1–9 as described in [“Add to Import from the Data Sources Tab or Available Tab” on page 50](#).
- 2 Select the table or file in the queue on the **Import** tab to copy with the **Create job** option.
- 3 Verify that the import properties on the right are correct for the copy that you will create. For example, the name specified in the **Target table name** field cannot be the same as the original unless you have selected the **Replace file** option.
- 4 When ready, right-click the table or file and select the **Create job** option. The Create Job window is displayed.
- 5 Specify a name and description for the job and click **OK**. The name cannot be longer than 100 characters. A status message is displayed, stating whether the job request was created successfully.

The next step is to access the job request in SAS Environment Manager, where you can run the job immediately or schedule it.

- 6 From the application bar, click the applications menu button at the top left corner. Select **Manage Environment**. SAS Environment Manager is displayed.
- 7 Click  (**Jobs**) in the navigation bar on the left. Any job requests that you have created are displayed on the **Scheduling** tab of the Jobs and Flows table. The name of the job is based on the name of the table or file that you imported. For more information, see [“Schedule Jobs.”](#)

After the import job runs, a copy of the table or file is loaded to memory on the CAS server that is specified in the caslib. The copy can be selected from the **Available** tab or the **Data Sources** tab.

Import a Snapshot of an ESP Project

SAS Event Stream Processing enables you to capture and analyze streaming data from sensors, IoT devices, and other data sources. You can add a user-defined connection to a project on an ESP server. The "windows" created by that project are displayed as unloaded tables on the **Data Sources** tab. If you import one of these tables and load it into memory, you have a snapshot of the data that was streamed through that window. You can then run a data profile and perform other tasks with this snapshot of streamed data.

To import a snapshot of an ESP project:

- 1 Add a user-defined connection to a project on an ESP server. See [“Example: User-Defined Connection to an ESP Project” on page 16](#).

The windows that are created by that project are displayed as unloaded tables in the **Data Sources** tab.

- 2 Import one of these tables and load it to memory. See [“Add to Import from the Data Sources Tab or Available Tab” on page 50](#).

Note: By default, the name of the source (ESP caslib from Step 1) is specified as the target in the **Target location** field for the import. ESP caslibs are read-only in this context. You must select a writeable caslib in the **Target location** field.

- 3 If the import is successful, you can right-click the loaded table and run a data profile and perform other tasks with this snapshot of streamed data. See [“Working with Tables” on page 37](#).

Importing Delimited Files from a Caslib Directory

Overview of Importing Delimited Files from a Directory

You can extract the contents of all delimited files in a caslib directory and copy this information into an in-memory table. SAS Viya applications can then analyze the information in this table. The delimited files must be in comma-separated value (CSV) format, and they must all have the same column metadata.

One way to import all delimited files in a directory is to select **Directory** ⇒ **Delimited files** on the **Import** tab. You are prompted to select a caslib that includes a directory of delimited files that you want to analyze. After you select caslib and directory, a **Merge files** item is added to the import queue.

Figure 22 Merge Files Item in the Queue on the Import Tab

The screenshot displays the SAS Enterprise Guide interface. On the left, the 'Import (1)' tab is active, showing a queue with one item: 'Merge files in "Public" cas-shared-default/Public'. On the right, the configuration panel for this item is open. The 'Target table name' is 'Public_Merged'. The 'Target location' is 'cas-shared-default/Public'. There is a checkbox for 'Save as an in-memory table only' which is unchecked. Under 'If target table name exists:', the 'Cancel import' radio button is selected. The 'Label' field is 'Enter label' and the 'Format' is 'sashdat'. Below these fields are tabs for 'File Specifications', 'Filter Rows (Off)', 'Select Columns (Off)', and 'Advanced'. The 'File Specifications' tab is active, showing a warning: 'The directory files must have an identical structure to merge successfully.' There is an unchecked checkbox for 'Include delimited files in subdirectories' and a dropdown for 'Input file delimiter' set to 'Comma'.

Import properties for the **Merge files** item appear on the right. You specify various options and import the files that are specified in the item. The contents of the files are copied into the table in the **Target table name** field and loaded to memory. The target table is in the caslib that is specified in the **Target location** field.

Prerequisites for Importing Delimited Files from a Directory

Perform the following tasks before importing a directory of delimited files:

- 1 Collect a set of delimited files that you want to analyze. The files must be in comma-separated value (CSV) format, and they must all have the same column metadata.
- 2 Copy these files to a directory that a CAS server can access.
- 3 Add the source caslib.

The source caslib is a caslib of type PATH, DNFS, or S3.

The root directory of the caslib can include delimited files and subdirectories of delimited file that you want to analyze.

Select the **Include subdirectories** check box when you add the caslib. This setting is required in order to support the Choose Directory window on the **Import** tab.

For more information about PATH and DNFS caslibs, see [“Connecting to Remote File Systems” on page 9](#). For more information about S3 caslibs, see [“Example: User-Defined Connection to S3” on page 17](#).

4 Add or identify the target caslib.

The target caslib is a global caslib that is on the same CAS server as the source caslib.

Import Delimited Files from a Directory

After you have prepared your source and target caslibs as previously described, you can extract the contents of all delimited files in a caslib directory.

1 To import delimited files from the root directory of a caslib or from all directories in the caslib:

- a Click the **Data Sources** tab.
- b Right-click a path-based caslib that contains the delimited files that you want to analyze. Select **Import from library** ⇒ **Delimited files**.
A **Merge files** item is added to the import queue, and import properties appear on the right. Go to Step 2.

Alternatively, to import delimited files from a specific directory in a caslib:

- a Click the **Import** tab.
- b Select **Directory** ⇒ **Delimited files**.
The Choose Directory window is displayed.
- c Select a caslib that contains a directory of delimited files that you want to analyze.
- d Select the directory of delimited files and click **Select**.

A **Merge files** item is added to the import queue, and import properties appear on the right.

2 The **Target table name** is the name of the caslib or the selected directory with the suffix **_Merged**. This name specifies the output table that contains the extracted information. Accept or change this name as appropriate.

If you change the name, follow the conventions for table names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). The target operating system and data source might have additional constraints on these names. For example, the target table name can be a maximum of 247 characters long. However, a shorter name might be required due to the data source settings for the caslib that is specified in the **Target location** field. Contact your data administrator for more information about the permitted length for target table names.

3 In the **Target location** field, select an existing global caslib or add a new one.

The caslib must be a global caslib on the same CAS server as the source caslib. If you have the privilege that is required to add a new global caslib, follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).

4 Specify what action the import operation should take if the target table name exists in the caslib that is specified in the **Target location** field. The options are to cancel the import or to replace the existing table that has the same name.

- 5 Specify a label that helps identify the output table, if desired. The caslib for the target table must support labels. Caslibs of type PATH, DNFS, and HDFS support labels. A caslib for a DBMS does not support labels.
- 6 A **Format** selector for the target table is available when certain types of caslibs are selected in the **Target location** field. The selector enables you to copy the source table to a target table in a different format. Select a format that is appropriate for your goals. For more information, see [“Copying Tables in One Format to Another Format” on page 51](#).
- 7 Use the options on the **File Specifications** tab as needed. Click ⓘ for information about these options.
- 8 Use the **Filter Rows** tab and **Select Columns** tab subset columns or rows from the original file. For more information about these tabs, see [“Filter Rows from an Imported Table or File” on page 53](#) and [“Select Columns from an Imported Table or File” on page 53](#).
- 9 Use the options on the **Advanced** tab as needed. Click ⓘ for information about these options. To add options, select **Create additional input options**. Options are entered as name/value pairs. Click **Add Option** to add the option to the target caslib.
- 10 When ready, you can right-click the **Merge files** item in the queue and select **Import Item**.
 If the import succeeds, contents from files in the source caslib are copied into a table in the target caslib. The output table is loaded to memory on the CAS server that is specified in the **Target location** field.
 To verify the output, select the output table on the **Available** tab or the **Data Sources** tab. Click the **Sample Data** tab to see a sample of the output.
 If the import fails, see [“If Data Access Fails” on page 73](#).

Importing Documents from a Caslib Directory

Overview of Importing Documents

You can extract text and metadata from all documents in a caslib directory and load this information to an in-memory table. Each row in the table represents one document and its metadata. You can use applications such as SAS Visual Text Analytics to analyze the information in this table.

One way to import documents is to select **Directory** ⇨ **Documents** on the **Import** tab. You are prompted to select a caslib that includes a directory of documents that you want to analyze. After you select a caslib and directory, a **Documents** item is added to the import queue.

Figure 23 Document Item in the Queue on the Import Tab

The screenshot displays the 'Import (1)' tab in a software application. On the left, a list of items is shown under 'Import (1)'. The selected item is 'Documents in "doc"', with a sub-path 'qst_13646_vb025/DOC_CONVERSI...'. On the right, the configuration panel for this item is visible. It includes fields for 'Target table name' (set to 'Table A'), 'Target location' (set to 'qst_13646_vb025/DOC_CONVERSI...'), and a 'Find' button. There is an unchecked checkbox for 'Save as an in-memory table only'. Under 'If target table name exists:', the 'Cancel import' radio button is selected. The 'Label' field is 'Enter label', and the 'Format' dropdown is set to 'sashdat'. At the bottom, there are tabs for 'File Specifications' and 'Advanced'. Under 'File Specifications', the 'Filter file extensions' checkbox is checked, with a text input showing 'txt, pdf, etc'. The 'Recurse through directories' checkbox is unchecked.

Import properties for the **Documents** item appear on the right. You specify various options and import the documents that are specified in the item. The software uses Apache Tika libraries to extract metadata and text from the documents. Text and metadata from the documents is copied into the table in the **Target table name** field and loaded to memory. The target table is in the caslib that is specified in the **Target location** field.

Prerequisites for Importing Documents

Perform the following tasks before importing a directory of documents:

- 1 Collect a set of documents that you want to analyze. You can extract information from documents that are in the following formats.

Table 3 Document Formats That Can Be Converted

Format	File Extensions
HTML	.html
XML	.xml
MS Office	.doc, .docx, .xls, .xlsx, .ppt, .pptx
Open Document Format	.odt, .ods, .odp
iWork document format	.pages, .numbers, .key

Format	File Extensions
WordPerfect	.wpd
PDF	.pdf
Electronic publication format	.epub
Rich Text Format	.rtf
Text formats	.txt, .csv
Mail formats	.mbox, .msg, .tnef, .pst

2 Copy these documents to a directory that a CAS server can access.

3 Add the source caslib.

This caslib is a caslib of type PATH whose root directory can include documents and subdirectories of documents that you want to analyze.

Select the **Include subdirectories** check box when you add the caslib. This setting is required in order to support the Choose Directory window on the **Import** tab. For more information about adding PATH caslibs, see [“Connecting to Remote File Systems” on page 9](#).

4 Add or identify the target caslib.

The target caslib is a global caslib on the same CAS server as the source caslib. It contains the text and metadata that is extracted from the source documents.

Import Documents from a Directory

After you have prepared your source and target caslibs as previously described, to extract text and metadata from a directory of documents:

1 To import documents from the root directory of a caslib or from all directories in the caslib:

- a Click the **Data Sources** tab.
- b Right-click a path-based caslib that contains the documents that you want to analyze. Select **Import from library** ⇒ **Documents**.
A **Documents** item is added to the import queue, and import properties appear on the right. Go to Step 2.

Alternatively, to import documents from a specific directory in a caslib:

- a Click the **Import** tab.
- b Select **Directory** ⇒ **Documents**.
The Choose Directory window is displayed.
- c Select a caslib that contains a directory of documents that you want to analyze.
- d Select the directory of documents and click **Select**.

A **Documents** item is added to the import queue, and import properties appear on the right.

- 2 The default **Target table name** is **Table A**. This is the name of the output table that contains the extracted text and metadata. Accept or change this name as appropriate.

If you change the name, follow the conventions for table names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). The target operating system and data source might have additional constraints on these names. For example, the target table name can be a maximum of 247 characters long. However, a shorter name might be required due to the data source settings for the caslib that is specified in the **Target location** field. Contact your data administrator for more information about the permitted length for target table names.

- 3 In the **Target location** field, select an existing global caslib or add a new one.

The caslib must be a global caslib on the same CAS server as the source caslib. If you have the privilege that is required to add a new global caslib, follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).

- 4 Specify what action the import operation should take if the target table name exists in the caslib that is specified in the **Target location** field. The options are to cancel the import or to replace the existing table that has the same name.
- 5 Specify a label that helps identify the output table, if desired. The caslib for the target table must support labels. Caslibs of type PATH, DNFS, and HDFS support labels. A caslib for a DBMS does not support labels.
- 6 A **Format** selector for the target table is available when certain types of caslibs are selected in the **Target location** field. The selector enables you to copy the source table to a target table in a different format. Select a format that is appropriate for your goals. For more information, see [“Copying Tables in One Format to Another Format” on page 51](#).
- 7 Use the options on the **File Specifications** tab as needed. Click ⓘ for information about these options.
- 8 Use the options on the **Advanced** tab as needed. Click ⓘ for information about these options. To add options, select **Create additional input options**. Options are entered as name/value pairs. Click **Add Option** to add the option to the target caslib.
- 9 When ready, you can right-click the **Documents** item in the queue and select **Import Item**.

If the import succeeds, text and metadata from documents in the source caslib are copied to a table in the target caslib. The output table is loaded to memory on the CAS server that is specified in the **Target location** field.

To verify the output, select the output table on the **Available** tab or the **Data Sources** tab. Click the **Sample Data** tab to see a sample of the output.

If the import fails, see [“If Data Access Fails” on page 73](#).

Importing Images from a Caslib Directory

Overview of Importing Images

If you license SAS Visual Data Mining and Machine Learning, an image import option becomes available on the **Data Sources** tab and the **Import** tab. This option extracts images and associated metadata from all images in a caslib directory. It loads this information to an in-memory table. Each row in the table represents one image and its metadata. You can use SAS Visual Data Mining and Machine Learning to analyze the information in this table.

One way to import images is to select **Directory** ⇒ **Image files** on the **Import** tab. You are prompted to select a caslib that includes a directory of images that you want to analyze. After you select a caslib and directory, an **Images** item is added to the import queue.

Figure 24 Images Item in the Queue on the Import Tab

The screenshot displays the SAS Visual Data Mining and Machine Learning interface. The top navigation bar includes 'Available', 'Data Sources', and 'Import (1)'. The left pane, titled 'Import (1) ?' with a plus icon, shows a list of items to import. One item is visible: 'Images in "images_dicom"' with a sub-path 'cas-shared-default/IMAGES_PROD/i...'. The right pane is titled 'Images in "images_dicom"' and contains configuration options. At the top right of the right pane are 'Import Item' and 'Import All' buttons. The configuration fields include:

- Target table name:** A text field containing 'images_dicom_Images'.
- Target location:** A text field containing 'cas-shared-default/Public' with a folder icon and a 'Find' button.
- Save as an in-memory table only:** An unchecked checkbox.
- If target table name exists:** Two radio buttons: 'Cancel import' (selected) and 'Replace file'.
- Label:** A text field with the placeholder 'Enter label'.
- Format:** A dropdown menu set to 'sashdat'.

 Below these fields are two tabs: 'Load Image Settings' and 'Advanced'. The 'Advanced' tab is active, showing:

- Include images contained in subdirectories:** A checked checkbox.
- Decompress images:** An unchecked checkbox.
- Add additional metadata:** An unchecked checkbox.
- Import images as a series:** An unchecked checkbox.

Import properties for the **Images** item appear on the right. You specify various options and import the images that are specified in the item. Images and metadata are copied into the table in the **Target table name** field and loaded to memory. The target table is in the caslib that is specified in the **Target location** field.

Prerequisites for Importing Images

Perform the following tasks before importing a directory of images:

- 1 Collect a set of images that you want to analyze. You can extract information from images that are in the following formats.

Table 4 Image Formats That Can Be Imported

Format	File Extensions
Windows bitmap	.bmp, .dib
Joint Photographic Experts Group (JPEG)	.jpeg, .jpg, .jpe
Portable Network Graphics	.png
Tagged Image File Format (TIFF)	.tiff, .tif
ZIP files of photos	.zip
Digital Imaging and Communications in Medicine (DICOM) standard	.dcm, none
Nearly raw raster data (NRRD)	.nrrd, .nhdr
Neuro-imaging Informatics Technology Initiative (NIfTI)	.nii
Analyze	.img
MetaImage	.mha, .mhd
WebP	.webp
Portable image format	.pbm, .pgm, .ppm
JPEG 2	.jp2

- 2 Copy these images to a directory that a CAS server can access.
- 3 Add the source caslib.

This caslib is a caslib of type PATH whose root directory can include images and subdirectories of images that you want to analyze.

Select the **Include subdirectories** check box when you add the caslib. This setting is required in order to support the Choose Directory window on the **Import** tab. For more information about adding PATH caslibs, see [“Connecting to Remote File Systems” on page 9](#).

- 4 Add or identify the target caslib.

The target caslib is a global caslib on the same CAS server as the source caslib. Each row in the output table represents one image and its metadata.

Import Images from a Directory

After you have prepared your source and target caslibs as previously described above, perform the following steps to extract images and metadata from a directory of images:

- 1 To import images from the root directory of a caslib or from all directories in the caslib:
 - a Click the **Data Sources** tab.
 - b Right-click a path-based caslib that contains the images that you want to analyze. Select **Import from library** ⇒ **Image files**.
An **Images** item is added to the import queue, and import properties appear on the right. Go to Step 2.

Alternatively, to import images from a specific directory in a caslib:

- a Click the **Import** tab.
 - b Select **Directory** ⇒ **Images files**.
The Choose Directory window is displayed.
 - c Select a caslib that contains a directory of images that you want to analyze.
 - d Select the directory of images and click **Select**.
An **Images** item is added to the import queue, and import properties appear on the right.
- 2 The default **Target table name** is the name of the selected directory. This name specifies the output table that contains the extracted information. Accept or change this name as appropriate.
If you change the name, follow the conventions for table names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). The target operating system and data source might have additional constraints on these names. For example, the target table name can be a maximum of 247 characters long. However, a shorter name might be required due to the data source settings for the caslib that is specified in the **Target location** field. Contact your data administrator for more information about the permitted length for target table names.
 - 3 In the **Target location** field, select an existing global caslib or add a new one.
The caslib must be a global caslib on the same CAS server as the source caslib. If you have the privilege that is required to add a new global caslib, follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).
 - 4 Specify what action the import operation should take if the target table name exists in the caslib that is specified in the **Target location** field. The options are to cancel the import or to replace the existing table that has the same name.
 - 5 Specify a label that helps identify the output table, if desired. The caslib for the target table must support labels. Caslibs of type PATH, DNFS, and HDFS support labels. A caslib for a DBMS does not support labels.
 - 6 A **Format** selector for the target table is available when certain types of caslibs are selected in the **Target location** field. The selector enables you to copy the source table to a target table in a

different format. Select a format that is appropriate for your goals. For more information, see [“Copying Tables in One Format to Another Format” on page 51](#).

- 7 Use the options on the **Load Image Settings** tab as needed.

Include images contained in subdirectories. This option is selected by default.

Decompress images. Step 9 describes how to view images after you import them. Some image formats cannot be displayed by some browsers unless they are decompressed. Select this check box to decompress images in the output table for the import.

Add additional metadata. If you select the **Decompress images** option, you can select this option to add additional metadata to the images in the output table for the import. Select the additional metadata items that you want to add.

Import images as a series.

- To combine multiple images that are not in DICOM format into a single 3-D image, select **Import images as a series**. Make sure that **Import as a DICOM series** is deselected. The import operation reads all images from the specified directories and attempts to layer them into one 3-D image. The images must be photographic images with the same width and height. The images are layered in the alphabetical order of their file names. Only those images that constitute a 3-D image are fully supported by this option.
 - To combine multiple images in DICOM format into a series, select both **Import images as a series** and **Import as a DICOM series**. The import operation reads all images from the specified directories and attempts to assemble them into a DICOM series. Each image in the series is assigned a unique path that is based on the DICOM Unique Identifier (UID) of the series.
- 8 Use the options on the **Advanced** tab as needed. Click ⓘ for information about these options. To add options, select **Create additional input options**. Options are entered as name/value pairs. Click **Add Option** to add the option to the target caslib.
 - 9 When ready, you can right-click the **Images** item in the queue and select **Import Item**.
If the import succeeds, text and metadata from images in the source caslib are written to a table in the target caslib. The output table is loaded to memory on the CAS server that is specified in the **Target location** field.
To verify the output, select the output table on the **Available** tab or the **Data Sources** tab. Click the **Sample Data** tab to see a sample of the output. Click an image to preview it. Some image formats cannot be displayed by some browsers unless they are decompressed. Other formats might not be displayable unless the browser is configured to launch an external viewer.
If the import fails, see [“If Data Access Fails” on page 73](#).

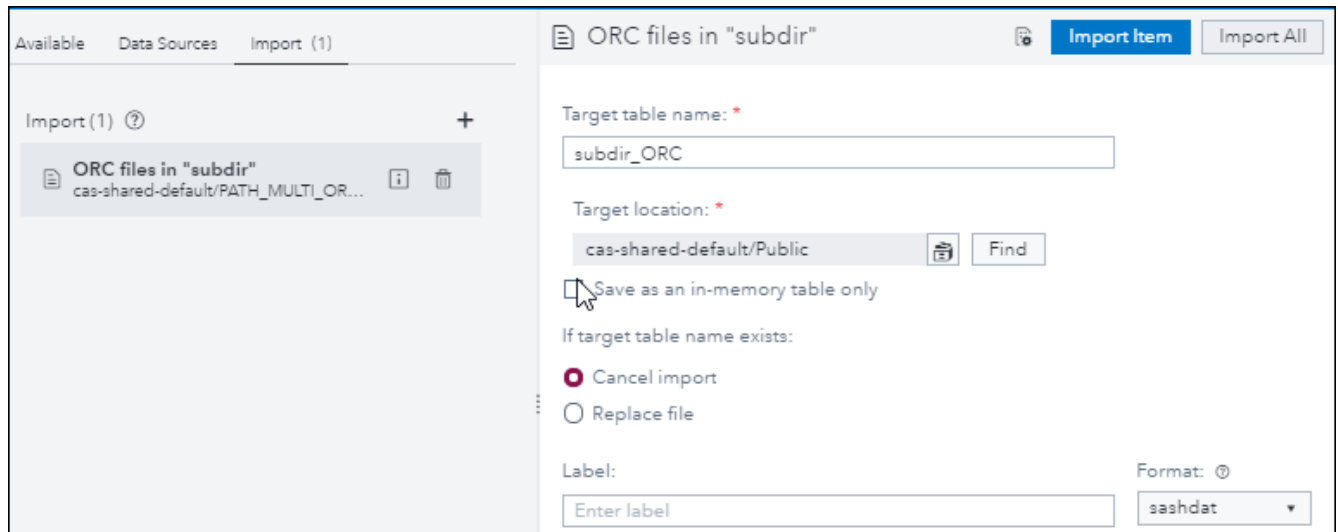
Importing ORC Files from a Caslib Directory

Overview of Importing ORC Files

ORC (Optimized Row Columnar) format is an open-source, data storage format for the Apache Hadoop ecosystem. You can extract the contents of all ORC files in a caslib directory and copy this information into an in-memory table. SAS Viya applications can then analyze the information in this table. The files must have the same column metadata.

You can use the **Directory** ⇒ **ORC files** option on the **Import** tab to select a caslib that contains a directory of ORC files. After you select a directory, an **ORC files** item is added to the import on the **Import** tab.

Figure 25 ORC Files Item in the Queue on the Import Tab



Import properties for the **ORC files** item appear on the right. You specify various options and import the files that are specified in the item. Output from the files is copied into the table in the **Target table name** field and loaded to memory. The target table is in the caslib that is specified the **Target location** field.

Prerequisites for Importing ORC Files

Perform the following tasks before using the **Directory** ⇒ **ORC files** option:

- 1 Collect a set of ORC files whose content is extracted for analysis. The files must have the same column metadata.
- 2 Copy these files to a directory that a CAS server can access.

Note:

The ORC files must be put into a subdirectory of the root of the source caslib. You cannot import ORC files from the root directory of a caslib.

- 3 Add the source caslib.

The source caslib is a caslib of type PATH or ADLS whose root directory includes one or more subdirectories of ORC files.

Select the **Include subdirectories** check box when you add the caslib. This setting is required in order to support the Choose Directory window on the **Import** tab. For more information about adding path-based caslibs, see [“Connecting to Remote File Systems” on page 9](#).

- 4 Add or identify the target caslib.

The target caslib is a global caslib that is on the same CAS server as the source caslib. It contains the information that is extracted from the ORC files.

Import ORC Files from a Directory

After you have prepared your source and target caslibs as previously described, to extract information from a directory of ORC files:

- 1 On the **Import** tab, select **Directory** ⇒ **ORC files**. The Choose Directory window is displayed.
- 2 Select a caslib that contains a directory of ORC files that you want to analyze.
- 3 Select the directory of ORC files and click **Select**.

An **ORC files** item is added to the Import queue. Import properties appear on the right.

- 4 The default **Target table name** is the name of the selected directory. This name specifies the output table that contains the extracted information. Accept or change this name as appropriate.

If you change the name, follow the conventions for table names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). The target operating system and data source might have additional constraints on these names. For example, the target table name can be a maximum of 247 characters long. However, a shorter name might be required due to the data source settings for the caslib that is specified in the **Target location** field. Contact your data administrator for more information about the permitted length for target table names.

- 5 In the **Target location** field, select an existing global caslib or add a new one.

The caslib must be a global caslib on the same CAS server as the source caslib. If you have the privilege that is required to add a new global caslib, follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 72](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 71](#).

- 6 Specify what action the import operation should take if the target table name exists in the caslib that is specified in the **Target location** field. The options are to cancel the import or to replace the existing table that has the same name.
- 7 Specify a label that helps identify the output table, if desired. The caslib for the target table must support labels. Caslibs of type PATH, DNFS, and HDFS support labels. A caslib for a DBMS does not support labels.
- 8 A **Format** selector for the target table is available when certain types of caslibs are selected in the **Target location** field. The selector enables you to copy the source table to a target table in a different format. Select a format that is appropriate for your goals. For more information, see [“Copying Tables in One Format to Another Format” on page 51](#).
- 9 Use the **Filter Rows** tab and **Select Columns** tab to subset columns or rows from the original files. For more information about these tabs, see [“Filter Rows from an Imported Table or File” on page 53](#) and [“Select Columns from an Imported Table or File” on page 53](#).
- 10 Use the **Advanced** tab with the **Additional Input Options** control as needed. A **fileType** option is provided by default. No additional options are available for ORC files.

- 11 When ready, you can right-click the **ORC files** item in the queue and select **Import Item**.

If the import succeeds, information from the ORC files in the selected directory is written to a table in the target caslib. The output table is loaded to memory on the CAS server that is specified in the **Target location** field.

To verify the output, select the output table on the **Available** tab or the **Data Sources** tab. Click the **Sample Data** tab to see a sample of the output.

If the import fails, see [“If Data Access Fails” on page 73](#).

Maintaining SAS Data Explorer

Administrative Tasks for SAS Data Explorer and the Data Selection Windows

Overview of the Tasks

The tasks described in this section must be performed by an administrator. They apply to the SAS Data Explorer application and to the embedded data selection windows unless stated otherwise.

Controlling Access to Features

Limit Access to the Download Table Option

By default, all authenticated users can access the **Download table** option. This option enables you to right-click a table on the **Available** or **Data Sources** tab and save it as a delimited file on your computer. Administrators can limit access to this option.

To limit access to the **Download table** option, an administrator finds the `dataPreparation_capabilities/exportTable` rule in SAS Environment Manager. The administrator changes the initial principal (Authenticated Users) to the ID of the group that should be able to perform the controlled function. For information about updating rules, administrators can refer to the topic for the “[Rules Page](#)” in *SAS Environment Manager: User’s Guide*.

Limit Access to the Import Tab

By default, all authenticated users can access the **Import** tab. Administrators can limit access to the **Import** tab or to individual social media feeds on this tab. For example, the **Import** tab requires users to add persistent data tables on a CAS server. You might want to limit who can do this.

To limit access to the **Import** tab, an administrator finds the relevant rule in SAS Environment Manager. The administrator changes the initial principal (Authenticated Users) to the ID of the group that should be able to perform the controlled function. The rule is as follows:

Import tab

`/casManagement_capabilities/importData`

For information about updating rules, administrators can refer to the topic for the “[Rules Page](#)” in *SAS Environment Manager: User’s Guide*.

CAS Management Service Options


Administrators can change how some features work in SAS Data Explorer and the embedded data selection windows by updating service options. Service options are updated from the Configuration page in SAS Environment Manager. For more information, see [“Configuration Page.”](#)

On the Configuration page, select All Services in the View selector. Enter *Explorer* in the **Filter** window. Select the SAS Data Explorer service and find the option that you want to update. For a description of the service options for SAS Data Explorer, see [“SAS Data Explorer” in SAS Viya: Configuration Reference.](#)

SAS Data Explorer Options

Options That Require a SAS Data Preparation License

SAS Environment Manager and other SAS Viya applications always include the embedded data selection windows that are described in [“Data Selection Windows and SAS Data Explorer” on page 2.](#) A SAS Data Preparation license is not required for these windows.


You do need a SAS Data Preparation license to display the SAS Data Explorer web application. To verify that SAS Data Preparation is licensed, click  at the upper left of SAS Viya applications and then select **Manage Data**. SAS Data Explorer is displayed. If you can display SAS Data Explorer, your site has licensed SAS Data Preparation. In that case, the tabs in both SAS Data Explorer and the embedded data selection windows have the same advanced options. For more information about advanced data profiling options, see [“Profiling Data” on page 42.](#)

If you have licensed SAS Data Preparation, you can use the `dataexplorer` command-line interface (CLI). For more information, see [“Using the SAS Data Explorer CLI” on page 75.](#)


For a description of all features that are available with a SAS Data Preparation license, see [“Additional Data Management Features Included with SAS Data Preparation” in Getting Started with SAS Data Preparation for SAS Viya,](#)

Modify SAS Data Explorer Settings

To modify settings for the SAS Data Explorer web application:

- 1 In the application bar, click your name, and then click **Settings**. The Settings window is displayed.
- 2 The following settings are available:
 - Global** settings. These settings are applied to all SAS web applications. Click  for details about each setting.
 - Data Explorer** settings. These settings are saved on a per-user basis. All of your settings persist between sessions.
 - General** settings.
 - There are two user preferences in this section.

Default target location

Click  to select a caslib where target tables are stored by default. This caslib serves as the default for various operations, including operations on the **Import** tab. A change to this setting will be applied to the next table that uses the default target location. A change is not applied retroactively.

If a default target location is not set in this field, the default CAS server and caslib for the CAS Management Service is used. For more information, administrators can refer to [“CAS Management Service” in SAS Viya: Configuration Reference](#).

Default locale for Quality Knowledge Base

Use this selector to specify the SAS Quality Knowledge Base (QKB) locale that is used by the **Analyze column contents while running profile** option. For more information about QKBs, see [“Quality Knowledge Base” in SAS Data Quality: Getting Started](#).

The locale should be appropriate for the data that you are profiling. For example, a table of names and addresses from the United States might be profiled with the **English-United States** locale. Locales from all QKBs on all CAS servers that appear on the **Data Sources** tab are listed here. Duplicate locales are filtered out.

If **Use the default server** is selected, the software uses the default locale that is specified for the CAS server for the profiled table.

If only one QKB is available, that will be considered the default QKB, regardless of the default setting.

If no default locale has been defined for this server, the profile fails.

Geographic Mapping settings.

Use the options in this section to accept the terms and conditions for Esri ArcGIS Online Services. For more information about these options, see [“Importing Esri Data for Geoenrichment and Geocoding” on page 28](#).

Profile settings.

There are two user preferences in this section.

Apply formats to variables when profiling data

Select this check box to apply formats to the output data for the **Run profile** option. Some data is more meaningful when it is formatted. For example, currency values might be more meaningful if they are formatted as currency rather than as integers. If formatting is not enabled, profiling could fail for some data, such as hexadecimal data. The impact on the performance of data profiling is usually acceptable. For more information about profiling, see [“Profiling Data” on page 42](#).

Analyze column contents while running profile

Select this check box to trigger column content analysis during profiling. If the analysis can determine what type of content is in the column, the column is tagged with the appropriate content tag. For example, a column that contains street addresses might get a **Street Address** tag. Other applications can use these tags to analyze the data in tables. The analysis is based on the locale that is specified in the **Default locale for Quality Knowledge Base** selector. Content analysis impacts profiling performance. For more information about column content analysis, see [“Enable Automatic Content Tagging for Columns in a Table” on page 43](#).

General Usage Notes

Caslibs on the Data Sources Tab and Import Tab

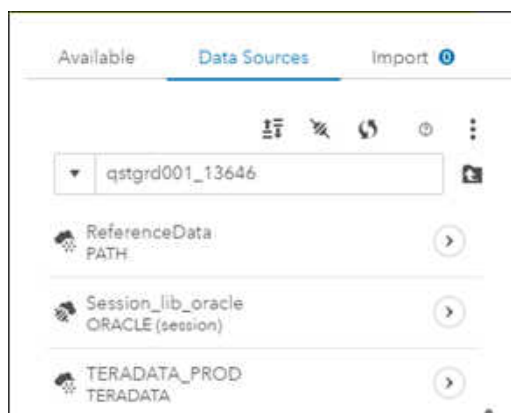
Overview



A caslib is an in-memory space to hold tables, access control lists, and data source information. One property of a caslib is scope. A caslib can have one of two scopes: session scope or global scope.

Session-based caslibs enable you to work with data during the current session only. They are dropped when you sign out of the current session. Session-based caslibs cannot share in-memory data with other CAS users or sessions, such as a session started by another application. Global caslibs persist as long as the CAS server is running. Most global caslibs enable you to share in-memory data with other CAS users, sessions, or applications.

The **Data Sources** tab displays the caslibs that have been defined on a CAS server, as shown in the next figure.

Figure 26 Caslibs on a Selected Server



Items with  beside them are global caslibs on the selected CAS server. Items with  beside them (and with the **session** label) are session-based caslibs on the selected CAS server.


Some features on the **Data Sources** tab or the **Import** tab are valid only for tables or files in global caslibs. For example, the following features are available only for global caslibs:

- **Load.** This option is used to load a table or file to memory on a CAS server.
- **Run profile.** A data profile report enables you to recognize data patterns, identify scarcity in the data, and review basic statistics for the selected table or file. For more information, see [“Profiling Data” on page 42](#).
- **Create job** on the **Import** tab. This option creates import jobs that can be scheduled in SAS Environment Manager. For more information, see [“Importing Data from the Available Tab or Data Sources Tab” on page 47](#).
- Only global caslibs can be selected in the **Target location** field when you are importing data from the **Import** tab.

Global caslibs with names such as CASUSER(*user-ID*) or CASUSERHDFS(*user-ID*) are personal caslibs. Personal caslibs are an optional feature that must be selected when the server is configured. When you start a session, personal caslibs are always available and have global scope. This enables you to access files and in-memory tables from any session that you start. However, they are personal and only your user ID can access the data. For more information, see [Personal, Pre-defined, and Manually Added Caslibs](#).

Caslibs and the Data Sources Tab

The **Data Sources** tab enables you to add a new caslib that connects to a data source. The privileges associated with your login determine what type of caslib you can add. Most application users have the privilege that is required to add new session-based caslibs. They do not have the privilege that is required to add new global caslibs.

If you click  on the **Data Sources** tab, the **Connection Settings** dialog box is displayed. This dialog box enables you to add a new caslib that connects to a data source.

The **Connection Settings** dialog box has an option **Persist this connection beyond the current session**. If you do not select this check box, and you specify a connection, the caslib that is added is a session-based caslib.

If you select the **Persist this connection beyond the current session** option, and you specify a connection, the caslib that is added is a global caslib. Your login must have the privilege that is required to add a global caslib, or the connection fails.

You can check to see what options are available for the tables or files in the new caslib. If the saved connection uses a global caslib, the following options are available when you right-click a table in the caslib: **Load**, **Delete**, **Add to import**, and **Run profile**. For information about these options, see [“Working with Tables” on page 37](#).

If the saved connection uses a session-based caslib, the following options are valid when you right-click a table in the caslib: **Delete** and **Add to import**. You do not have access to the **Load** option, which is used to load a table or file to memory on a CAS server. The **Run profile** option is available, but it is not valid for tables in a session-based caslib.

However, if you can access an appropriate global caslib, you can use the **Add to import** option to copy the table to that global caslib. From there, you have access to the **Load** option and the **Run profile** option. For information about the **Add to import** option, see [“Importing Data from the Available Tab or Data Sources Tab” on page 47](#).

Names for Caslibs, Tables, and Columns

When adding or renaming caslibs, tables, or columns, follow the SAS conventions for these names as described in [Rules for Caslib, Table, and Column Names](#). The target operating system and data source might have additional constraints on these names. For example, a colon cannot be used in a Windows file name.

Some special characters, such as backslash (\) and semicolon can be specified in a programming interface. However, many SAS Visual Analytics visual interfaces and REST clients do not accept special characters. To ensure that a name can be used by the greatest number of interfaces, avoid special characters.

Interfaces such as the **Data Sources** tab will not display all tables in a caslib where some table names differ only in case. Try to avoid such table names when adding tables to a caslib.

Specifying a Default Target Location

You can select a caslib where target tables is stored by default. This caslib serves as the default for various operations, including operations on the **Import** tab.

For the embedded Choose Data window or the SAS Data Explorer application: right-click a global caslib on the **Data Sources** tab and select **Set as default target location**.

For the SAS Data Explorer application only, see the **Default target location** setting in [“Modify SAS Data Explorer Settings” on page 69](#).

Enabling Other Users to Access a Global Caslib

When you add a global caslib, other people are not granted access to the new caslib by default. To enable users and groups to access a global caslib, right-click the caslib and select **Edit authorization**. Use the Authorization window to grant access to the caslib. For more information, see [“CAS Authorization Window” in SAS Viya: CAS Authorization Window](#).


If Data Access Fails

If you fail to access data from the **Data Sources** tab or the **Import** tab, an error message is displayed. Reasons for failure include:

- A hardware or software resource that the caslib requires is not available.
- You do not have permission to access a resource that the caslib requires.
- The caslib or the CAS server is incorrectly configured.

If the problem persists, contact an administrator.

Refresh to Update Information about Caslibs and Tables

Information about caslibs and tables on the **Available** and **Data Sources** tabs is stored in the cache for your web browser. If you think this information does not reflect the current state of your system, click  in the toolbar.

Support for Third-Party Software

Unless otherwise noted, SAS Data Explorer supports the databases, browsers, and other third-party software that are supported by SAS Viya. For more information, see [Third-Party Software Requirements for Use with SAS Viya](#).

Performance Options for the Available Tab

On large systems, the **Available** tab might be slow to refresh. An administrator can set the following SAS Data Explorer service options for the **Available** tab.

filterAvailableTab

Specifies whether the **Available** tab should display only loaded tables from the default caslib. You can use the **Data Sources** tab to work with any tables that are subsequently loaded from other caslibs.

filterAvailableTabByCaslibList

Specifies whether the caslib filter in the **filterAvailableTabCaslibList** tab should take effect. It includes users' personal caslibs in the filter.

filterAvailableTabCasliblist

Specifies a comma-delimited string of caslibs that should appear on the **Available** tab when the **filterAvailableTabByCaslibList** option is selected. It must be specified as a comma-delimited string with each caslib in the format `cas-server/caslib`. Use two consecutive commas to escape a comma in a server or library name.

useBatchQueryForAvailableTab

Enhances the performance of the **Available** tab. On by default. Available as of 2021.2.1.

To set these options, see [“CAS Management Service Options” on page 69](#).

No Tables Are Displayed on the Available Tab

If the **Available** tab does not display any tables, ask an administrator to check your SAS Viya environment. Determine whether the number of processor cores minus two is less than, or equal to, the number of CAS servers that are registered in the **dataSources** service. If so, the **Available** tab might not display any tables. The **Data Sources** tab might not display the list of initial CAS servers.

The preferred way to address this issue is to edit the SAS Data Explorer service option **jvm**. The administrator would add a configuration entry that looks like the following:

Name: `java_option_fjpcp`

Value: `-Djava.util.concurrent.ForkJoinPool.common.parallelism=16`

The `fjpcp` part of the name can be variable as long as it is unique. The value of `java.util.concurrent.ForkJoinPool.common.parallelism` should be set to at least the number of cas servers plus 2, but 16 could be considered a good minimum. To update the **jvm** option, see [“CAS Management Service Options” on page 69](#).

Using the SAS Data Explorer CLI

Overview

In SAS Viya, a command-line interface (CLI) is a user interface to the SAS Viya REST services. You enter commands on a command line and receive a response back from the system. You can use a CLI to interact directly with SAS Viya programmatically without a GUI. If you license SAS Data Preparation, you can use the `dataexplorer` CLI to automate data management.

To use any CLI, you must create a profile and sign in to SAS Viya. For instructions, see [“Command-Line Interface: Preliminary Instructions” in SAS Viya: Using the Command-Line Interface](#). You can then enter commands as shown in the examples below.

Note: In [“Command-Line Interface: Preliminary Instructions” in SAS Viya: Using the Command-Line Interface](#), commands begin with `sas-viya`. When using the SAS Data Explorer CLI, begin your commands with `sas-dataexplorer-cli`, as shown in the following examples.

For information about other CLIs that are provided with SAS Viya, see [“Command-Line Interface: Overview” in SAS Viya: Using the Command-Line Interface](#).

Examples

Note: Enter all commands on a single line. Multiple lines are used in some examples to improve readability.

Examples: List commands for the `dataexplorer` plug-in to the admin CLI.

```
sas-dataexplorer-cli --help
sas-dataexplorer-cli data-sources tables help create
```

Example: List the data sources.

```
sas-dataexplorer-cli data-sources list
```

Example: List the connections for the specified data source.

```
sas-dataexplorer-cli data-sources connections list --source dsName
```

Example: List the tables for the specified data source and connection.

```
sas-dataexplorer-cli data-sources tables list --source dsName
--connection connection
```

Example: List the rows for the specified table for the specified data source and connection.

```
sas-dataexplorer-cli data-sources rows list --source dsName
--connection connection --table tableName
```

Example: Delete a table from the specified data source and connection.

```
sas-dataexplorer-cli data-sources tables delete --source dsName
--connection connection --table tableName
```

Example: Create a new table from the specified data source and connection.

```
sas-dataexplorer-cli data-sources tables create --target-data-source dsName
--target-connection connection --target-table-name tableName --source-data-source
dsName
--source-connection connection
--source-table-path path-to-table
```

Example: Create a new table from a local file.

```
sas-dataexplorer-cli data-sources tables create --target-data-source dsName
--target-connection connection --target-table-name tableName
--local-file path-to-file
```

Example: Create a new table from a table URI.

```
sas-dataexplorer-cli data-sources tables create --target-data-source dsName
--target-connection connection
--target-table-name tableName
--source-table-uri uri-for-table
```

Example: Query for the specified job.

```
sas-dataexplorer-cli jobs query --job-id jobId
```

Example: Copy a table from a database using a generic data connector.

This command uses a generic data connector to copy a table from a Snowflake database.

```
sas-dataexplorer-cli --verbose dataexplorer data-sources tables create --source-data-
source cas-shared-default
--source-connection CLI_OTHER_SNOWFLAKE
--source-content-type other
--source-table-path CARS
--target-data-source cas-shared-default
--target-connection 'CASUSER(myID) '
--target-table-name cli-other-snowflake
```

CLI_OTHER_SNOWFLAKE is the name of the connection to a Snowflake database.

--source-content-type *other* specifies this connection as generic (user-defined).

CARS is the source table.

cas-shared-default is a CAS server.

Generic connectors enable the CLI to access any data for which there is a CAS connection type. You can specify a generic data connector for any of the data sources that are described in [“Quick Reference for Data Connector Syntax” in SAS Cloud Analytic Services: User’s Guide](#).

Example: Create a table from a subset of values from another table.

This command uses a generic data connector and import options to copy a subset of values from a SAS LASR Analytic Server table. The command then writes these values to a new table.

```
sas-dataexplorer-cli --verbose dataexplorer data-sources tables create
--source-data-source cas-shared-default
--source-connection CLI_LASR
--source-content-type other
--source-table-path CAKECLI
--import-options '{"fileType": "LASR", "where": "Age < 20"}'
```

```
--target-data-source cas-shared-default  
--target-connection 'CASUSER(myID) '  
--target-table-name cli-lasr-where-agelessthan20
```

The import options `'{"fileType": "LASR", "where": "Age < 20"}'` filter out rows when Age < 20 in the source table CAKECLI. For more information about the import options that are available for a data source, see [Common Parameter: importOptions](#).

