



SAS[®] Data Explorer 2.2: User's Guide

What's New in SAS Data Explorer 2.2

Cloud Data Exchange Is Available on the Data Sources Tab

If you have licensed SAS Data Preparation, connections to Cloud Data Exchange are available on the **Data Sources** tab. Cloud Data Exchange enables a CAS server to read and write data stored on-premises, behind a firewall. For more information about these connections, see ["Connecting to Databases" on page 12](#)

New Databases Are Available on the Data Sources Tab

User-defined data connections and connections to the RedShift database are now supported on the **Data Sources** tab. For more information, see ["Connecting to Databases" on page 12](#) and ["Adding User-Defined Connections" on page 17](#).

Advanced Input Options and Unique ID Columns

An **Advanced Settings** tab now displays the connection properties or import properties for some data sources. If the tab is present, it enables you to specify additional input options, a unique ID column, or both. Additional input options might be needed to accurately copy the contents of an imported file. The unique ID control enables you to add a column to the target table that contains a unique value for each row. These unique row identifiers can be used in text topics in SAS Visual Analytics.

The **Add to Import** option for CAS delimited files and Cloud Data Exchange tables enables you to specify additional input options. For more information, see ["Overview of the Add to Import Option" on page 50](#).

The **Local files** import option for Excel files, SAS data sets, or SASHDAT files enables you to specify a unique ID column. The **Local files** import option for delimited files enables you to specify additional input options, a unique ID column, or both. For more information, see ["Overview of Importing Local Data Files" on page 20](#).

New Import Feature Extracts Text and Metadata from a Collection of Documents

The **Documents Directory** option on the **Import** tab extracts text and metadata from a collection of documents on a CAS server. The text and metadata is written to a table on a CAS server. SAS Visual Text Analytics can then analyze this information. For more information, see [“Converting Documents for Analysis” on page 55](#).

New Import Feature Supports 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. For more information, see [“Geocoding” on page 38](#).

Improved Management of Tags for Tables and Columns

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. There are now more ways to add tags. You can manage tags from the **Details** tab for a table. For more information, see [“Managing Tags for Tables and Columns” on page 43](#).

New Data Profiling Features

If you license the SAS Data Preparation offering, the following new features are available:

- 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 will be added to the column.
- The **Run profile and save** option 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.

For more information, see [“Profiling Data” on page 46](#).

Improved Management of Tables and Caslibs

If you have appropriate privilege, you can view and edit the permissions that each user or group has for the selected caslib or table. You can change the physical path and description for PATH, HDFS, and DNFS caslibs. For more information, see [“Working with Caslibs and Tables” on page 41](#).

Improved Social Media Import Performance

Social media import features have been enhanced to support larger sets of data. Import jobs can now run longer before timing out. For more information, see [“Importing Data from Social Media” on page 25](#).

SAS Projects and Job Monitoring Have Moved

SAS projects are now maintained from the **Projects** tab in SAS Drive. For more information, see [“Projects” in SAS Drive: Getting Started](#). Job monitoring is now performed in the **Jobs** section of SAS Environment Manager. For more information, see [SAS Viya Administration: Jobs](#).

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. For more information about these features, see [“Profiling Data” on page 46](#).

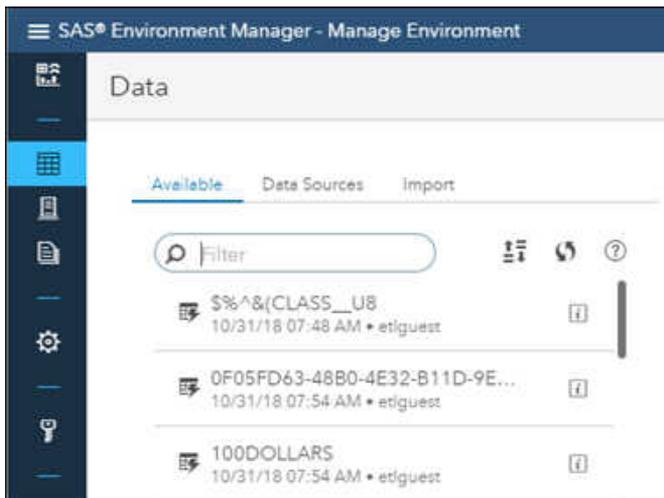
Working with SAS Data Explorer

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.

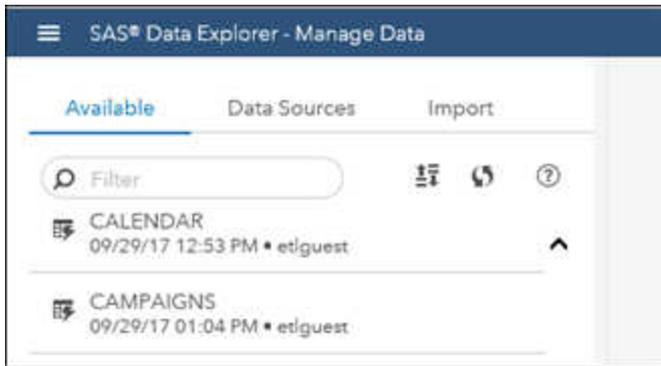
Figure 1 Data Selection Window in SAS Environment Manager



SAS Data Studio, SAS Visual Analytics, Model Studio, and SAS Decision Manager 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 4](#).

If you license SAS Data Preparation, the tabs in both SAS Data Explorer and the embedded data selection windows will have the same advanced options. For more information about advanced data profiling options, see [“Profiling Data” on page 46](#). For more information about Cloud Data Exchange connections, see [“Overview of Database Connections” on page 12](#).

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 7](#).

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 41](#).

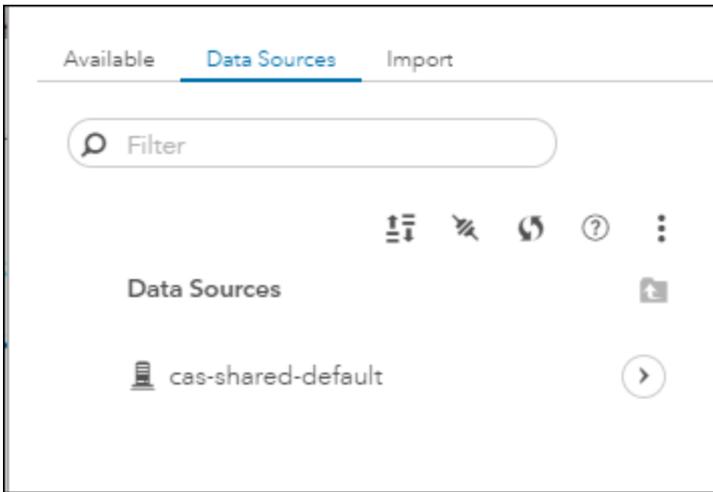
Data Sources Tab: Work with Databases or Remote File Systems

The **Data Sources** tab enables you to create a connection to a database server or a remote file system, such as a Hadoop Distributed File System (HDFS). If the connection is successful, tables that you are authorized to access will be available on the **Data Sources** tab. For more information about these tasks, see [“Connecting to Databases” on page 12](#) and [“Connecting to Remote File Systems” on page 15](#).

The **Data Sources** tab also enables you to work with caslibs and tables on a CAS server. For more information about these tasks, see [“Working with Caslibs and Tables” on page 41](#).

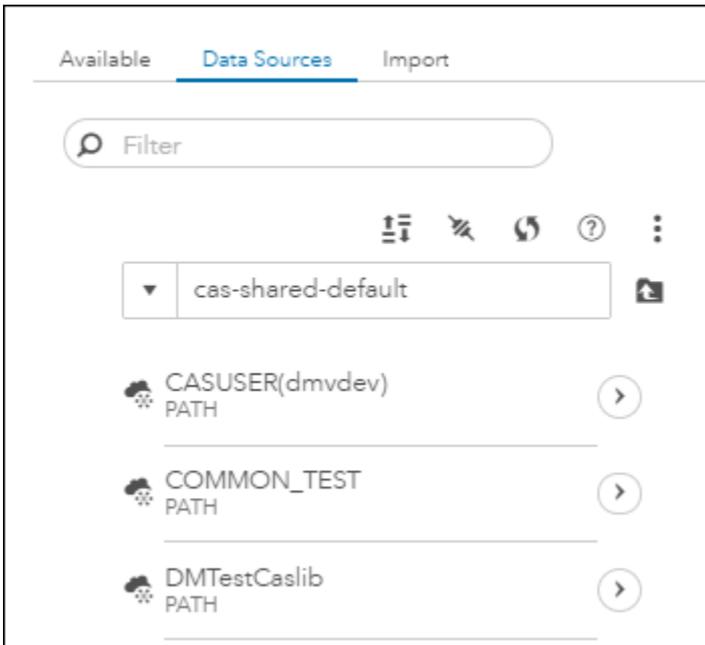
Here is an overview of how the **Data Sources** tab displays CAS servers, caslibs, and tables.

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

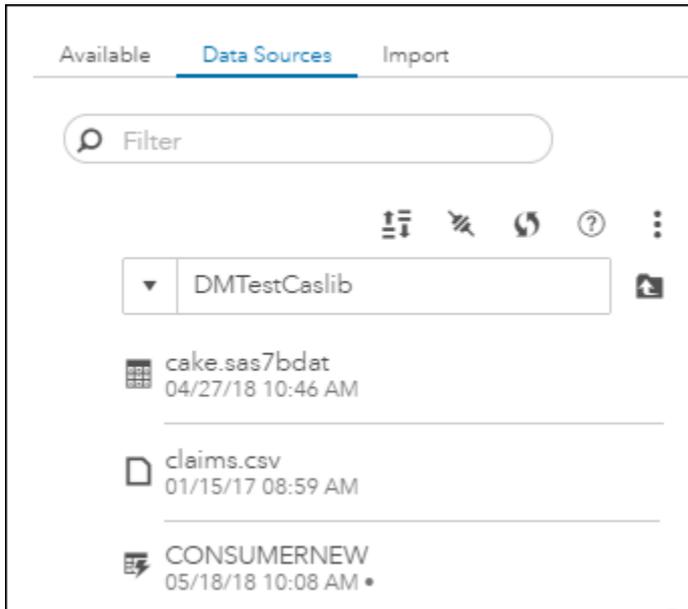


Items with  beside them are global caslibs on the selected CAS server. Global caslibs with names such as CASUSER(**user-ID**) or CASUSERHDFS(**user-ID**) are personal caslibs. Only your user ID can access the data in your personal caslib.

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 9](#). 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 41](#).

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 41](#).

Import Tab: Work with Local Files, Social Media Content, or Esri Data

The **Import** tab enables you to copy documents, local files, social media content, or Environmental Systems Research Institute (Esri) data to a caslib. It also enables you to copy a table or a file on the **Data Sources** tab or the **Available** tab to a caslib.

For more information, see the following topics:

- [“Converting Documents for Analysis” on page 55](#)
- [“Importing Local Files” on page 20](#)
- [“Importing Data from Social Media” on page 25](#)
- [“Importing Esri Data for Geo-enrichment and Geocoding” on page 34](#)
- [“Copying Data from the Available Tab or Data Sources Tab” on page 50](#)

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 7](#).

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 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. You might want to do this for two reasons:

- You do not want all users to be able to import data from Twitter, Facebook, Google Analytics, YouTube, or Google Drive.
- 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 or to individual social media feeds on this tab, an administrator will find the relevant rule in SAS Environment Manager. The administrator will then change the initial principal (Authenticated Users) to the ID of the group that should be able to perform the controlled function. The rules are as follows:

Import tab

/casManagement_capabilities/importData

Twitter Feed

/webDataAccess_capabilities/twitterImport

Facebook Feed

/webDataAccess_capabilities/facebookImport

Google Analytics Feed

/webDataAccess_capabilities/googleanalyticsImport

YouTube Feed

/webDataAccess_capabilities/youtubeImport

Google Drive Feed

/webDataAccess_capabilities/googledriveImport

For more information about updating rules, administrators can refer to [General Authorization: How to \(Rules Page\)](#) in *SAS Viya Administration: General Authorization*.

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 instructions on how to access the Configuration page, see [“Configuration Properties: How to Configure Services”](#) in *SAS Viya Administration: Configuration Properties*.

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 Administration: Configuration Properties*.

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 3](#). 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 will have the same advanced options. For more information about advanced data profiling options, see [“Profiling Data” on page 46](#). For more information about Cloud Data Exchange connections, see [“Overview of Database Connections” on page 12](#).

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 will be stored by default. This caslib will serve 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 Administration: Configuration Properties*](#).

Default locale for Quality Knowledge Base

Use this selector to specify the locale that is used by the **Analyze column contents while running profile** option. 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 no default locale has been defined for this server, the profile will fail. You can use this control to select a locale.

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.

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 Geo-enrichment and Geocoding” on page 34](#).

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. The impact on data profiling performance is usually acceptable. For more information about profiling, see [“Profiling Data” on page 46](#).

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 will be 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 46](#).

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. 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 or sessions.

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

Figure 6 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 46](#).
- **Create job** on the **Import** tab. This option creates import jobs that can be scheduled in SAS Environment Manager. For more information, see [“Copying Data from the Available Tab or Data Sources Tab” on page 50](#).
- Only global caslibs can be selected in the **Target destination** 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 will be a global caslib. Your login must have the privilege that is required to add a global caslib, or the connection will fail.

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 41](#).

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 [“Copying Data from the Available Tab or Data Sources Tab” on page 50](#).

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.

Specifying a Default Target Location

You can select a caslib where target tables will be stored by default. This caslib will serve 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 8](#).

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: How to \(Authorization Window\)” in SAS Viya Administration: Cloud Analytic Services Authorization](#).

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](#).

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.

One way to address this issue is to set the SAS Data Explorer service option **filterAvailableTab** to `true`. A value of `true` populates the **Available** and **Data Sources** tabs with tables only from the default caslib for your site. You can use the **Data Sources** tab to work with any tables that are subsequently loaded from other caslibs. To set the **filterAvailableTab** option, see [“CAS Management Service Options” on page 7](#).

Another 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 7](#).

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 will be 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 9](#).
- Identify the data source that you want to access.

You can create a connection between a caslib and any of these databases: DB2, Hadoop Hive, Impala, SAS LASR Analytic Server, ODBC, Oracle, PostgreSQL, Redshift, and Teradata. 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 [SAS Viya Support for Databases](#).

You can add a connection for a database other than the databases listed above. See [“Adding User-Defined Connections” on page 17](#).

If you have licensed SAS Data Preparation, connections to Cloud Data Exchange are available on the **Data Sources** tab. This software enables a CAS server to read and write data that is stored on-premises, behind a firewall. For more information about this software, see [“About Cloud Data Exchange” in *Cloud Data Exchange for SAS Viya: Administrator’s Guide*](#).

- 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](#).

Connection options for Cloud Data Exchange data sources are described in [“Data Agent CAS Library Options” in *Cloud Data Exchange for SAS Viya: Administrator’s Guide*](#).

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 3](#).
- 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 10](#). 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 9](#).
- 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 will be 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 9](#).
- 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 41](#).

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 [“Copying Data from the Available Tab or Data Sources Tab” on page 50](#).

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 will be 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:** `cftblal`.

JDBC Connection Pooling for SAS Data Agent in Cloud Data Exchange

The dataAgentManagement and dataAgentContent microservices use JDBC connection pooling for their connections to the SAS Data Agent. This improves performance and allows for reusing resources. However, sometimes, connection pooling can cause unusual behavior. For example, a pooled connection to a BASE data service will display only the catalogs and schemas that are available when the connection was made. If other catalogs and schemas were added to or removed from the BASE data service after the connection was made, those changes will not be reflected in the pooled connection until the connection times out and a new connection is made.

To disable connection pooling, an administrator can set the service option **dataSourcePoolMaxIdle** to 0. Service options are updated from the Configuration page in SAS Environment Manager. For details about how to access the Configuration page, see [“Configuration Properties: How to Configure Services” in SAS Viya Administration: Configuration Properties](#).

On the Configuration page, select All Services in the View selector. Enter **dataAgentContent** in the **Filter** window. Select the **dataAgentContent** service and set the service option **dataSourcePoolMaxIdle** to 0.

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.
- HDFS (Hadoop Distributed File System).
- 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 [“Adding User-Defined Connections” on page 17](#).

If the connection is successful, the tables that you are authorized to access will be 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, an HDFS, 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 9](#).
- 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 will fail. 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 will be 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 includes an **Include subdirectories** check box. 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. The **Data Sources** tab shows tables and files in a directory but not subdirectories in a directory, for example.

However, if you are creating a path-based caslib as input to the **Documents Directory** feature on the **Import** tab, and you want to access documents in subdirectories, select the **Include subdirectories** check box. Selecting this check box enables the **Select from Documents Directory** window to use a caslib to navigate subdirectories.

- 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](#)

[HDFS 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 3](#).
- 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 10](#). 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**, **HDFS**, 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 9](#).
- 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 will be created. If the new caslib is a session-based caslib, you must specify an existing directory, or the connection will fail. For more information about global and session-based caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 9](#).
- 10 Enter a **Description** for the connection, if desired.
- 11 In most cases, do not select the **Include subdirectories** check box. However, if you are creating a path-based caslib as input to the **Documents Directory** feature on the **Import** tab, and you want to access documents in subdirectories, select the **Include subdirectories** check box. Selecting this check box enables the **Select from Documents Directory** window to use a caslib to navigate subdirectories.
- 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 will be 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 9](#).
- 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 41](#).

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 [“Copying Data from the Available Tab or Data Sources Tab” on page 50](#).

Adding User-Defined Connections

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 7 Redshift Source Type Selected on the Connection Settings Window

The screenshot shows the 'Connection Settings' dialog box. 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.' with a help icon. At the bottom, there are two tabs: 'Settings' (selected) and 'Advanced'. Under the 'Settings' tab, the text 'Specify the Redshift connection information.' is highlighted in yellow, followed by a 'User ID:' label and an input field containing 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. This 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.

To add a 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](#). Plan the connection settings that you want to specify.

Here is a set of example settings for a database connection in the Connection Settings window:

- **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:**
 - Data transfer mode:** Default
 - Number of read nodes:** 3
 - Number of write nodes:** 3
 - Additional Parameters**
 - **Name:** *server-name* **Value:** VTserver
 - **Name:** *user-name* **Value:** user1
 - **Name:** *password* **Value:** myPwd
 - **Name:** *schema-name* **Value:** mySchema

Some of the connection settings are entered as name/value pairs in the **Additional Parameters** section of the **Advanced** tab. In the **Name** field, enter the name of the option that you want to specify, as described in the SAS data connector 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`, as shown above.

To add a user-defined connection to a remote file system, ask the file system administrator for the information required to connect to the target directory. 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](#).

Plan the connection properties that you want to specify. Here is a set of example settings for a remote file system connection in the Connection Settings window:

- **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:**
 - Data transfer mode:** Parallel
 - Number of read nodes:** 3
 - Number of write nodes:** 3
 - Additional Parameters**

- **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*

Add a User-Defined Connection

Use information that you previously gathered to add a new caslib with user-defined connection properties for a database or remote file system.

- 1 Display the window that contains the **Data Sources** tab. See [“Data Selection Windows and SAS Data Explorer” on page 3](#).
- 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 10](#). 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 previous topic.
- 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 9](#).
- 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 **Authentication domain**, **Data transfer mode**, **Number of read nodes**, and **Number of write nodes** according to the information that you previously gathered.
- 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*. See the examples in the overview topic above. You can also click  beside the **Source type** field for details about options for user-defined data sources.
- 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 will be 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 9.

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 41.

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 “[Copying Data from the Available Tab or Data Sources Tab](#)” on page 50.

Importing Local Files

Overview of Importing Local Data Files

Local data 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`.

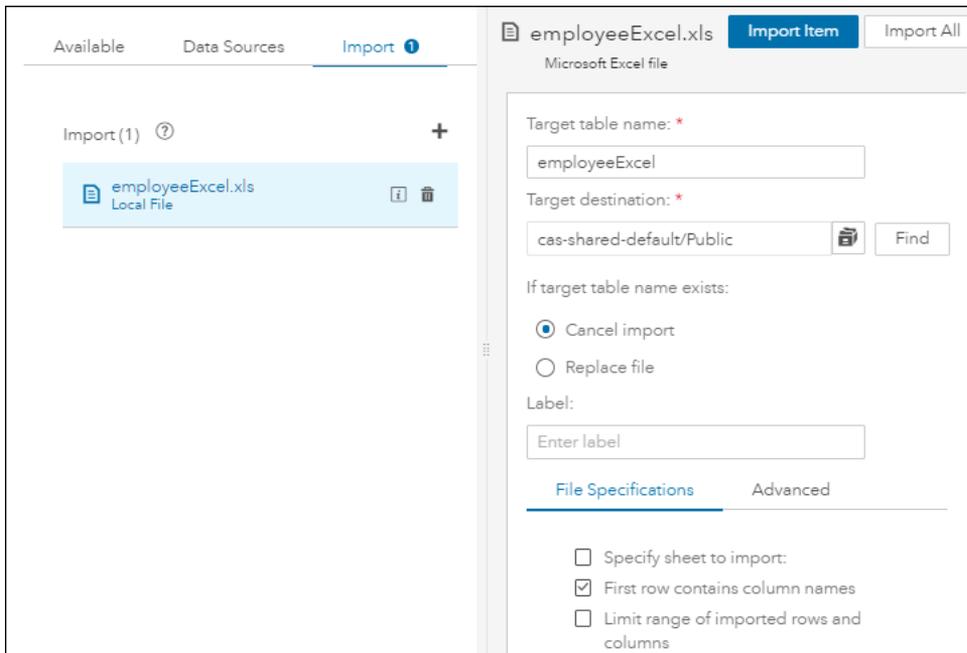
The **Local File** option on the **Import** tab enables you to copy a local file and load it to memory on a CAS server.

Here are the supported file 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.

If you drag and drop a local file onto the **Import** tab, it is added to the queue on the left side, as shown in the following figure.

Figure 8 Excel Spreadsheet in the Queue on the Import Tab



Import properties for the selected file appear on the right. A default table name and caslib for the copy that you will create are provided in the **Target table name** field and the **Target destination** field. Use the **File Specifications** tab to specify options that might be needed to accurately copy the contents of the imported file.

The **Advanced** tab displays the **Unique ID** control for all types of local import files. This control enables you to add a column to the target table that contains a unique value for each row. These unique row identifiers can be used in text topics in SAS Visual Analytics. A tag is added to the new column. You can view the new column and tag on the **Details** tab for the target table. See [“Manage Tags from the Details Tab” on page 43](#).

For delimited text files, the **Advanced** tab also displays the **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 options for local delimited text files.

When ready, you can right-click the file in the import queue and select **Import Item**. A copy of the file 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 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.

Gather the following information to import a local file:

- Identify the local file that you want to load to memory. Verify that the file is in a supported file format.
- Identify the physical path to the local file. Note the size of the file that you want to import. Importing a large table or file can be time-consuming. See [“Notes for All Local Data Files” on page 24](#).
- Gather information about any file specifications that might be needed to accurately copy the contents of the imported file. For example, to read a delimited text file, you might have to specify a delimiter value other than the default delimiter, which is a comma. To review the import options for a file type, see Step 8 in [“Import Local Microsoft Excel, Text, or SAS Data Set Files” on page 22](#).

Note: Any items in the queue on the **Import** tab will be dropped from the queue when you log off. The target files that you load to memory will 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 9](#).

Import Local Microsoft Excel, Text, or SAS Data Set Files

Use information that you previously gathered to copy a local file and load it to memory on a CAS server:

- 1 Display the window that contains the **Import** tab.
See [“Data Selection Windows and SAS Data Explorer” on page 3](#).
- 2 Click the **Import** tab.
- 3 Drag and drop a local file onto the **Import** tab. Alternatively, click **Local File** on the **Import** tab. 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.

- 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 10](#). 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 destination** 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 24](#).

- 5 A default caslib is provided in the **Target destination** 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 10](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 9](#).

- 6 Specify what action the import operation should take if the target filename exists in the caslib that was specified in the **Target destination** field. The options are to cancel the import or to replace the existing item 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 **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 will be 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 will fail to import.

Encryption key

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

- 9 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.

10 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 Data Files](#)” on page 24. See also “[General Usage Notes](#)” on page 9.

Usage Notes for Importing Local Data Files

Notes for 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 `maxFileSizeUploadSize` 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 `maxFileSizeUploadSize` 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 would enter *CAS Management* in the **Filter** window for the **Environment** section of SAS Environment Manager. They would then open the CAS Management service and find the `maxFileSizeUploadSize` option. For more information, see “[Configuration Properties: How to Configure Services](#)” in *SAS Viya Administration: Configuration Properties*.

Notes for 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 will be 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 (.) will be imported as a character type.

Notes for Importing SAS Data Sets

Here are some key points about importing SAS data sets:

- Before importing a SAS data set that uses user-defined formats, ensure that the custom format catalog is available to the CAS server. For more information, see [Manage User-Defined Formats](#).
- Importing indexed SAS data sets is not supported.
- The default value of the **Character multiplier** field is 1. If your data contains international characters, and the data is not importing correctly, you might want to specify a higher value.

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

In general, you can import files that use blanks and special characters in the filenames 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 filename.
- 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 filename.

- If the spreadsheet contains multiple worksheets, then the output table name is initially set as the filename 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 Data from Social Media

Overview of Importing Data from Social Media

The **Social Media** folder on the **Import** tab enables you to import data from social media accounts and to load it to memory on a CAS server. The data can then be analyzed in SAS Visual Analytics and other SAS Viya applications.

For example, if you select the **Facebook feed** option in the **Social Media** folder, a **Facebook feed** item is added to the queue on the left side, as shown in the next figure.

Figure 9 Facebook Feed Item in the Queue on the Import Tab

The screenshot shows the SAS Viya interface. On the left, the 'Import' tab is active, displaying a queue with one item: 'Facebook Facebook feed'. On the right, the 'Facebook' import properties window is open. It includes fields for 'Target table name' (set to 'Facebook'), 'Target destination' (set to 'qstgrd001_37437/Public'), and 'Facebook fan page name' (with an example '(Example: SASsoftware)'). There are also date range fields for 'From' (2017-11-0) and 'To' (2017-11-1). Buttons for 'Import Item' and 'Import All' are visible at the top right of the window.

An import properties window appears on the right side. A default table name and a caslib are provided in the **Target table name** and **Target destination** fields. You can use the Facebook options in the window to select the data that you want from the Facebook feed.

When ready, you can right-click the **Facebook feed** item in the import queue and select **Import Item**. A copy of the specified Facebook data is loaded to memory on the CAS server that is specified in the caslib. Alternatively, you can create multiple import items and import them all at once by clicking **Import All** at 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.

If the import fails, see [“General Usage Notes” on page 9](#).

Note: Any items in the queue on the **Import** tab will be removed from the queue when you log off. The target tables that you load to memory will 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 9](#).

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 7](#).

Import Data from a Facebook Page

The **Facebook feed** option on the **Import** tab enables you to copy data from a Facebook account and load it to memory on a CAS server.

Before you can import information from Facebook, you must grant SAS the right to access your account. Perform these steps:

- 1 Display the window that contains the **Import** tab.
See [“Data Selection Windows and SAS Data Explorer” on page 3](#).

- 2 Click the **Import** tab.

- 3 Expand the **Social Media** folder and select **Facebook feed**.

A message is displayed, indicating that you must grant SAS Visual Analytics the right to access Facebook information.

- 4 Click **OK**.

To load data from a Facebook page to memory on a CAS server:

- 1 On the **Import** tab, expand the **Social Media** folder and select **Facebook feed**.

A **Facebook feed** item is added to the queue on the left side. Import properties appear on the right side.

- 2 The **Target table name** is **Facebook** by default. Accept or change this name 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 10](#). 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 destination** field. Contact your data administrator for more information about the permitted length for target table names.

- 3 A default caslib is provided in the **Target destination** field. To store the data from Facebook elsewhere you can select an existing global caslib or you can add a new one.

If you have the access rights that are required to add a new global caslib, follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 10](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 9](#).

- 4 Specify what action the import operation should take if the target filename exists in the caslib that is specified in the **Target destination** field. The options are to cancel the import operation or to replace the existing item that has the same name.
- 5 Specify a label that will help identify the output table of Facebook information, 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.
- 6 Use the settings under **Facebook Options** to select data from a Facebook page.

Use the **Facebook fan page name** field to specify the name of the page from which you will import data. The name appears at the end of the Facebook URL, such as **MyFanPage** at the end of this URL: <https://www.facebook.com/MyFanPage>.

Use other Facebook options as needed. If your site uses an HTTP Proxy Server to access the internet, enter the proxy host and port.

The **Clear authorization** option enables you to delete the SAS authorization for this Facebook account. If you delete this information, you must reauthorize the account before you can import Facebook data.

- 7 When ready, you can right-click the **Facebook feed** item in the queue and select **Import Item**.

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

If the import fails, see [“General Usage Notes” on page 9](#).

Importing Data from Twitter

Grant Access to Your Twitter Account

The **Twitter feed** option on the **Import** tab enables you to copy data from a Twitter account and load it to memory on a CAS server.

Before you can import information from Twitter, you must grant SAS the right to access your account. Perform these steps:

- 1 Display the window that contains the **Import** tab.
See [“Data Selection Windows and SAS Data Explorer” on page 3](#).
- 2 Click the **Import** tab.
- 3 Expand the **Social Media** folder and select **Twitter feed**.
A message is displayed, indicating that you must grant SAS Visual Analytics the right to access your Twitter account.
- 4 Click **Obtain Access Code**.
A new browser tab opens. Respond as prompted to sign in to your Twitter account, to provide authorization to SAS, and then to copy the access code.
- 5 Return to the tab where you clicked **Obtain Access Code**. Paste the access code into the text box on this tab.
- 6 If your site uses an HTTP Proxy Server to access the internet, enter the proxy host and port. Otherwise, ignore the proxy fields.
- 7 When ready, click **OK**.
A **Twitter feed** item is added to the queue on the **Import** tab. If you are ready to import information, see [“Import Data from Google Analytics” on page 29](#). Otherwise, click  to remove the **Twitter feed** item from the queue.

Import Data from Twitter

After you have granted SAS the right to access a Twitter account, perform the following steps to load data from that account to memory on a CAS server:

- 1 On the **Import** tab, expand the **Social Media** folder and select **Twitter feed**.
A **Twitter feed** item is added to the queue on the left side. An import properties window appears on the right.
- 2 The **Target table name** is **Twitter** by default. 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 10. 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 destination** field. Contact your data administrator for more information about the permitted length for target table names.

- 3 A default caslib is provided in the **Target destination** field. To store the data from Twitter elsewhere, you can select an existing global caslib or you can add a new one.

If you have the privilege required to add a new global caslib, follow the conventions for caslib names as described in “[Names for Caslibs, Tables, and Columns](#)” on page 10. For more information about global caslibs, see “[Caslibs on the Data Sources Tab and Import Tab](#)” on page 9.

- 4 Specify what action the import operation should take if the target filename exists in the caslib that is specified in the **Target destination** field. The options are to cancel the import operation or to replace the existing item that has the same name.
- 5 Specify a label that will help identify the output table of Twitter information, 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.
- 6 Use the settings under **Twitter Options** to select data from the Twitter account.

Use the **Search term** field to select a list of tweets that include the specified term. The search operators that you can use are described on the [Twitter Search page](#).

Use other Twitter options as needed. If your site uses an HTTP Proxy Server to access the internet, enter the proxy host and port. Otherwise, ignore the proxy settings.

The **Clear authorization** option enables you to delete the SAS authorization for this Twitter account. If you delete this information, you must reauthorize the account before you can import Twitter data.

- 7 When ready, you can right-click the **Twitter feed** item in the queue and select **Import Item**.

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

If the import fails, see “[General Usage Notes](#)” on page 9.

Importing Data from Google Analytics

Grant Access to Your Google Analytics Account

The **Google Analytics feed** option on the **Import** tab enables you to copy data from a Google Analytics account and load it to memory on a CAS server.

Before you can import information from Google Analytics, you must grant SAS the right to access your account. Perform these steps:

- 1 Display the window that contains the **Import** tab.
See “[Data Selection Windows and SAS Data Explorer](#)” on page 3.
- 2 Click the **Import** tab.
- 3 Expand the **Social Media** folder and select **Google Analytics feed**.
A message is displayed, indicating that you must grant SAS Visual Analytics the right to access your Google Analytics account.
- 4 Click **Obtain Access Code**.

A new browser tab opens. Respond as prompted to sign in to your Google Analytics account, to provide authorization to SAS, and then to copy the access code.

- 5 Return to the tab where you clicked **Obtain Access Code**. Paste the access code into the text box on this tab.
- 6 If your site uses an HTTP Proxy Server to access the internet, enter the proxy host and port. Otherwise, ignore the proxy fields.
- 7 When ready, click **OK**.

A **Google Analytics feed** item is added to the queue on the **Import** tab. If you are ready to import information, see [“Import Data from Google Analytics” on page 29](#).. Otherwise, click  to remove the **Google Analytics feed** item from the queue.

Import Data from Google Analytics

After you have granted SAS the right to access your Google Analytics account, perform the following steps to load data from that account to memory on a CAS server:

- 1 On the **Import** tab, expand the **Social Media** folder and select **Google Analytics feed**.
A **Google Analytics feed** item is added to the queue on the left. Import properties appear on the right.
- 2 The **Target table name** is **Google Analytics** by default. 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 10](#). 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 destination** field. Contact your data administrator for more information about the permitted length for target table names.
- 3 A default caslib is provided in the **Target destination** field. To store the data from Google Analytics elsewhere, you can select an existing global caslib or you can 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 10](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 9](#).
- 4 Specify what action the import operation should take if the target filename exists in the caslib that is specified in the **Target destination** field. The options are to cancel the import operation or to replace the existing item that has the same name.
- 5 Specify a label that will help identify the output table of Google Analytics information, 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.
- 6 Use the settings under **Google Analytics Options** to select data from the account. To choose useful options, you must understand the information that is available in this account.
If your site uses an HTTP Proxy Server to access the internet, enter the proxy host and port. Otherwise, ignore the proxy settings.
The **Clear authorization** option enables you to delete the SAS authorization for this Google Analytics account. If you delete this information, you must reauthorize the account before you can import Google Analytics data.
- 7 When ready, you can right-click the **Google Analytics feed** item in the queue and select **Import Item**.

If the import is successful, the table of Google Analytics data is loaded to memory on the CAS server that is specified in the caslib. The table of Google Analytics data can be selected from the **Available** tab or the **Data Sources** tab.

The access token for a Google Analytics file will expire after an hour. If your connection to Google Analytics times out, try extracting smaller amounts of data. If the import fails for other reasons, see [“Usage Notes for Importing Local Data Files” on page 24](#).

Importing Data from YouTube

Grant Access to Your YouTube Account

The **YouTube feed** option on the **Import** tab enables you to copy data from a YouTube account and load it to memory on a CAS server.

Before you can import information from YouTube, you must grant SAS the right to access your account. Perform these steps:

- 1 Display the window that contains the **Import** tab.
See [“Data Selection Windows and SAS Data Explorer” on page 3](#).
- 2 Click the **Import** tab.
- 3 Expand the **Social Media** folder and select **YouTube feed**.
A message is displayed, indicating that you must grant SAS Visual Analytics the right to access your YouTube account.
- 4 Click **Obtain Access Code**.
A new browser tab opens. Respond as prompted to sign in to your YouTube account, to provide authorization to SAS, and then to copy the access code.
- 5 Return to the tab where you clicked **Obtain Access Code**. Paste the access code into the text box on this tab.
- 6 If your site uses an HTTP Proxy Server to access the internet, enter the proxy host and port. Otherwise, ignore the proxy fields.
- 7 When ready, click **OK**.
A **YouTube feed** item is added to the queue on the **Import** tab. If you are ready to import information, see [“Import Data from YouTube” on page 30](#). Otherwise, click  to remove the **YouTube feed** item from the queue.

Import Data from YouTube

After you have granted SAS the right to access your YouTube account, perform the following steps to load data from that account to memory on a CAS server:

- 1 On the **Import** tab, expand the **Social Media** folder and select **YouTube feed**.
A **YouTube feed** item is added to the queue on the left side. Import properties appear on the right.
- 2 The **Target table name** is **YouTube** by default. 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 10](#). 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**

destination field. Contact your data administrator for more information about the permitted length for target table names.

- 3 A default caslib is provided in the **Target destination** field. To store the data from YouTube elsewhere, you can select an existing global caslib or you can 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 10](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 9](#).
- 4 Specify what action the import operation should take if the target filename exists in the caslib that is specified in the **Target destination** field. The options are to cancel the import operation or to replace the existing item that has the same name.
- 5 Specify a label that will help identify the output table of YouTube information, 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.
- 6 Use the settings under **YouTube Options** to select data from the account. To choose useful options, you must understand the metrics, the dimensions, and other options that are available for this account.
If your site uses an HTTP Proxy Server to access the internet, enter the proxy host and port. Otherwise, ignore the proxy options.
The **Clear authorization** option enables you to delete the SAS authorization for this YouTube account. If you delete this information, you must reauthorize the account before you can import YouTube data.
- 7 When ready, you can right-click the **YouTube feed** item in the queue and select **Import Item**.
If the import succeeds, the table of YouTube data is loaded to memory on the CAS server that is specified in the caslib. The table of YouTube data can be selected from the **Available** tab or the **Data Sources** tab.
If the import fails, see [“General Usage Notes” on page 9](#).

Importing Files from Google Drive

Grant Access to Your Google Drive Account

The **Google Drive feed** option on the **Import** tab enables you to copy a file from Google Drive and load it to memory on a CAS server. Supported file formats are:

- Comma-delimited (CSV) text files, TXT files, and Google Sheets (gsheet) files.
- SAS data sets (SASHDAT or SAS7BDAT).
- Excel workbook (XLSX) files, Excel 97-2003 workbook (XLS) files, and macro-enabled spreadsheets (XLSM) files. You cannot import XLST, XLSB, or other Excel file types. You cannot import pivot tables.

Before you can import a file from Google Drive, you must grant SAS the right to access your account. Perform these steps:

- 1 Display the window that contains the **Import** tab.
See [“Data Selection Windows and SAS Data Explorer” on page 3](#).
- 2 Click the **Import** tab.
- 3 Expand the **Social Media** folder and select **Google Drive feed**.
A message is displayed, indicating that you must grant SAS Visual Analytics the right to access your Google Drive account.
- 4 Click **Obtain Access Code**.

A new browser tab opens. Respond as prompted to sign in to your Google Drive account, to provide authorization to SAS, and then to copy the access code.

- 5 Return to the tab where you clicked **Obtain Access Code**. Paste the access code into the text box on this tab.
- 6 If your site uses an HTTP Proxy Server to access the internet, enter the proxy host and port. Otherwise, ignore the proxy fields.
- 7 When ready, click **OK**.

The **Import from Google Drive** window appears. If you are ready to import a file, see [“Import a File from Google Drive” on page 32](#). Otherwise, click **Cancel** from this window.

Import a File from Google Drive

After you have granted SAS the right to access your Google Drive account, you can copy a file from Google Drive and load it to memory on a CAS server. Perform these steps:

- 1 On the **Import** tab, expand the **Social Media** folder and select **Google Drive feed**.
The Import from Google Drive window appears.
- 2 Navigate to the file that you want to import. Click the file, and then click **Select**.
A **Google Drive feed** item for the selected file is added to the queue on the left side. Import properties appear on the right.
- 3 The **Target table name** is the name of the selected table by default. 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 10](#). 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 destination** field. Contact your data administrator for more information about the permitted length for target table names.
- 4 A default caslib is provided in the **Target destination** field. To store the file from Google Drive 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 10](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 9](#).
- 5 Specify what action the import operation should take if the target filename exists in the caslib that is specified in the **Target destination** 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 of Google Drive information, 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.
- 7 Use the settings under **File Specification Options** to specify options for the imported file.

Spreadsheet options for Microsoft Excel or Google Sheets:

The following options are available for these 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 will be the first row in the specified range.

Text file options:

The following options are available for 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 will fail to import.

Encryption key

If the file is encrypted, enter the encryption key for the file. Otherwise, the file will fail 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 2.

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 will fail to import.

If your site uses an HTTP Proxy Server to access the internet, enter the proxy host and port. Otherwise, ignore the proxy fields.

The **Clear authorization** option enables you to delete the SAS authorization for this Google Drive account. If you delete this information, you must reauthorize the account before you can import Google Drive data.

- 8 When ready, you can right-click the **Google Drive feed** item in the queue and select **Import Item**.

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

If the import fails, see [“Usage Notes for Importing Local Data Files”](#) on page 24.

Importing Esri Data for Geo-enrichment and Geocoding

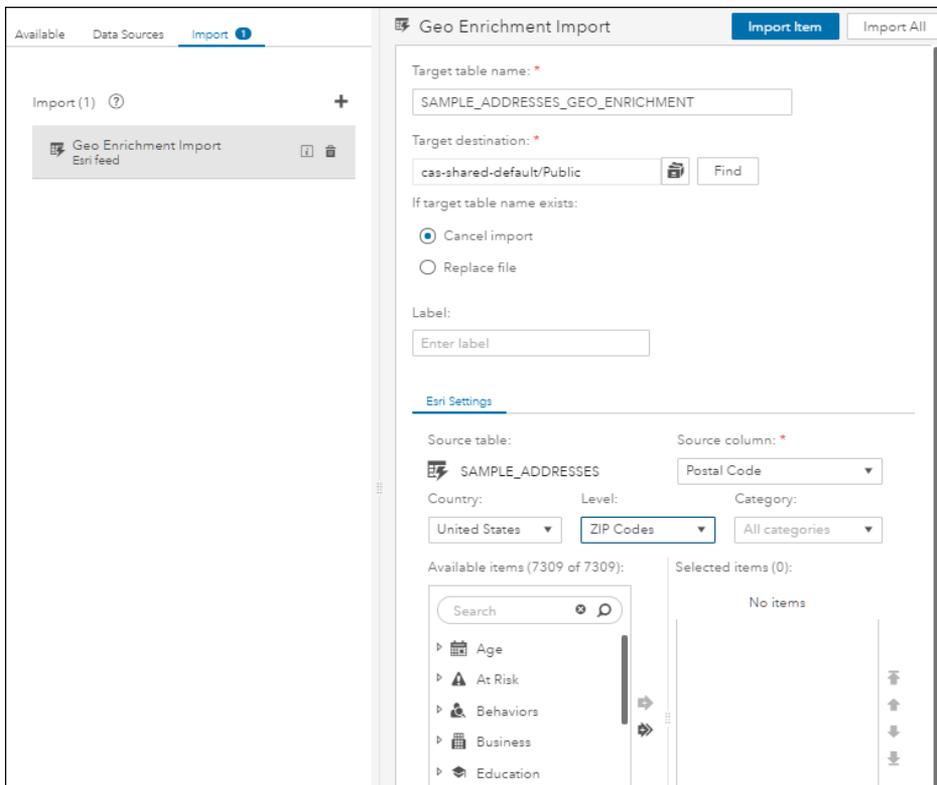
Geo-enrichment

Overview of Esri Geo-enrichment

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 10 Geo Enrichment Import Item in the Queue on the Import Tab



Import properties appear on the right. A default table name and caslib are provided in the **Target table name** and **Target destination** fields. The default table name and the caslib of the output table will combine Esri enrichment data with data from the selected source table. 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 will be dropped from the queue when you log off. The target tables that you load to memory will 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 9.

Prerequisites for Geo-enrichment

Perform the following tasks before using the **Geo-enrichment** 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 [“Identity Management: How To \(SAS Environment Manager\)”](#) in *SAS Viya Administration: Identity Management*.
- 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 **Geo-enrichment 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 Geo-enrichment

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 **Geo-enrichment**.

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 10. 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 destination** field. Contact your data administrator for more information about the permitted length for target table names.

4 A default caslib is provided in the **Target destination** 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 required to add a new global caslib, follow the conventions for caslib names as described in “[Names for Caslibs, Tables, and Columns](#)” on page 10. For more information about global caslibs, see “[Caslibs on the Data Sources Tab and Import Tab](#)” on page 9.

5 Specify what action the import operation should take if the target filename exists in the caslib specified in the **Target destination** 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 will be 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 will be 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 will cost.
- 10 Click **Yes** to continue or **No** to cancel the import.

Esri enrichment data will be combined with the data in the source table and written to the output table. The output table will be loaded to memory on the CAS server that is specified in the **Target destination** 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 Data Files”](#) on page 24.

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 11 Geocode Import Item in the Queue on the Import Tab

The screenshot displays the 'Geocode Import' configuration interface. On the left, the 'Import' tab shows a queue with one item: 'Geocode Import' from 'cas-shared-default/Public'. The main configuration area on the right includes:

- Target table name:** SAMPLE_ADDRESSES_GEO_CODE
- Target destination:** cas-shared-default/Public
- If target table name exists:**
 - Cancel import
 - Replace file
- Label:** Enter label
- Geocode Settings:**
 - Source table:** SAMPLE_ADDRESSES
 - Available columns (0 of 5):** City, State/Province, Postal Code, Country
 - Selected columns (5):** Postal Code, Street Address, City, State/Province, Country
- Sample address:** 24253 5270 Lectus. St. Evansville Florida United States
- Coordinates:** x=-87.56 y=37.98 score=79
- Test** button

Import properties appear on the right. A default table name and caslib are provided in the **Target table name** and **Target destination** fields. These fields specify the name and caslib of the output table that will combine Esri geocode data with data from the selected source table. 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 will be dropped from the queue when you log off. The target tables that you load to memory will 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 9](#).

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 [“Identity Management: How To \(SAS Environment Manager\)” in SAS Viya Administration: Identity Management](#).
- 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 10](#). 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 destination** field. Contact your data administrator for more information about the permitted length for target table names.

- 4 A default caslib is provided in the **Target destination** 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 10](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 9](#).

- 5 Specify what action the import operation should take if the target filename exists in the caslib that is specified in the **Target destination** 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 will cost.
- 11 Click **Yes** to continue or **No** to cancel the import.

Esri geocoding data will be combined with the data in the source table and written to the output table. The output table will be loaded to memory on the CAS server that is specified in the **Target destination** 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 Data Files” on page 24](#).

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 name, the CAS server, and the source type. 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 documents from library	Extracts metadata and text from documents that are accessible from the selected caslib. The extracted information can be analyzed by SAS Visual Text Analytics. This option is available only for PATH caslibs. See “Converting Documents for Analysis” on page 55 .
View authorization	Displays the permissions that each user or group has for the selected caslib. See “CAS Authorization: How to (Authorization Window)” in SAS Viya Administration: Cloud Analytic Services Authorization .
Edit authorization	Enables you to change the permissions that each user or group has for the selected caslib. This option is available if you have ManageAccess permission for the selected caslib. See “CAS Authorization: How to (Authorization Window)” in SAS Viya Administration: Cloud Analytic Services Authorization .
Edit	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.
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 43](#).
- **Sample Data.** Displays sample data from the selected 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 46](#).

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, etc.	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 9 .
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	Displays the permissions each user or group has for the selected table. See “CAS Authorization: How to (Authorization Window)” in SAS Viya Administration: Cloud Analytic Services Authorization .
Edit authorization	Enables you to change the permissions that each user or group has for the selected table. This option is available if you have ManageAccess permission for the selected table. See “CAS Authorization: How to (Authorization Window)” in SAS Viya Administration: Cloud Analytic Services Authorization .
Delete	For in-memory tables: Delete unloads the table from memory and removes its associated physical table. For physical tables: Delete removes the physical table. Refresh the view to verify your changes. Select  from the tool bar.
Add to import	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 “Copy Data from the Data Sources Tab or Available Tab” on page 52 .
Run profile	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 46 .
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. For more information, see “Run Profile and Save” on page 49 .

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 43](#).
- 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 will be added to the new column. See [“Importing Local Files” on page 20](#).
- 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 will be added to the column. See [“Enable Automatic Content Tagging for Columns in a Table” on page 46](#).
- 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

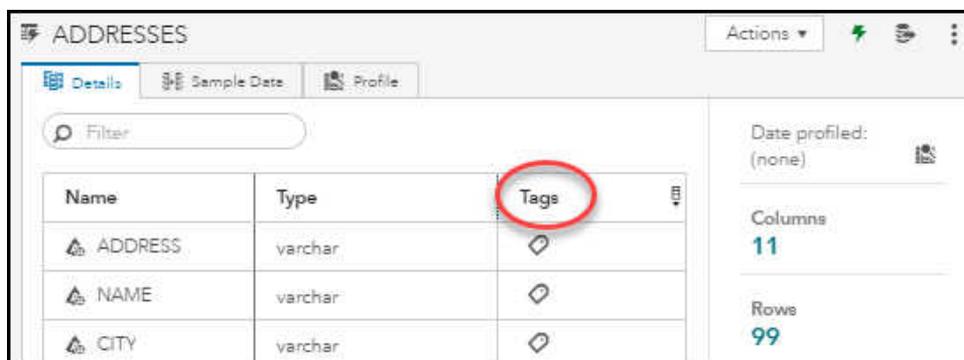
Manage Tags for a Column

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 12 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 13 Choose Tags Window for a ZIP Code Column

The screenshot shows a window titled "Choose Tags" with a close button (X) in the top right corner. Below the title, the "Item name" field contains the text "ZIP". Underneath, the "Selected tags" field is empty. Below that is a search bar labeled "Filter". At the bottom, there is a list of predefined tags, each with a plus sign (+) to its right:

Blank	+
City	+
City-State/Province-Postal Code	+
Company Name	+

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 14 Tags Section at Lower Right on the Details Tab

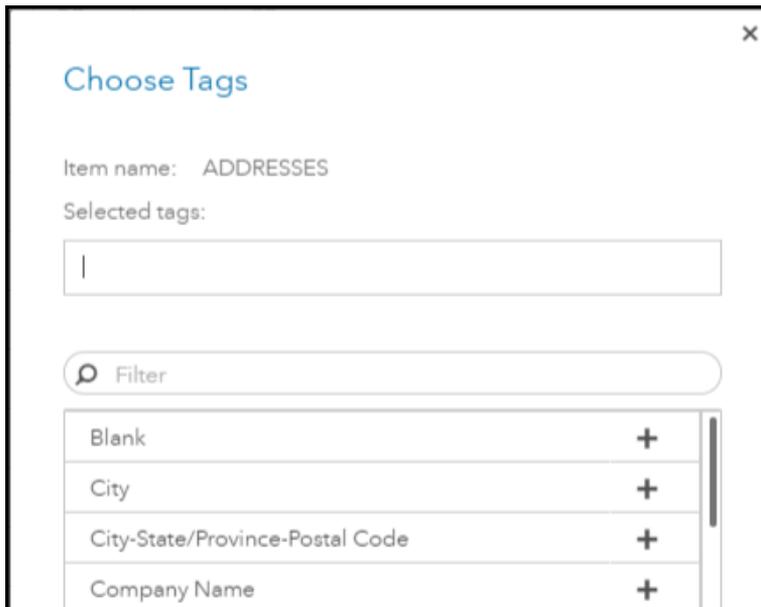


This section displays any tags that have been added to the selected table. It does not display tags that have been added to columns in the table.

- 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 15 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 16 Data Profile Report

Column	Unique	Null	Blank	Pattern Count	Mean	Median
Brand Names	66.67%...	25.00% (3)		6		
ID	100.00% (12)				6.50	6.50
MyDate	87.50% (7)	33.33%...		1		
Numeric Entries	58.33...				1.17	1.00

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.

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 48](#). 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 will be 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 43](#).

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 [“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 8](#).

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 will be analyzed for content. Other columns are ignored.
- Columns with an **ID Analysis** value of UNKNOWN will not get a content tag. For more information about this value, see [“Profile Report for a Column” on page 49](#).

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 17 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 48](#).

- 3 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 which 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 will be 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.

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 displays.
- 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 will persist 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 9](#).

Profiling Tables with a Large Number of Columns

If you profile a table that has thousands of columns of data, the profile might require more memory than is available on your CAS server. In that case, you could try profiling a version of the table that had fewer columns. For information about the **Select Columns** tab, see [“Copying Data from the Available Tab or Data Sources Tab” on page 50](#).

Frequency Distribution and Pattern Metrics Might Be Truncated

When you profile a table, the length of values for the frequency distribution and pattern metrics will be 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 will decide 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 will 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, so 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.

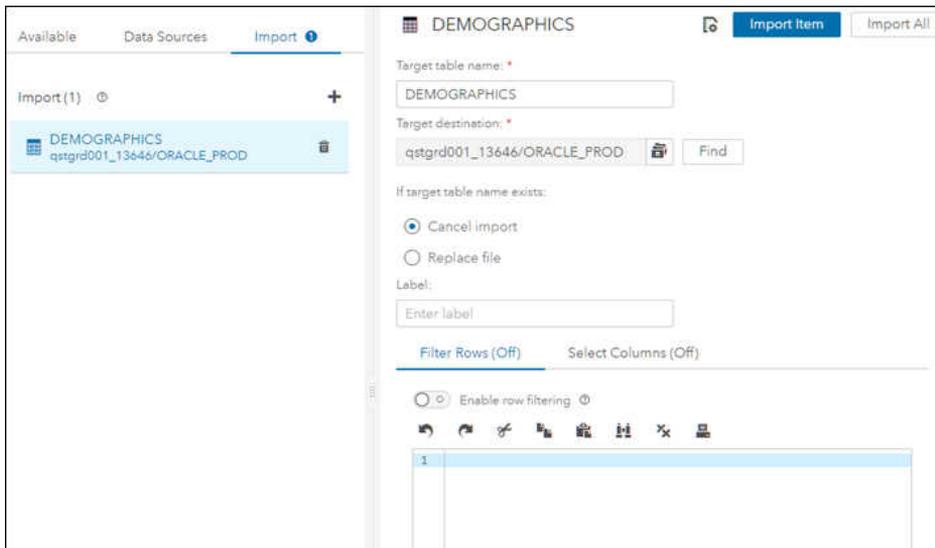
Copying Data from the Available Tab or Data Sources Tab

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 18 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 **Target table name** field is set to the default target location. For more information about the default target location, see [“Modify SAS Data Explorer Settings” on page 8](#). You can accept or change this name as needed. If the source caslib is global in scope, its name will be copied into the **Target destination** field. You can accept or change this name as needed.

Note the **Filter Rows** tab and **Select Columns** tab in the previous figure. 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.

Delimited files stored on a CAS server and Cloud Data Exchange 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.

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](#).

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

- 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.
- If the source caslib is global in scope, its name will be copied into the **Target destination** 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](#).
- The name of the source table or file will be 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.

- 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 54](#).

Note: Any items in the queue on the **Import** tab will be dropped from the queue when you log off. The target tables or files that you load to memory will 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 9](#).

Copy Data 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 3](#).
- 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 set to the default target location. 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 10](#). 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 destination** 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 destination** 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 10](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 9](#).

- 7 Specify what action the import operation should take if the target filename exists in the caslib that is specified in the **Target destination** field. The options are to cancel the import or to replace the existing item that has the same name.
- 8 Specify a label that will help identify the copy, 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 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 54](#).

Delimited files that are stored on a CAS server and Cloud Data Exchange 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. Options are entered as name/value pairs, such as **Name: useNarrowCharacterTypes, Value: TRUE**. Click **Add Option** to add the option to the import record.

10 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 9](#).

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 window in SAS Environment Manager. For more information about using this tab, see [“Jobs: How To” in SAS Viya Administration: Jobs](#).

- 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.

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 [“Copy Data from the Data Sources Tab or Available Tab” on page 52](#) and specify a caslib in the **Target destination** field that is different from the source caslib. This will create 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 [“Additional Considerations for Cross-Caslib Data” in SAS Viya Administration: Cloud Analytic Services Authorization](#).

Filter Rows from an Imported Table or File

As described in [“Copy Data from the Data Sources Tab or Available Tab” on page 52](#), 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 Cloud Data Exchange, 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 [“Copy Data from the Data Sources Tab or Available Tab” on page 52](#).
- 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 Data Files”](#) on page 24.

Select Columns from an Imported Table or File

As described in [“Copy Data from the Data Sources Tab or Available Tab”](#) on page 52, 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 will be 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 [“Copy Data from the Data Sources Tab or Available Tab”](#) on page 52.
- 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 destination** 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 will begin 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 9.

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 [“Copy Data from the Data Sources Tab or Available Tab”](#) on page 52.

- 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 window. The name of the job is based on the name of the table or file that you imported. For more information about using this window, see [“Jobs: How To” in SAS Viya Administration: 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.

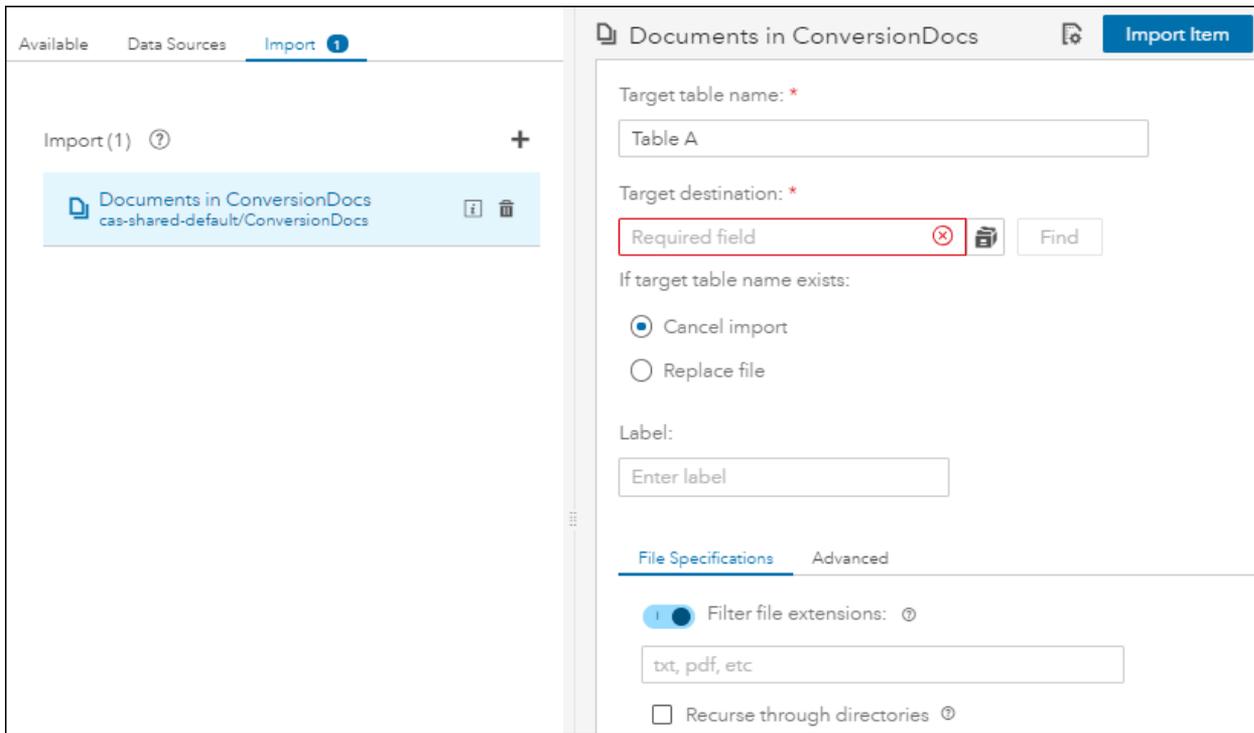
Converting Documents for Analysis

Overview of Document Conversion

You can extract text and metadata from a collection of documents in a caslib and write this information to a table in a caslib. SAS Visual Text Analytics can then analyze the text and metadata in this table.

If you right-click a path-based caslib on the **Data Sources** tab, you can select **Import documents from library**. This adds a **Documents** item to the import queue on the **Import** tab, shown as follows:

Figure 19 Document Item in the Queue on the Import Tab



Alternatively, you can select **Documents Directory** on the **Import** tab, and you are prompted to select a caslib. This caslib points to a directory that contains a set of documents whose content will be extracted for analysis. After you select the caslib, a **Documents** item is added to the import queue on the left, as shown in the previous figure.

Import properties for the **Documents** item appear on the right. You specify various options and import the item. The software uses Tika libraries to extract metadata and text from the documents in the caslib that is specified in the **Documents** item. The Tika libraries were installed with your SAS Viya offering. Output from the extraction is written to the table that you specify in the **Target table name** field. This table is in the caslib that you specify in the **Target destination** field.

Note: Any items in the queue on the **Import** tab will be dropped from the queue when you log off. The target tables that you load to memory will 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 9.

Perform the following tasks before using the **Documents Directory** option:

- 1 Collect a set of documents whose content will be extracted for analysis.

The **Documents Directory** option can be used to extract information from the document formats that are specified in the following table.

Table 3 Document Formats That Can Be Converted

Format	File Extensions
HTML	.html
XML	.xml
MS Office	.doc, .docx, .xls, .xlsx, .ppt, .pptx

Format	File Extensions
Open Document Format	.odt, .ods, .odp
iWork document format	.pages, .numbers, .key
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 is a path-based caslib that points to the directory of documents from which text and metadata will be extracted. To access documents in subdirectories, select the **Include subdirectories** check box when you add the caslib. For more information about adding path-based caslibs, see [“Connecting to Remote File Systems” on page 15](#).

4 Add or identify the target caslib.

This is global caslib on the same CAS server as the source caslib. It will contain the text and metadata that is extracted from the source documents.

Convert Documents for Analysis

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

1 On the **Data Sources** tab, right-click a path-based caslib that contains a set of documents whose content will be extracted for analysis. Select **Import documents from library**.

Alternatively, on the **Import** tab, select **Documents Directory**. The Select from Documents Directory window is displayed. Select a caslib that contains a set of documents whose content will be extracted for analysis.

After you select the caslib, a **Documents** item is added to the queue on the Import queue, and Import properties appear on the right.

2 The **Target table name** is **Table A** by default. This is the name of the output table that will contain 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 10](#). 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 destination** field. Contact your data administrator for more information about the permitted length for target table names.

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

This must be a global caslib on the same CAS server as the source caslib. If you have the privilege required to add a new global caslib, follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 10](#). For more information about global caslibs, see [“Caslibs on the Data Sources Tab and Import Tab” on page 9](#).

- 4 Specify what action the import operation should take if the target filename exists in the caslib that is specified in the **Target destination** field. The options are to cancel the import or to replace the existing item that has the same name.
- 5 Specify a label that will help identify the output 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.
- 6 Use the options on the **File Specifications** tab as needed. Click ⓘ for information about these options.
- 7 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 caslib.
- 8 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 will be written to a table in the target caslib. The output table will be loaded to memory on the CAS server that is specified in the **Target destination** 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 11](#).