



SAS® Data Hub: User's Guide

2026.09

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

Overview of SAS Data Hub	2
About SAS Data Hub	2
Accessing SAS Data Hub	2
Understanding the SAS Data Hub User Interface	3
Work with File Lists on the Data Explorer Page	6
Work with In-Memory Tables and Files	10
Sources Tab: Connect to External Data Sources	10
Import Tab: Copy Data from a Caslib	12
Pop-Up Menu Options for Items on the Data Explorer Page	13
Using SAS Data Hub as a Progressive Web Application	15
Using SAS Viya Copilot for SAS Data Hub	16
What is SAS Viya Copilot for SAS Data Hub?	16
Access the SAS Viya Copilot for SAS Data Hub	16
How SAS Viya Copilot Can Assist You	17
Third-Party Models	17
Best Practices for Using SAS Viya Copilot for SAS Data Hub	19
Get Help with SAS Data Hub	19
Making Data Available to CAS	20
Understanding the Connections Window	20
Connecting to Databases	24
Connecting to Remote File Systems	28
User-Defined Connections for Data Sources	30
Importing Data from SAS Folders	36
Importing Local Files	40
Importing Esri Data for Geoenrichment and Geocoding	47
Working with Data in CAS	55

Working with Caslibs and Tables	55
Copying Data from the In-Memory Data Item or the Sources Tab	59
Import a Snapshot of an ESP Project	67
Importing Delimited Files from a Caslib Directory	68
Importing Documents from a Caslib Directory	71
Importing Images from a Caslib Directory	74
Importing ORC Files from a Caslib Directory	78
Managing the Access Policy for a Connection	81
Access Type Values	81
Maintaining SAS Data Hub	82
Administrative Tasks for SAS Data Hub and the Data Selection Windows	82
SAS Data Hub Options	83
General Usage Notes	84

Overview of SAS Data Hub

About SAS Data Hub

SAS Data Hub is a tool in the SAS Viya platform that enables you to access, manage, and explore data on a SAS Cloud Analytic Services (CAS) server. It provides an easy-to-use interface for loading data from various sources, including on-premises systems and cloud-based platforms. SAS Data Hub can be used independently as a web application or accessed from within other SAS Viya applications.

SAS Data Hub enables you to manage data connections on the Manage Connections page and on the **Sources** tab of the Data Explorer page.

SAS Data Hub enables you to access data assets in the following ways on the Data Explorer page:

- Click the **In-memory data (available)** item on the **Sources** tab to access tables and files that have been loaded to memory from any CAS server to which you have access.
- Review and manage the sources listed on the **Sources** tab. The sources are grouped by the servers that contain them.
- Click the appropriate items on the **Import** to import the data assets that you need.

Accessing SAS Data Hub

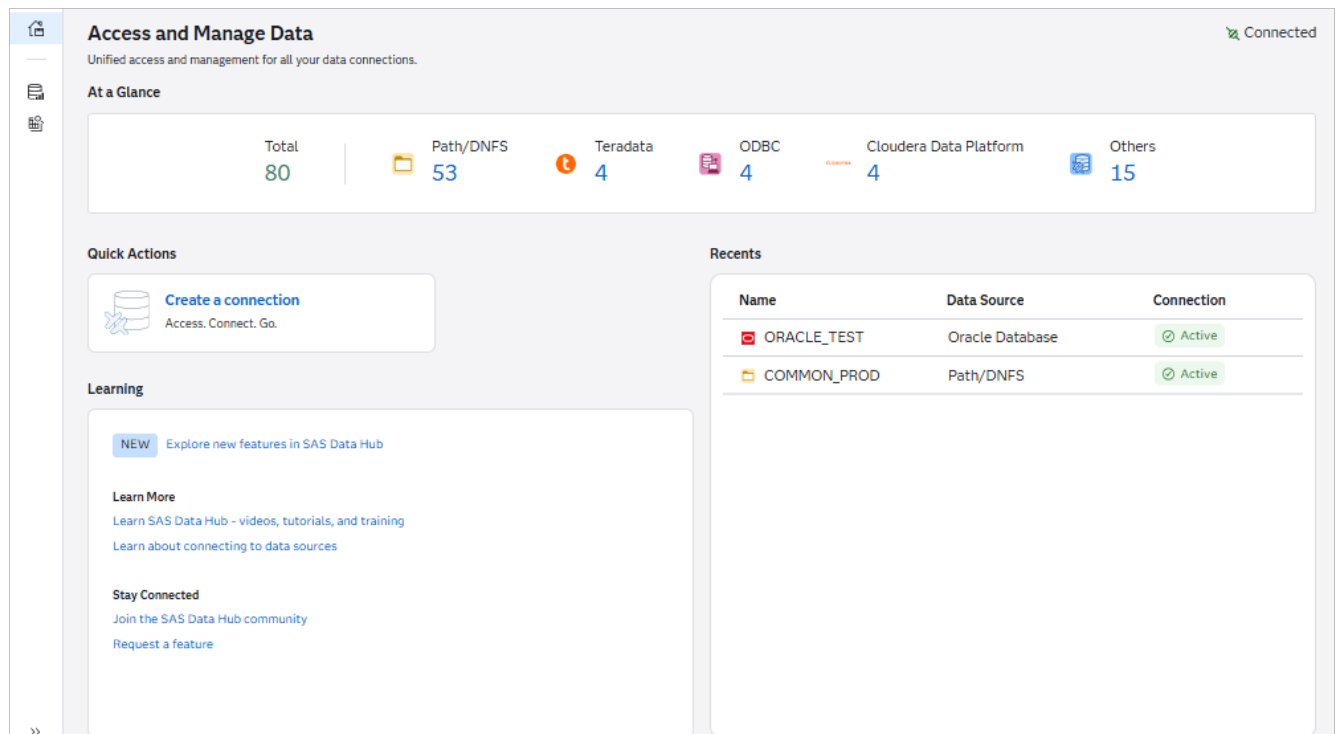
To display SAS Data Hub, click  at the upper left of SAS Viya platform applications and then select **Access and Manage Data**. SAS Data Hub is displayed.

Understanding the SAS Data Hub User Interface

About the Home Page

After you sign in to SAS Data Hub, the home page appears. The home page provides an overview of your connections and links to resources.

Here is an example of a home page:



The home page displays the following elements:

At a glance

displays a summary of the connections that you can access.

Quick Actions

provides shortcuts to actions such as creating a new connection.

When your environment contains multiple CAS servers, the default CAS server is used for the quick actions. If the default server is not available, then the first available server in alphabetical order is used.

Learning

displays a set of links to documentation and other resources that can help you learn SAS Data Hub.

Recents

displays the connections that you have created or modified recently.

About the Manage Connections Page

The Manage Connections page provides an overview of all of your data connections, and enables you to create, view, and manage them.

Here is an example of a Manage Connections page:

Manage Connections

At a Glance

Total: 75

Path/DNFS: 51

Teradata: 4

ODBC: 4

Cloudera Data Platform: 4

Others: 12

Name	Short Name	Connection	Date Modified	Data Source
TERADATA_TEST_P...	C3_0U6T3	Active	Aug 29, 2026, 2:00:09 AM	Teradata Vantage on Azure
TERADATA_TEST_S...	C3_3J705	Active	Aug 29, 2026, 1:49:54 AM	Teradata Vantage on Azure
TERADATA_TEST	C3_1Q8V2	Active	Aug 29, 2026, 1:49:41 AM	Teradata Vantage on Azure
TERADATA_PROD...	C3_GC2E1	Active	Aug 29, 2026, 1:49:28 AM	Teradata Vantage on Azure
DB_PGRES	DB_PGRES	Active	Aug 29, 2026, 1:49:10 AM	PostgreSQL
VERTICALIB_NEW	C3_AQ5L1	Active	Aug 29, 2026, 1:48:56 AM	Vertica
ORACLE_TEST	C3_H2092	Active	Aug 29, 2026, 1:48:43 AM	Oracle Database
ORACLE_PROD	C3_XX7Z6	Active	Aug 29, 2026, 1:48:30 AM	Oracle Database
ODBC_SQLS_PROD	C3_HT182	Active	Aug 29, 2026, 1:48:17 AM	ODBC
ODBC_SQLS_TEST...	C3_SX403	Active	Aug 29, 2026, 1:48:04 AM	ODBC
ODBC_SQLS_TEST	C3_6G750	Active	Aug 29, 2026, 1:47:51 AM	ODBC
ODBC_SQLS_TARG...	C3_3G811	Active	Aug 29, 2026, 1:47:37 AM	ODBC
LASR_TEST	C3_9Z1Y4	Active	Aug 29, 2026, 1:47:24 AM	SAS LASR
LASR_PROD	C3_TP1X3	Active	Aug 29, 2026, 1:47:11 AM	SAS LASR
IMPALA_TARG	C3_6O993	Active	Aug 29, 2026, 1:46:57 AM	Impala

Overview (Active)

Description

Properties

- Data source: Oracle Database
- Connection: ORACLE_TEST
- Short name: C3_H2092
- Scope: Global
- CAS server: cas-shared-default
- Compute context: Not available
- Created by: sas.datasources
- Date created: Aug 29, 2026, 1:48:43 AM
- Modified by: sas.datasources
- Date modified: Aug 29, 2026, 1:48:43 AM

The Manage Connections page displays a list of metrics that summarize your connections and a table with details about your connections. For details about creating connections, see [“Making Data Available to CAS” on page 20](#).

Note: New caslibs that were created in SAS Data and AI Studio might not appear on the Manage Connections page. To enable them to appear, edit and then save the connections on the Data Explorer page.

The connections table can display the following status values in the **Connection** column:

Active

The connection is active in the current session.

Active (Out-of-Sync)

The connection has been edited, but the changes have not been applied to the current connection. The connection is still active using the previous settings. The test status and any logs for this connection are cleared.

To synchronize the connection settings and return the connection to **Active** status, right-click the connection and select **Reconnect**.

Active (Warnings)

The connection is active, but warnings were found while connecting. Click **View** in the **Log** column to see the warnings.

Inactive (Error)

The connection is inactive and issues were found when connecting. Click **View** in the **Log** column to see the errors.

Inactive

The connection is inactive.

Connecting

The connection is in the process of connecting.

You can right-click a connection in the table to perform actions such as editing properties, testing the connection, disconnecting the connection, and so on. Depending on your permissions, some actions might not be available for some of the connections.

The **Log** column in the connections table shows any logs that are available. Click **View** to see the details.

You can access the side pane to view the properties, destinations, and access policy for the selected connection. Click one of the following tabs to open the side pane:

 Overview

displays the description and properties for the selected connection.

 Destinations

displays the destinations for the selected connection. Destinations can apply the connection settings to multiple CAS and compute servers.

You can click **Edit** to manage the server destinations for the selected connection.

 Access Policy

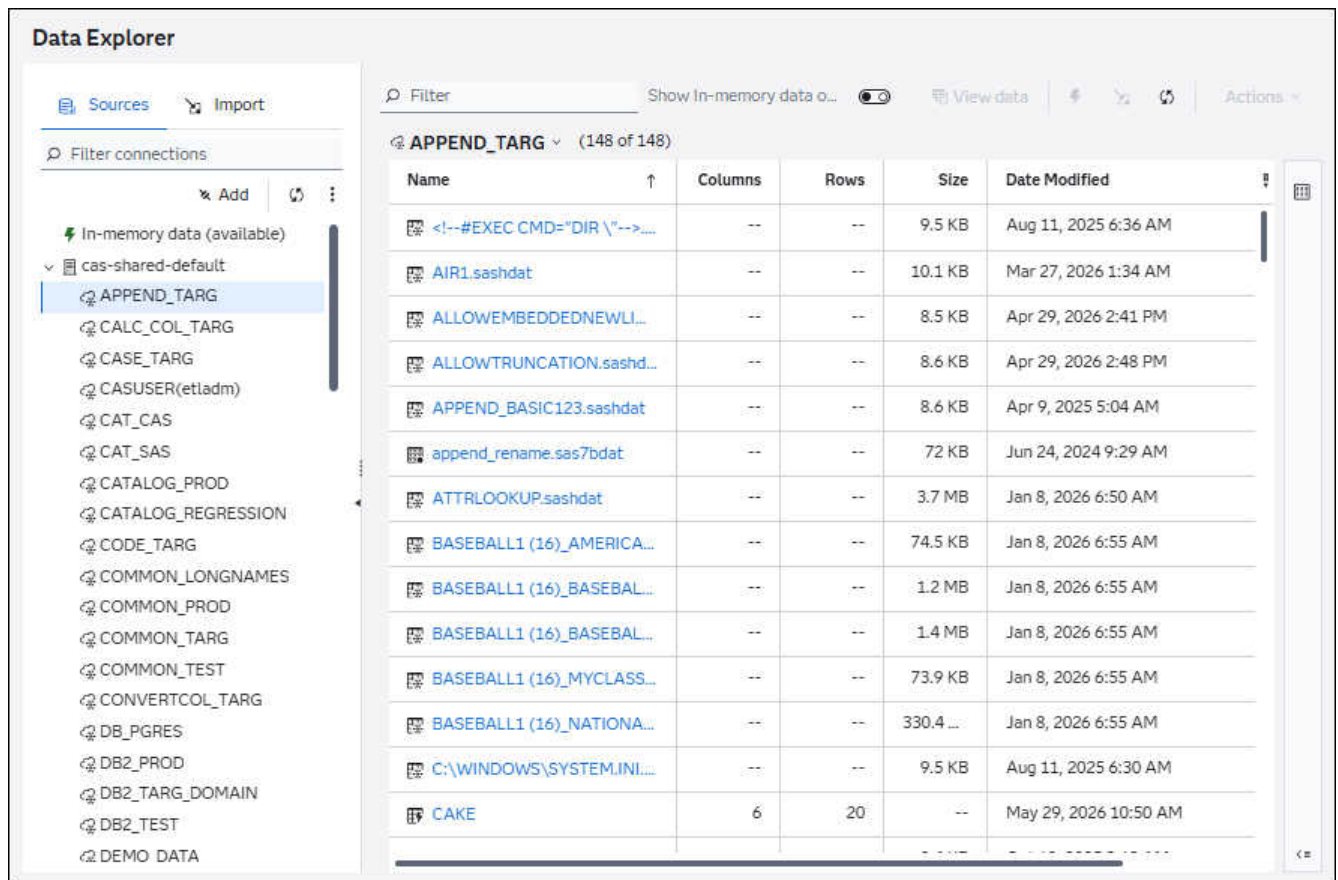
displays the access policy settings for the selected connection.

You can click **Edit** to open the Access Policy window and manage the access policy settings for the connection. For more information, see [“Managing the Access Policy for a Connection” on page 81](#).

About the Data Explorer Page

The Data Explorer page provides the most detailed view of your data assets and enables you to import local files.

Here is an example of the Data Explorer page:



When the **Sources** tab is selected, the page displays a list of all the connections in the left pane, and a list of assets in right pane.

Click the **In-memory data (available)** item in the left pane to access tables and files that have been loaded to memory from any CAS server to which you have access.

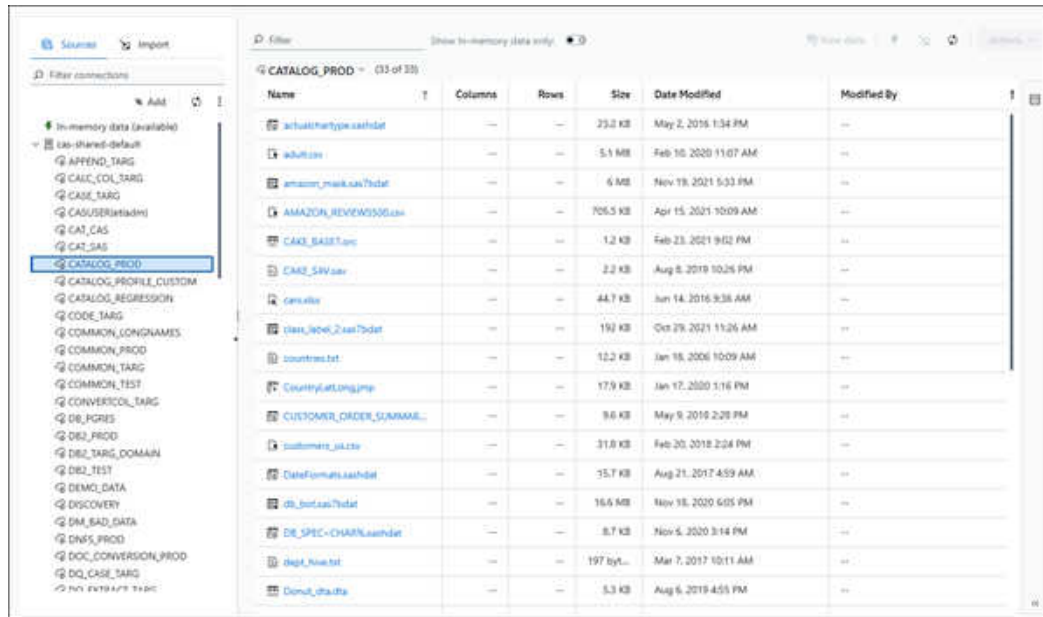
To load an asset into memory, navigate the connections in the left pane and then select the asset that you want to load in the right pane. Click ⚡ to load the selected asset.

When the **Import** tab is selected, you can import data assets by clicking the links in the left pane or by dragging files onto the right pane. For more information, see [“Importing Local Files” on page 40](#).

Work with File Lists on the Data Explorer Page

When you click **In-memory data (available)** or **Sources**, the data assets are displayed in a list to the right of the **Sources** tab.

Figure 1 Sources File List



A set of management tools is displayed above the list of files. Slightly different tools are displayed for the **Sources** tab list and the list that is displayed when you click the **in-memory data (available)** item because connected sources need different tools than in-memory items.

Table 1 File List Tools

Name	Function
Filter field	Enables you to filter on the names of the items in the caslib. Enter a text string that is contained in the name of the item or items that you need.
Show in memory data only button	When clicked, only in-memory data items are displayed in the list for the selected caslib.
View data button	When clicked, displays the Details tab for the item selected in the list.
Load into memory (⚡) button	When clicked, loads the selected item into memory.
Add to import (📁) button	When clicked, adds the selected item to the current import.
Refresh (↻) button	When clicked, refreshes the list for the selected caslib.
Actions	Enables you to choose an action to perform for the selected item. See “Pop-Up Menu Options for Items on the Data Explorer Page” on page 13.

Name	Function
Manage columns ()	When clicked, displays the Manage Columns window. Use this window to hide or display columns in the current list.
Connection properties ()	When clicked, displays the connection properties for the selected item.

Click  to see the connection properties for the selected caslib.

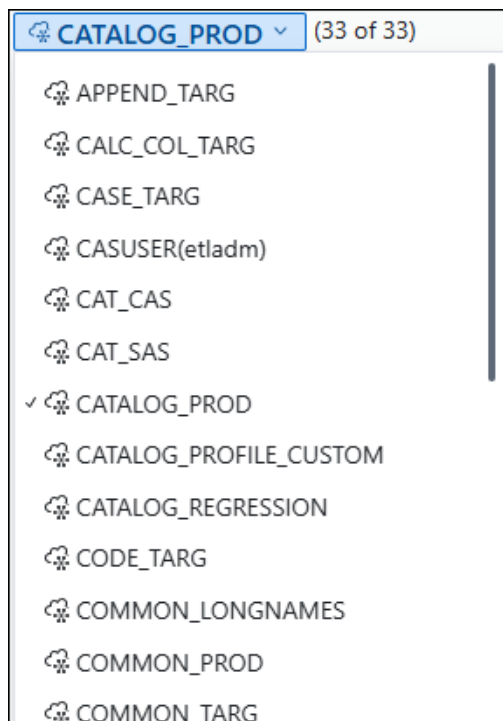
Figure 2 *Connection Properties*

Connection Properties

Name:	CATALOG_PROD
Description:	--
Server:	cas-shared-default
Source type:	PATH
Personal:	false
Path:	/dmtesting/SASViya/Data/Catalog_Prod/
Include subdirectories:	false

When you display the list for a caslib, you can click the down arrow next to the name to see a list of the other caslibs available on the currently selected server.

Figure 3 Caslibs Breadcrumbs



Breadcrumbs are also available from the **Details** or **Sample data** tab view for a selected table.

Figure 4 Table Breadcrumbs

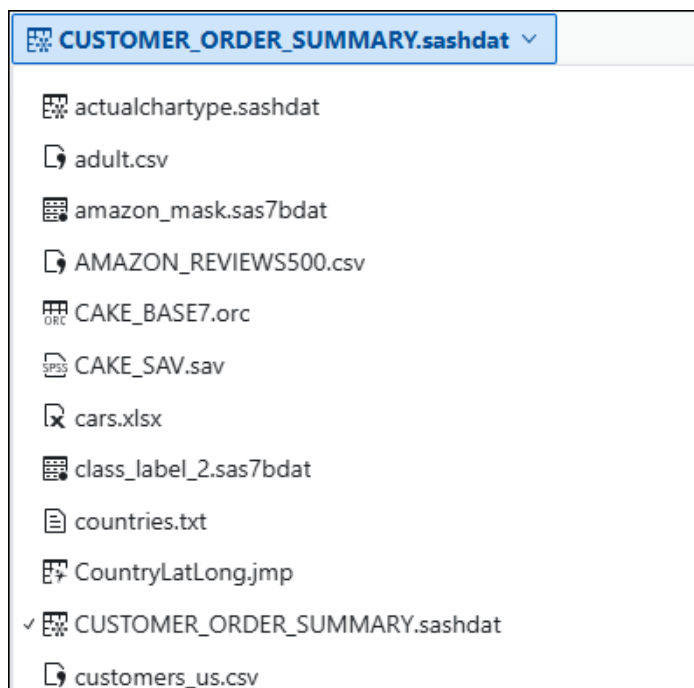
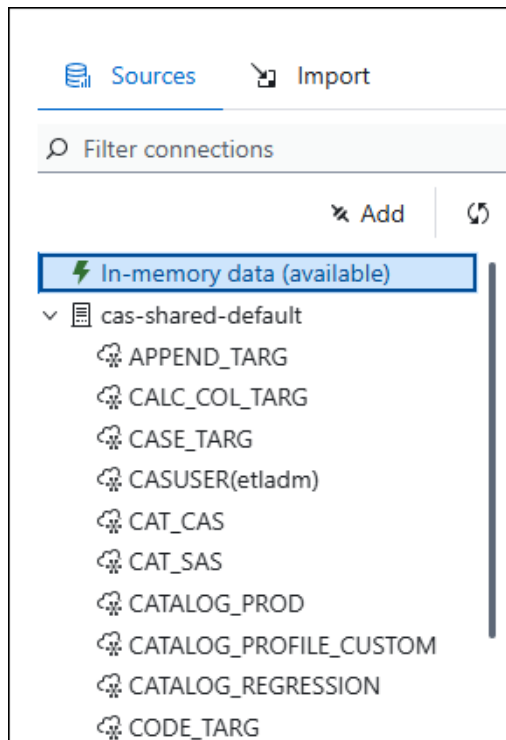
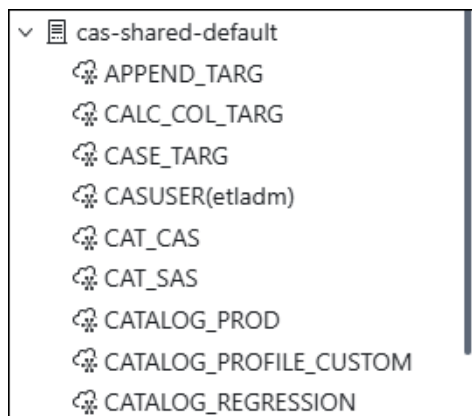


Figure 6 Data Selection Window with the Sources Tab Selected



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 7 Caslibs on a Selected Server

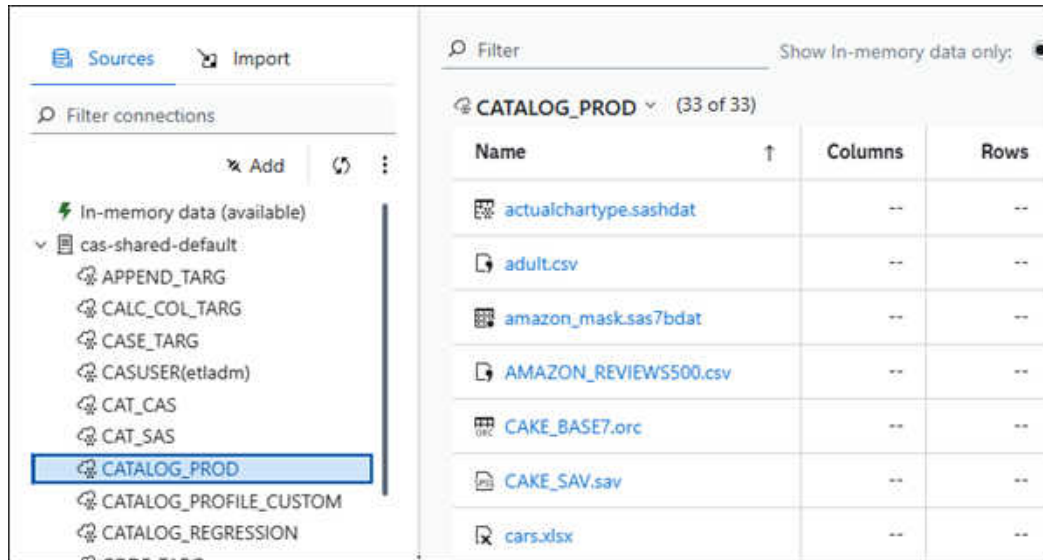


Items with  beside them are global caslibs on the selected CAS server. Global caslibs with names such as CASUSER(**user-ID**) or CASUSERHDFS(**user-ID**) are personal caslibs. Only your user ID can access the data in your personal caslib.

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 Sources Tab and Import Tab” on page 84](#). If you select a caslib, you can click  to see connection properties for that caslib. 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 55](#).

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

Figure 8 Tables in a Selected Caslib



Items with  beside them are tables that are accessible to the CAS server that have not been loaded to memory. Note that the physical table icons in the listing above are marked with badges that identify the type of data that is contained in the table. For example, the row for CAKE_BASE7.orc in the listing above is displayed with the , which identifies the file as an ORC table. Similarly, the row for adult.csv is displayed with the , which identifies the file as a CSV table.

Items with  are tables that have been loaded to memory. If you select a table, two tabs of information are displayed: **Details** and **Sample data**. If you right-click a table, you can select various options, such as **Load into memory**. For more information about these tabs and options, see [“Working with Tables” on page 56](#).

Import Tab: Copy Data from a Caslib

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

- [“Copying Data from the In-Memory Data Item or the Sources Tab” on page 59](#)
- [“Import a Snapshot of an ESP Project” on page 67](#)
- [“Importing Delimited Files from a Caslib Directory” on page 68](#)
- [“Importing Documents from a Caslib Directory” on page 71](#)
- [“Importing Images from a Caslib Directory” on page 74](#)
- [“Importing ORC Files from a Caslib Directory” on page 78](#)

Administrators can change the default rules and deny access to the **Import** tab. For more information, see [“Limit Access to the Import Tab” on page 82](#).

Pop-Up Menu Options for Items on the Data Explorer Page

If you right-click a table the **In-memory data (available)** list or the **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 In-Memory Data List or Sources Tab

Name	Action
Actions ⇨ ...	Opens the selected table in another application such as SAS Visual Analytics. For more information, see the documentation for that application.
Load into memory	Loads the selected table to memory on a CAS server. The Load into memory option is available only for tables in a global caslib. For more information about global caslibs, see “Caslibs on the Sources Tab and Import Tab” on page 84 .
Unload from memory	Unloads the selected table from memory on a CAS server. The table must be loaded again if you want to work with the table in CAS.
View authorization	If available, displays the permissions that each user or group has for the selected table. See “CAS Authorization Window” in SAS Viya: CAS Authorization Window .
Edit authorization	If available, enables you to change the permissions that each user or group has for the selected table. The ManageAccess permission is required to modify access controls. See “CAS Authorization Window” in SAS Viya: CAS Authorization Window .
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 toolbar. Note: A SAS data set (.sas7bdat file) deletion might fail if the file was created on a different platform than the one that the CAS server runs on. For example, your attempt to delete a Windows file from a CAS server that runs on Linux might fail. In this case, you can use the DATASETS procedure to delete the data set.

Name	Action
Remove the access controls associated with the source file	The Delete window that is displayed when you delete either an in-memory table or a physical table includes the Remove the access controls associated with the source file check box. Select this option to disable the authorization that is set with the Edit authorization option for the table. The check box is disabled when Edit authorization is not granted to the user. The Edit authorization option is explained earlier in this table of pop-up menu options. Note: The access controls applied to the source table are persisted independently of the source table itself. For example, delete a source table without selecting Remove the access controls associated with the source file. If you then create a new source table in that caslib with the same name as the deleted one, the new source table has the same persisted access controls applied to it as the original source table.
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 In-memory data (available) list or the Sources tab. For more information, see “Add to Import from the Sources Tab or the In-Memory Data List” on page 62.
Download table	If available, downloads the selected in-memory table to your computer and saves it as a delimited file. You can select CSV or TSV format. Administrators can limit access to this option. See “Limit Access to the Download Table Option” on page 82. Note: The selected table must contain data that can be saved to CSV or TSV format. For example, you cannot download a binary image file to a CSV file and get useful data.
Assume Superuser role	If available, enables you to assume the Superuser role. This role provides you with the maximum ability to review and manage data. Click Relinquish Superuser role when you no longer need this level of control of the data.
Show hidden sources	If available and selected, displays hidden data sources. You must click  on the Sources tab to display the hidden sources.

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 66.](#)

Using SAS Data Hub as a Progressive Web Application

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

- Application persistence – By default, your session will never time out, so you can restart your work more quickly. Your administrator can enable a time-out for the PWA using a configuration property in SAS Environment Manager. For more information, see [“sas.htmlcommons” in SAS Viya Platform: Configuration Properties](#).
- Performance – When installed as a PWA, SAS Data Hub typically starts faster than when accessed in the browser.
- Desktop experience – Installing SAS Data Hub as a PWA confers all the benefits of a traditional installation. You can launch SAS Data Hub from the **Start** menu or taskbar, and you do not need to sort through many tabs to find the correct instance of SAS Data Hub.
- Renaming – You can rename each PWA instance to quickly access different environments, such as development, test, or production servers.

To install the PWA:

- 1 Open any other SAS web app (such as SAS Lineage) in a Chromium-based web browser (such as Chrome).
- 2 Do one of the following:
 - Click the appropriate icon at the end of the address bar, and then select **Install SAS**.
 - Open the web browser’s **More** menu and select **Install SAS**.

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

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

To uninstall the PWA:

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

Using SAS Viya Copilot for SAS Data Hub

What is SAS Viya Copilot for SAS Data Hub?

SAS Viya Copilot is a set of software as a service (SaaS) features and capabilities that use large language models (LLMs) to provide users with a more intuitive and accessible way to work with the SAS Viya platform.

SAS Viya Copilot for SAS Data Hub has capabilities that are tailored for data management. The copilot uses an LLM that includes the official SAS documentation for SAS Data Hub to help ensure that the responses are relevant and accurate for users.

The copilot capabilities in SAS Data Hub are surfaced through a conversational chat interface in the **SAS Viya Copilot** pane.

For an overview of SAS Viya Copilot, see [SAS Viya Copilot: Foundation Overview](#). SAS Viya Copilot is not enabled by default. It must be enabled by your SAS Viya administrator in SAS Environment Manager. See “[Enable SAS Viya Copilot in SAS Environment Manager](#)” in [SAS Viya Copilot: Foundation Overview](#). When SAS Viya Copilot is enabled, all of its capabilities are available for your use.

Access the SAS Viya Copilot for SAS Data Hub

IMPORTANT English is the only officially supported language for prompts and output. There are no localized versions of SAS Viya Copilot.

To use the copilot:

- 1 Click  on the application bar to open the **SAS Viya Copilot** pane.
- 2 Enter request criteria in the text field or select one of the example prompts, and then click ➤.

Note: Getting the best AI results can be an iterative process. Your first prompt might not get you the best results.

- 3 (Optional) Clear the current conversation and start a new one with the copilot by clicking ✖.

How SAS Viya Copilot Can Assist You

SAS Viya Copilot for SAS Data Hub can assist you by retrieving help from the user documentation. For more information, see [“Get Help with SAS Data Hub” on page 19](#).

IMPORTANT It is your responsibility to review any output generated with AI to ensure that the AI-generated content that you select is appropriate and accurate for its intended purpose.

Third-Party Models

Model Overview

SAS Viya Copilot is powered by a combination of proprietary SAS technology and Microsoft Azure technology to help you securely access, use, and manage your data. The Microsoft Azure technology used for SAS Viya Copilot centers on cloud-based infrastructure for building and deploying scalable applications.

For a list of the Microsoft Azure models that are used by SAS Viya Copilot, see “Models Used by SAS Viya Copilot” in [SAS Viya Copilot: Foundation Overview](#).

Input and Output Modalities

Input modalities and size limit for each modality as used by SAS:

- gpt-4.1: Text, 1,047,576 tokens
- text-embedding-3-large: Text, 8,192 tokens

Output modalities and size limit for each modality as used by SAS:

- gpt-4.1: Text, 32,768 tokens
- text-embedding-3-large: Embedded, Fixed-size vector (1536 dimensions)

.....
Note: Because embedding models return a fixed-dimension vector rather than text tokens, there is no published output token size.
.....

Prerequisites

- SAS Data Hub 2026.09 or later

Using the Model

- Acceptable use policy: <https://www.sas.com/acceptableuse>.
- Intended use: SAS Viya Copilot provides a chat interface that enables you to retrieve information or instructions from the product documentation.

Type and Nature of AI Systems

The models support features in SAS software in a specific context but cannot be more broadly incorporated into customer AI systems.

Best Practices

When conversing with SAS Viya Copilot, follow these best practices for writing your prompts:

- Be as clear and specific as possible in your prompts.
- Be intentional with the tone of your prompts because the AI-generated suggestions often match the tone of the original prompt.
- Ensure that your prompts are safe and respectful.
- Enter your prompts only in natural language English.
- Avoid words or phrases in your prompts that are culturally specific, such as slang or idioms that are not easily understood by others.
- If you are not getting good results, try rephrasing the prompt.
- If you need more specific information, ask follow-up questions.

Known Issues or Limitations

- SAS does not guarantee that the AI responses are correct, or that the generated suggestions are optimal.
- LLMs are known to hallucinate.
- SAS Viya Copilot supports only English inputs and output.

For more information, see “[SAS Viya Copilot Limitations](#)” in *SAS Viya Copilot: Foundation Overview*.

Best Practices for Using SAS Viya Copilot for SAS Data Hub

Getting the best AI suggestions from SAS Viya Copilot can be an iterative process. Your first prompt, or question, might not get you the best results. If the reply is not what you are looking for, you can try rewriting your question or asking follow-up questions.

You can improve the quality of the AI-generated suggestions from SAS Viya Copilot by keeping the following best practices in mind when writing your prompts or questions.

IMPORTANT It is your responsibility to review any output generated with AI to ensure that the AI-generated content that you select is appropriate and accurate for its intended purpose.

- SAS Data Hub users interact with the copilot using the **SAS Viya Copilot** pane, which assists users in data management. Through natural language input, users can ask for help with using the product.
- When conversing with the copilot, follow the [best practices on page 18](#) for writing prompts.
- Ensure that the prompts are safe and respectful.
- Avoid words or phrases in the prompts that are culturally specific, such as slang or idioms that are not easily understood by others.
- If the results to a prompt are not sufficient, try rephrasing the prompt.
- If more specific information is needed, ask follow-up questions.

Get Help with SAS Data Hub

You can ask the copilot how to perform tasks in SAS Data Hub by entering natural-language questions in the **SAS Viya Copilot** pane. Here are some example questions:

- `What is Data Hub?`
- `Do I need to be an administrator to manage connections?`
- `What data sources are supported?`
- `How do I secure a connection?`
- `Where can I see why my connection failed?`
- `How do I view the lineage of my data?`
- `How do I load a table into memory?`
- `How do I import a CSV file?`

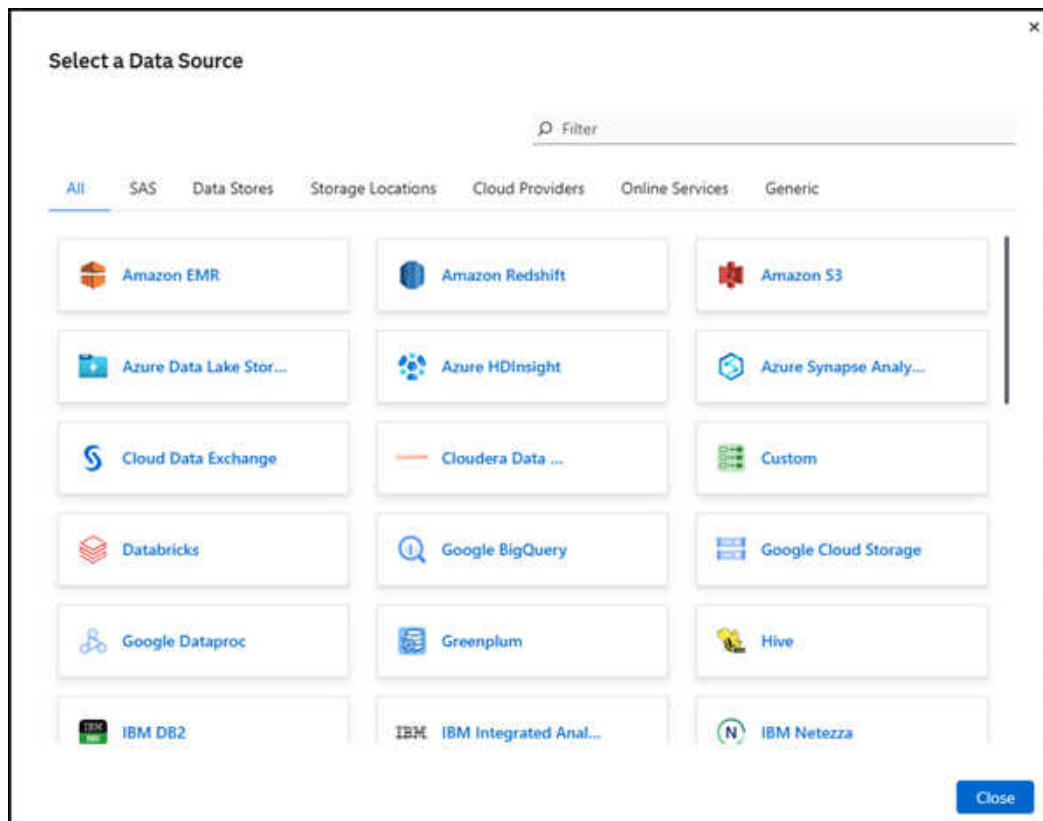
At the bottom of the response, the copilot provides reference links to the documentation sources. You can follow the links to read additional information and to verify the copilot output.

Making Data Available to CAS

Understanding the Connections Window

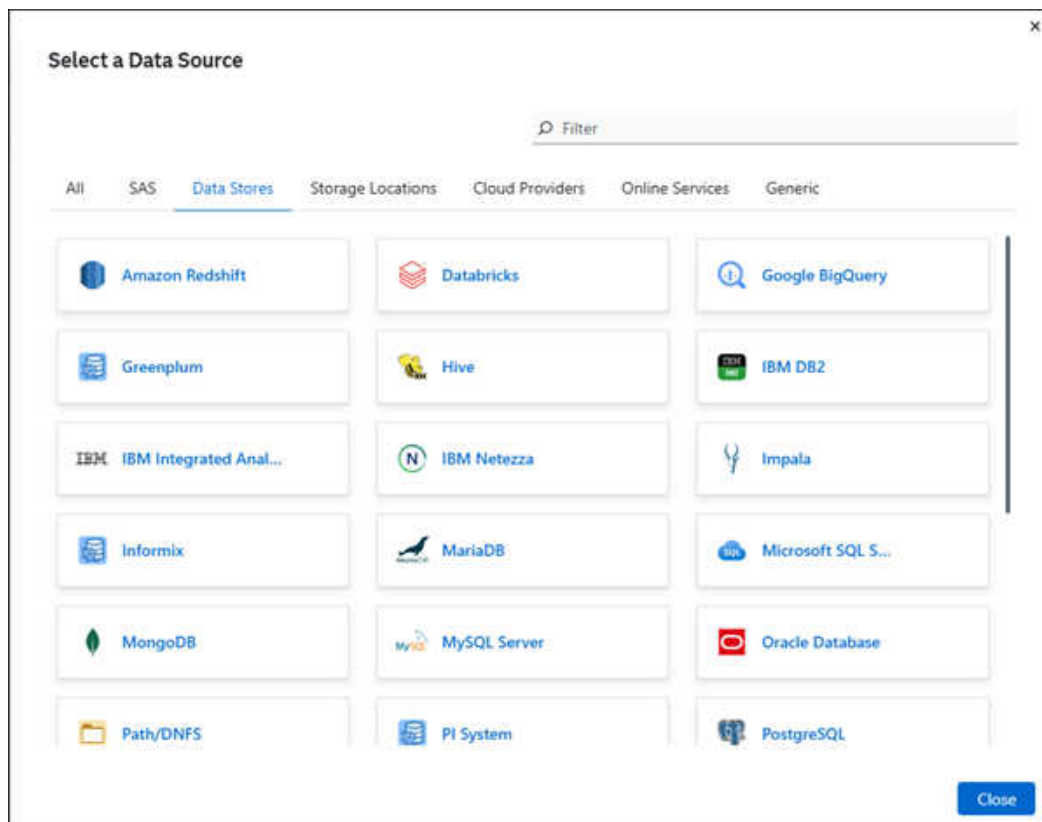
When you click **Add** on the Data Explorer page or **Create** on the Manage Connection page, the Select a Data Source window is displayed.

Figure 9 Select a Data Source Window



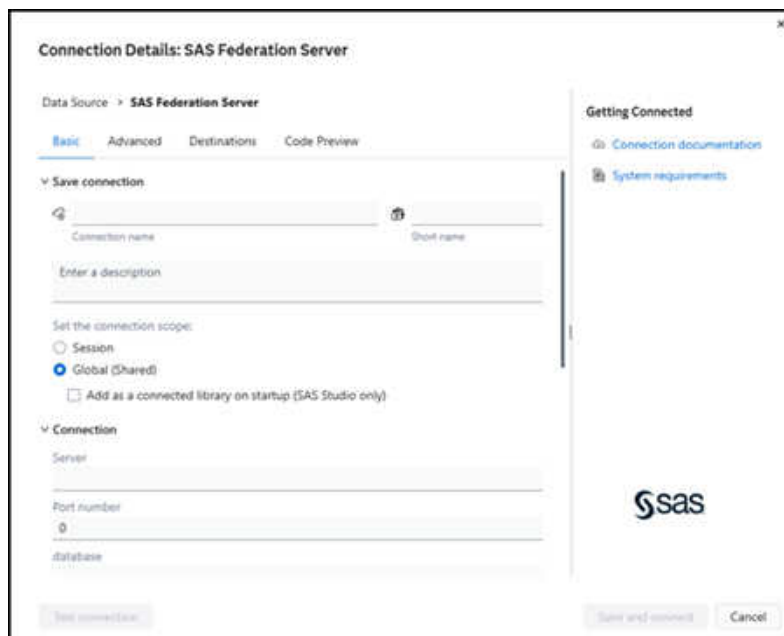
The window contains buttons that you can use to directly access the details windows for commonly used connection types. It also contains tabs across the top to access the window's specific types of connections.

Figure 10 Data Stores Tab



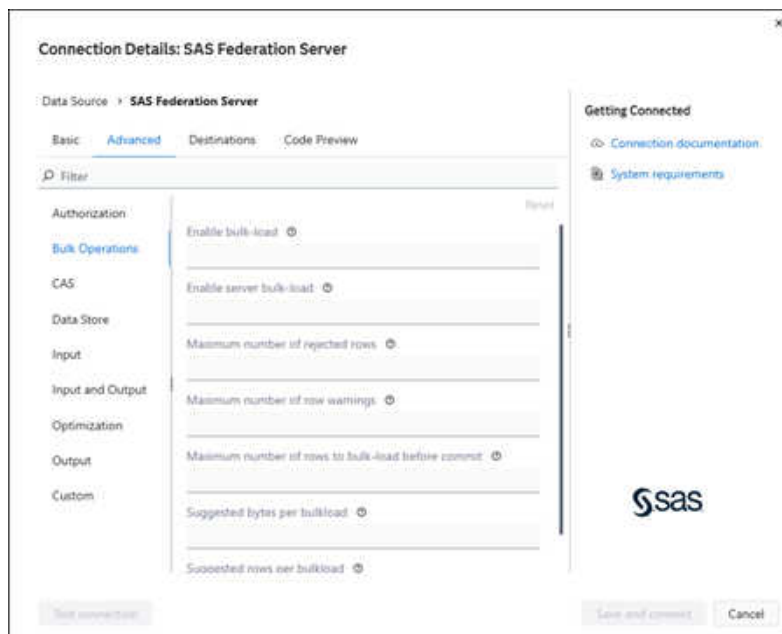
For example, this tab contains a set of buttons to access data store-based connections, such as Greenplum and Impala.

Figure 11 Connection Details: Basic Tab



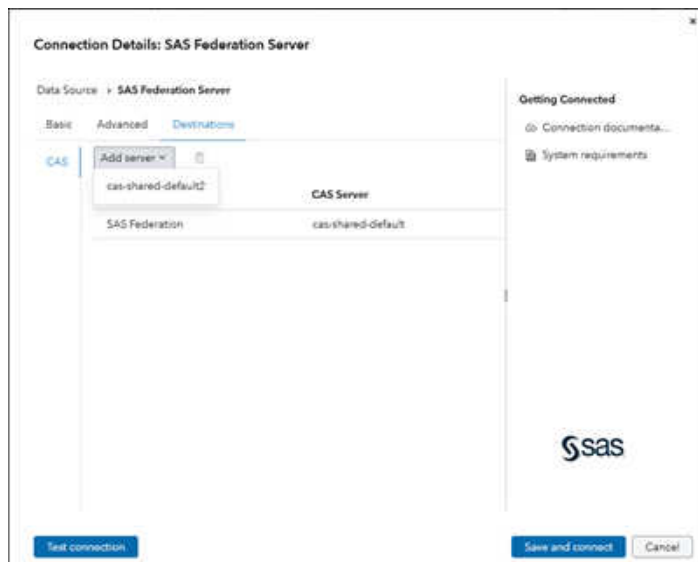
The **Basic** tab enables you to specify the basic settings for the selected connection type.

Figure 12 Connection Details: Advanced Tab



The **Advanced** tab enables you to specify advanced settings that are need for the selected connection type. The links on the left of the window display different categories of settings. The Getting Connected section of the **Basic** and **Advanced** tabs provides access to connection documentation and system requirements. You can use this information when you enter the settings for the connection.

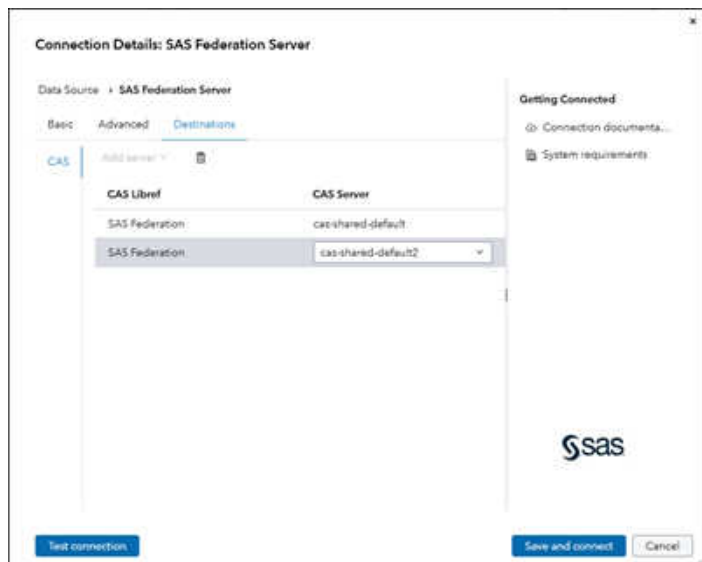
Figure 13 Connection Details: Destinations Tab



The **Destinations** tab enables you to apply the settings from the **Basic** and **Advanced** tabs to one or more additional servers. This ability enables you to add, edit, and delete connections on multiple servers.

You must set the connection scope to **Global (shared)** on the **Basic** tab to access the **Destinations** tab. To select servers from the list of available servers, click **Add server** and click the items that you need in the drop-down menu.

Figure 14 Server Selected on the Destinations Tab

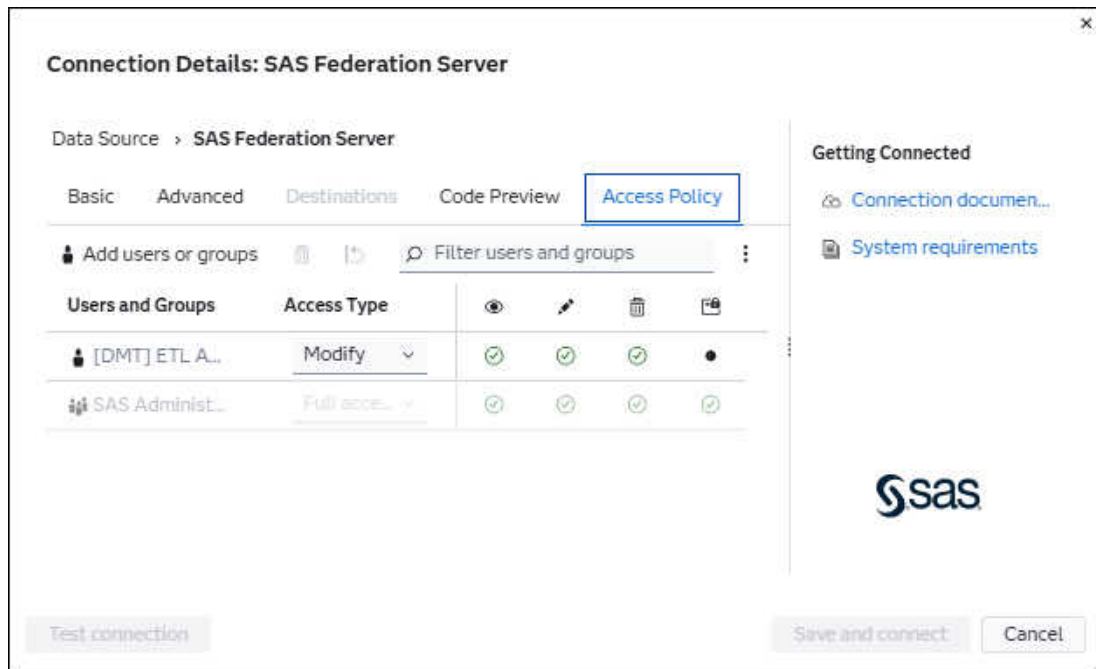


The caslib for the cas-shared-default2 has been selected in the display above. When multiple servers are added to the **Destinations** tab, the following statements are true:

- Connections that are created on one server are created and established on all the selected servers.
- When the same caslib is deleted or edited from one server, it is also deleted or edited from the other servers.
- A connection can be created on multiple servers. However, you can remove one of the servers by editing the connection and saving the changes. In this case, the caslib for the removed server is deleted.

Click  to remove a selected server from the list of destinations.

Figure 15 Connection Details: Access Policy Tab



The **Access Policy** tab enables you to manage the access to a connection for users and groups.

For more information, see [“Managing the Access Policy for a Connection”](#) on page 81.

Connecting to Databases

Overview of Database Connections

The **Sources** tab enables you to add a new caslib that connects to a database server. If the connection is successful, tables that you are authorized to access in the database are available on the **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.
- If you are administrative user, you can specify the scope of the connection. The Connection Details window includes two radio buttons for this purpose:
 - ☐ **Session:** Select this check box to add a session-based caslib for this connection.
 - ☐ **Global (Shared):** Select this check box to add a global caslib for this connection.

If you are not an administrative user, these radio buttons are not available and a session connection is created.

For more information about connection scope, see [“Caslibs on the Sources Tab and Import Tab”](#) on page 84.

- Default caslibs are provided in the **Location (library)** fields on the **Import** tab when you import data assets. The default caslib is removed by resetting the **Location (library)** field to a blank field when you
 - ☐ change the caslib from global scope to session scope
 - ☐ change the name of the caslib
 - ☐ delete the default caslib and create a new one with the same details

If you change properties such as the path or description of the caslib, it remains the default caslib.

- Identify the data source that you want to access.

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

You can add a connection for a database other than the databases listed above. See [“Data Source Requirements” in System Requirements for the SAS Viya Platform](#).

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

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

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

Create a Database Connection

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

- 1 Click  to open the Data Explorer page.
- 2 Click the **Sources** tab.
- 3 Click **Add** on the **Sources** tab. The Select a Data Source window is displayed.
- 4 Navigate to the database type that you need in Select a Data Source window. Click the button named for that database type to open the **Basic** tab in its Connection Details window.
- 5 Use the Save connection section of the window to enter the name, short name, and descriptive information about the connection that you want to create.
- 6 Enter a name for the caslib in the **Connection name** field in the Save connection section of the window. Note that you must also enter a value in the **Short name** field.

Follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 86](#). The target operating system and data source might have additional constraints on these names. If desired, you can enter a short name and description of the caslib for your reference.

- 7 If you are administrative user, you can specify the scope of the connection. The Connection Details window includes two radio buttons for this purpose:

- **Session:** Select this check box to add a session-based caslib for this connection.
- **Global (Shared):** Select this check box to add a global caslib for this connection.

If you are not an administrative user, these radio buttons are not available and a session connection is created.

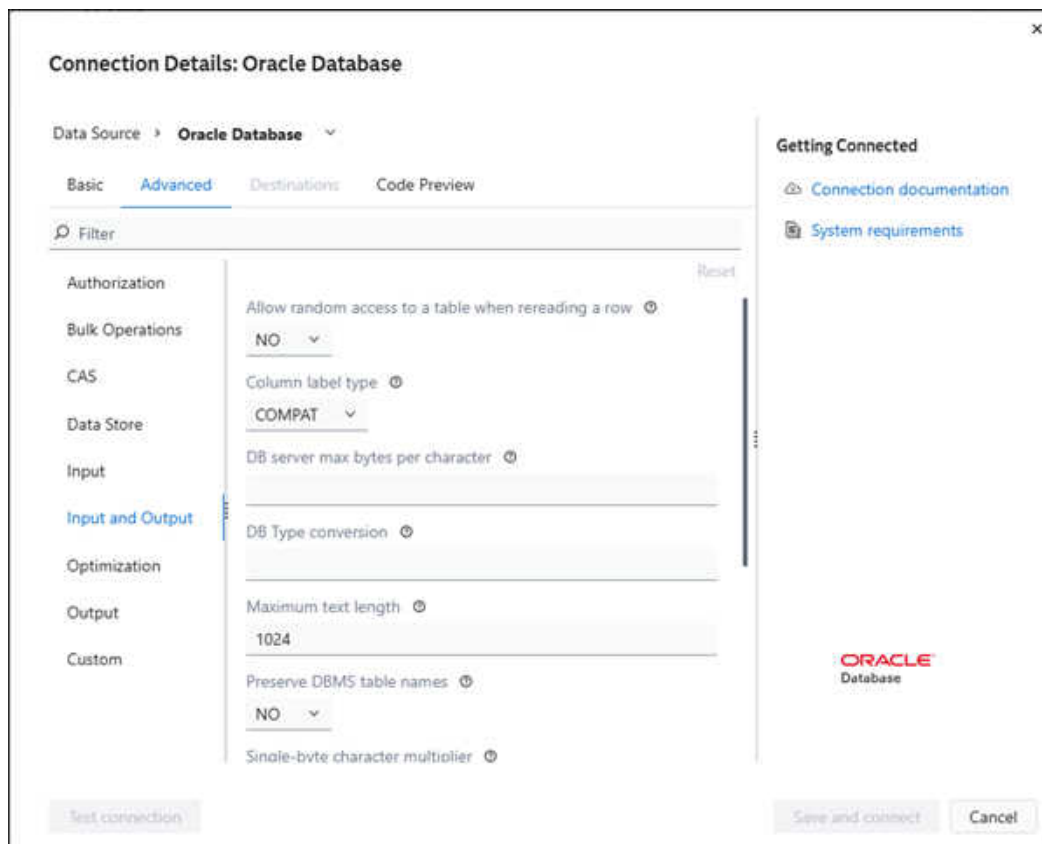
For more information about connection scope, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).

- 8 Use the Connection section of the window to enter information about the connection. The **Connection documentation** link in the right pane takes you directly to detailed information about the selected connection.

The fields in this section vary depending on the data type. For example, the Connection section for Oracle databases contains fields for **Connection options**, **Path**, and **Schema name**.

- 9 Use the Authentication section to specify a user name and password for the connection. Click **Authentication domain** to access the **Authentication Domain** field.
- 10 Click the **Advanced** tab to specify additional options for the current data source. Note that the advanced options are divided into groups that you can access from the links at the left of the window. Click ⓘ beside each option for more information about the option.

Figure 16 Advanced Options



- 11 Click the **Access Policy** tab to assign access to this connection.

To add users or groups, click **Users and Groups** and then select the users or groups that you want to add.

For each user and group, click **Access Type** to select the type of access that is granted. You can override individual permissions by clicking the icons in the table.

12 Click **Test connection** to test your connection.

13 When ready, click **Save and connect** to save your connection.

If the connection succeeds, tables that you are authorized to access in the database are available from the caslib that you specified in Step 6. Information about caslibs and tables on the **In-memory data (available)** list and the **Sources** tab 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 84](#).

14 If the target caslib is not visible from your current view, open the specified CAS server to find it.

Usage Notes for Database Connections

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 1024, which is a large value and could cause memory allocation problems and decreased performance. In most cases, you should specify a value that is more accurate for the actual length of the string rather than using the default.

Managing Table Names for SAS LASR Analytic Server

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

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

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

Connecting to Remote File Systems

Overview of Remote File System Connections

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

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

You can also add a connection for a remote file system other than the systems listed here. See [“User-Defined Connections for Data Sources” on page 30](#).

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

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

- Identify a name for the new caslib and the CAS server to be used in the connection.
- Identify the type of remote file system that you want to access: a DNFS or a server-side directory (PATH).
- The **Save connection** section of the Connection Details window enables you to set a connection scope. Select **Global Shared** to add a global caslib for this connection. Select **Session** to add a session-based caslib for this connection. For more information about this option, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).
- Ask the file system administrator for the physical path and other information that is necessary to access an appropriate directory on that system. If the new caslib is a session-based caslib, you must specify an existing directory, or the connection fails. If the new caslib is global in scope, you can specify a directory that does not exist on the target file system and create that directory. 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 Details** window can contain an **Subdirectories** check box in the **CAS** section of the **Advanced** tab. Select this check box only if you are creating a path-based caslib for a feature that requires directory support. Features that require a path-based caslib with directory support include tables in Parquet format, tables in ORC format, and the **Directory Import** data sources on the **Import** tab. In most cases, do not select this check box. Each **File System** connection to a directory usually provides access to data in that directory only, not to data in any subdirectories. For example, the **Sources** tab shows tables and files in a directory but not subdirectories in a directory.
- The **Connection Details** window includes an **Advanced** tab where you can specify advanced connection settings. See the documentation for any special connection options that you want to use.

[DNFS Data Source](#)

[Path-Based Data](#)

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

Connect to a Remote File System

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

- 1 Click  to open the Data Explorer page.
- 2 Click the **Sources** tab.
- 3 Click **Add** on the **Sources** tab. The Select a Data Source window is displayed.
- 4 Click **Storage Location**. Then click **Path/DNFS**.
- 5 Use the Save connection section of the window to enter the name, short name, and descriptive information about the connection that you want to create.

- 6 Enter a name for the caslib in the **Connection name** field in the Save connection section of the window.

Follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 86](#). The target operating system and data source might have additional constraints on these names. If desired, you can enter a short name and description of the caslib for your reference.

- 7 If you are administrative user, you can specify the scope of the connection. The Connection Details window includes two radio buttons for this purpose:

- **Session:** Select this check box to add a session-based caslib for this connection.
- **Global (Shared):** Select this check box to add a global caslib for this connection.

If you are not an administrative user, these radio buttons are not available and a session connection is created.

For more information about connection scope, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).

- 8 Use the Location and library types section to specify the location of the source the type of library.
- 9 Enter the physical path to the remote file system in the **Location** field. If the connection needs to use a distributed server, click **Enable DNFS**.

If the new caslib is global in scope, you can specify a directory that does not exist on the target file system, and that directory are created. If the new caslib is a session-based caslib, you must specify an existing directory, or the connection fails. For more information about global and session-based caslibs, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).

- 10 The **Short Name** field for SAS is populated from the **Short name** field in the Save connection section. Enter a short name for the Parquet directory.
- 11 Click the **Advanced** tab to specify additional options for the current data source. Click ⓘ beside each option for details about the option.

- 12 Click the **Access Policy** tab to assign access to this connection.

To add users or groups, click **Users and Groups** and then select the users or groups that you want to add.

For each user and group, click **Access Type** to select the type of access that is granted. You can override individual permissions by clicking the icons in the table.

For more information, see [“Managing the Access Policy for a Connection” on page 81](#).

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

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

If the connection succeeds, tables that you are authorized to access in the remote directory are available from the caslib that you specified in Step 6. Information about caslibs and tables on the **In-memory data (available)** list or the **Sources** tab 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 84](#).

- 15 If the target caslib is not visible from your current view, open the specified CAS server to find it .

- 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 into memory**, **Delete**, and **Add to import**. For information about these options, see [“Working with Tables” on page 56](#).

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 into memory** option, which is used to load a table or file to memory on a CAS server.

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 into memory** option. For information about the **Add to import** option, see [“Copying Data from the In-Memory Data Item or the Sources Tab” on page 59](#).

User-Defined Connections for Data Sources

Overview of User-Defined Connections

When you add a connection from the Select a Data Source window, you typically navigate to the connection type that you need, such as a specific database or a remote file system. Sometimes, though, you need to connect to a database or file system that is not listed in the window. In these situations, you can click the **Generic** tab and click the **Custom** tile to access the Connection Details: Custom window.

Use the following topics to help you plan the connection settings that you want to specify.

General Steps for Adding a User-Defined Caslib Connection

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

- 1 Click  to open the Data Explorer page.
- 2 Click the **Sources** tab.
- 3 Click **Add** on the **Sources** tab. The Select a Data Source window is displayed.
- 4 Click **Generic**. Then click **Custom**.
- 5 Use the Save connection section of the window to enter the name, short name, and descriptive information about the connection that you want to create.
- 6 Enter a name for the caslib in the **Connection name** field in the Save connection section of the window. Note that you must also enter a value in the **Short name** field.

Follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 86](#). The target operating system and data source might have additional constraints on these names. If desired, you can enter a short name and description of the caslib for your reference.

- 7 If you are administrative user, you can specify the scope of the connection. The Connection Details window includes two radio buttons for this purpose:

- **Session**: Select this check box to add a session-based caslib for this connection.
- **Global (Shared)**: Select this check box to add a global caslib for this connection.

If you are not an administrative user, these radio buttons are not available and a session connection is created.

For more information about connection scope, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).

- 8 Use the Connection section of the window to specify the data connector source type and the engine name for the connection.
- 9 Click the **Advanced** tab to specify connection options for the current data source. This tab is populated with a set of fields that you use to enter name and value pairs to specify **Additional Parameters** for this connection. For example, you can specify **Data transfer mode**, **Number of read nodes**, and **Number of write nodes** according to the information that you previously gathered.

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

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 **Type** field, enter the data type of the parameter. In the **Value** field, enter the value that you want to specify for that option. For example, you could specify the option *server*, the type *string*, and a value of **VTserver**. You can also click **Connection documentation** on the right side of the window for details about options for user-defined data

sources. If you have not specified an Authentication Domain above, you can specify a server name, user credentials, and other attributes as name/value pairs in this area.

- 10 After you have entered a parameter, click **Add Parameter** to add it to the caslib.

- 11 Click the **Access Policy** tab to assign access to this connection.

To add users or groups, click **Users and Groups** and then select the users or groups that you want to add.

For each user and group, click **Access Type** to select the type of access that is granted. You can override individual permissions by clicking the icons in the table.

For more information, see [“Managing the Access Policy for a Connection” on page 81](#).

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

If the connection succeeds, tables that you are authorized to access in the remote directory are available from the caslib that you specified in Step 6. Information about caslibs and tables on the **In-memory data (available)** list and the **Sources** tab 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 84](#).

- 13 If the target caslib is not visible from your current view, open the specified CAS server to find it .

- 14 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 into memory**, **Delete**, and **Add to import**. For information about these options, see [“Working with Tables” on page 56](#).

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.

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 into memory** option. For information about the **Add to import** option, see [“Copying Data from the In-Memory Data Item or the Sources Tab” on page 59](#).

Example: User-Defined Connection to a Database

Ask the database administrator for the information required to connect to the target database. Refer to the SAS data connector syntax for the target database for connection options. See [“Quick Reference for Data Connector Syntax” in SAS Cloud Analytic Services: User’s Guide](#).

You will need user credentials, for example. You can specify a user name and password as name and value pairs in the **Additional Parameters** section of the **Advanced** tab, as shown below.

Alternatively, a SAS administrator can create an Authentication Domain for a connection, and you can specify `authenticationdomain` and the domain name as a name and value pair under **Additional Parameters** on the **Advanced** tab. For more information about SAS Authentication Domains, see [“Authentication Domain” in SAS Viya Administration: External Credentials](#).

Follow the steps described in [“General Steps for Adding a User-Defined Caslib Connection” on page 31](#). Here is an example of a user-defined connection to a database.

■ Save connection section

- ☐ **Connection name:** Vertica Connection
- ☐ **Short name:** Vertica
- ☐ **Description:** Access Vertica tables
- ☐ **Set the connection scope:**Global (Shared)

■ Connection section

- ☐ **Data connector srcType:** vertica
- ☐ **Engine name:** cas_server02

■ Advanced Tab

If your SAS administrator has provided an Authentication Domain for this data source, you can specify the domain name instead of specifying your credentials under **Additional Parameters**.

Table 3 Parameters with an Authentication Domain

Name	Type	Value
Data transfer mode	String	Default
Number of read nodes	Integer	3
Number of write nodes	Integer	3

If you have not specified an Authentication Domain above, you can specify a server name, user credentials, and other attributes as name and value pairs in this field. Here are some parameters:

Table 4 Parameters without an Authentication Domain

Name	Type	Value
server-name	String	VTserver
user-name	String	user1
password	String	myPwd
schema-name	String	mySchema

Example: User-Defined Connection to a Remote File System

Ask the file system administrator for the information required to connect to the target directory on the remote file system. Refer to SAS data connector syntax or CASLIB statement syntax for the target file system for connection options. See [“Quick Reference for Data Connector Syntax” in SAS Cloud Analytic Services: User’s Guide](#) or [Required Arguments for the CASLIB Statement](#).

You will need user credentials, for example. You can specify a user name and password as name and value pairs in the **Additional Parameters** section of the **Advanced** tab, as shown below.

Alternatively, a SAS administrator can create an Authentication Domain for a connection, and you can specify `authenticationdomain` and the domain name as a name and value pair under **Additional Parameters** on the **Advanced** tab. For more information about SAS Authentication Domains, see [“Authentication Domain” in SAS Viya Administration: External Credentials](#).

Follow the steps described in [“General Steps for Adding a User-Defined Caslib Connection”](#) on page 31. Here is an example of a user-defined connection to a remote file system.

■ Save connection section

- ☐ **Connection name:** Spark_Connection
- ☐ **Short name:** Spark
- ☐ **Description:** Access Hadoop Hive tables with Spark
- ☐ **Set the connection scope:** Global (Shared)

■ Connection section

- ☐ **Data connector srcType:** spark
- ☐ **Engine name:** cas_server02

■ Advanced Tab

If your SAS administrator has provided an Authentication Domain for this data source, you can specify the domain name instead of specifying your credentials under **Additional Parameters**.

Table 5 Parameters with an Authentication Domain

Name	Type	Value
Data transfer mode	String	Default
Number of read nodes	Integer	3
Number of write nodes	Integer	3

If you have not specified an Authentication Domain above, you can specify a server name, user credentials, and other attributes as name and value pairs in this field. Here are some sample parameters:

Table 6 Parameters without an Authentication Domain

Name	Type	Value
hadoopjarpath	String	/hadoop/jars:hadoop/ spark/jars
hadoopConfigDir	String	/hadoop/conf
server	String	thriftserver

Name	Type	Value
schema	String	mySchema
username	String	user1
password	String	myPwd
dbmaxText	String? or Integer?	2048

Example: User-Defined Connection to an ESP Project

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

Ask the ESP system administrator for the information that is required to connect to the desired server and project. For more information about connection options for an ESP server, "[SAS Event Stream Processing Data Source](#)." Follow the steps described in "[General Steps for Adding a User-Defined Caslib Connection](#)" on [page 31](#). Here is an example of a user-defined connection to an ESP project.

■ Save connection section

- ☐ **Connection name:** ESP_Project_Connection
- ☐ **Short name:** ESP
- ☐ **Description:** Access ESP Project
- ☐ **Set the connection scope:** Global (Shared)

■ Connection section

- ☐ **Data connector srcType:** ESP
- ☐ **Engine name:** cas_server02

■ Advanced Tab

Table 7 Parameters

Name	Type	Value
Data transfer mode	String	Default
Additional Parameters Specify the port and name of the ESP server.		
Port	Integer	1111
Server	String	esp_server_name

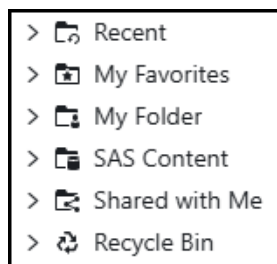
After you are connected to a project on an ESP server, see [“Import a Snapshot of an ESP Project” on page 67](#).

Importing Data from SAS Folders

Overview of SAS Folders

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

Figure 17 SAS Folders



SAS Viya platform applications save their content to SAS folders. You can select **New** ⇒ **Upload** ⇒ **Files to My Folder** to upload external files to a SAS folder.

Import Files from SAS Folders

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

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

- 1 Click  to open the Data Explorer page.
- 2 Click the **Import** tab.
- 3 Click **Folder Files** in the import queue.

A selection window for SAS folders is displayed.

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

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

Figure 18 Import a File from a SAS Folder

- 5 In the **Step 1 - File Import Options** section, 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 a worksheet to import

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

First row contains column names

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

Limit range of imported rows and columns

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

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

Delimited text file options:

The following options are available for delimited text files:

Input file delimiter

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

Scanned rows

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

Locale

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

For examples of locale code values, see [Available languages for Windows](#) or the [Windows Language Code Identifier \(LCID\) Reference](#).

Encoding

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

First row contains column names

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

Convert character columns to variable size

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

CAUTION

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

SAS data set, SAS7BDAT options:

The following options are available for SAS7BDAT files:

Password

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

Encryption key

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

Character multiplier

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

- 6 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.

- 7 In the **Step 2 - Output File Options** section, the name of the source was copied into the **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 86](#). The target operating system and data source might have additional constraints on these names. For example, the target table name can be a maximum of 247 characters long. However, a shorter file name might be required due to the data source settings for the caslib that is specified in the **Location (library)** 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 46](#).

- 8 A default caslib is provided in the **Location (library)** field. To override the default, select an existing global caslib or add a new one.

If you have the privilege that is required to add a new global caslib, follow the name conventions described in [“Names for Caslibs, Tables, and Columns” on page 86](#). For more information about global caslibs, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).

- 9 Specify what action the import operation should take if the target file name exists in the caslib that was specified in the **Location (library)** field. The options are to cancel the import or to replace the existing file that has the same name.
- 10 Specify a label that helps identify the copied table, if desired. The caslib for the target table must support labels. Caslibs of type PATH and DNFS (SASHDAT only) support labels, for example. Caslibs for DBMS and HDFS do not support labels.
- 11 Use the **Format** selector to select the desired format for the target table. Only SASHDAT and CSV formats are supported.
- 12 Select **Save as in-memory data only** if you need to save the output in memory without creating a physical file.
- 13 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 **In-memory data (available)** list or the **Sources** tab.

If the import fails, see [“Usage Notes for Importing Local Files” on page 45](#). See also [“General Usage Notes” on page 84](#).

[“Usage Notes for Importing Local Files” on page 45](#) also apply to files imported from SAS folders.

Importing Local Files

Overview of Importing Local Files

Local files are local to the browser where your SAS Viya platform 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 Import section of the **Import** tab supports two methods for copying a local file and loading it to memory on a CAS server:

- Use the **Multi-Sheet Excel File** option to import Excel files with multiple worksheets. Each worksheet is loaded to memory as a separate table.
- Use the **Local Files** option to import other local files. You can click **Local Files** to use a file selection window or drag a local file to the **Drag local file here** drag and drop target area.

Use the Local File Option

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

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

- 1 Click  to open the Data Explorer page.
- 2 Click the **Import** tab.
- 3 Drag and drop a local file onto the **Import** tab. Alternatively, click **Local files** 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, as shown in the next figure,

Figure 19 Import a Local File

The screenshot shows the 'Feedback.xlsx' import dialog. On the left, a sidebar lists 'Sources' and 'Import (1)' with 'Feedback.xlsx' as a local file. The main area is titled 'Feedback.xlsx' and contains two sections: 'Step 1 - File Import Options' and 'Step 2 - Output File Options'. In 'Step 1', the 'File Specifications' tab is selected, showing three options: 'Specify a worksheet to import' (unchecked), 'First row contains column names' (checked), and 'Limit the range of imported rows and columns' (unchecked). In 'Step 2', the 'Name' field is 'Feedback', 'Location (library)' is 'cas-shared-default/Public', 'Label' is 'Enter label', and 'Format' is 'sashdat'. There are also radio buttons for 'Cancel import' (selected), 'Replace file', and a checkbox for 'Save as in-memory data only'.

- 4 In the **Step 1 - File Import Options** section, 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 a worksheet to import

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

First row contains column names

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

Limit range of imported rows and columns

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

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

Delimited text file options:

The following options are available for delimited text files:

Input file delimiter

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

Scanned rows

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

Locale

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

For examples of locale code values, see [Available languages for Windows](#) or the [Windows Language Code Identifier \(LCID\) Reference](#).

Source encoding

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

First row contains column names

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

Convert character columns to variable size

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

CAUTION

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

SAS data set, SAS7BDAT options:

The following options are available for SAS7BDAT files:

Password

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

Encryption key

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

Character multiplier

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

- 5 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.

- 6 In the **Step 2 - Output File Options** section, the name of the source was copied into the **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 86](#). 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 **Location (library)** 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 46](#).

- 7 A default caslib is provided in the **Location (library)** 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 86](#). For more information about global caslibs, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).

- 8 Specify what action the import operation should take if the target file name exists in the caslib that was specified in the **Location (library)** field. The options are to cancel the import or to replace the existing file that has the same name.

- 9 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 and DNFS (SASHDAT only) support labels, for example. Caslibs for DBMS and HDFS do not support labels.

- 10 Use the **Format** selector to select the desired format for the target table. Only sashdat and CSV formats are supported.

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

If the import succeeds, a copy of the file is loaded to memory on the CAS server that is specified in the caslib. The copy of the file can be selected from the **In-memory data (available)** list or the **Sources** tab.

If the import fails, see [“Usage Notes for Importing Local Files” on page 45](#). See also [“General Usage Notes” on page 84](#).

Use the Excel Spreadsheet (Multiple Worksheets) Option

Use the **Multi-Sheet Excel File** option in the Local Files section of the **Import** tab to import Excel files with multiple worksheets. Each worksheet is loaded to memory as a separate table.

- 1 Click ⓘ to open the Data Explorer page.

- 2 Click the **Import** tab.
- 3 Click the **Multi-Sheet Excel File** option in the Local Files section of the **Import** tab to display a file selection window. Select the Excel file that contains the worksheets that you need.

The **Add to Import** window is displayed with a list of the worksheets that are available in the file. You are prompted to specify which worksheets to import. You can also change some default import settings.

- 4 When you are ready to import the worksheets, click **Add**.

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

Figure 20 Import an Excel File with Multiple Worksheets

The screenshot displays the SAS Import Wizard interface. On the left, a 'Queue (3)' panel lists three import jobs: 'Feedback (2)_2018', 'Feedback (2)_2017', and 'Feedback (2)_2016', all located in '/My Folder'. The 'Feedback (2)_2018' job is selected. The main panel on the right shows the configuration for this job, titled 'Feedback (2)_2018'. It includes two tabs: 'File Specifications' and 'Advanced'. The 'File Specifications' tab is active, showing options for 'Specify a worksheet to import' (checked) with 'Worksheet name:' set to '2018', and 'First row contains column names' (checked). There is also an unchecked option for 'Limit the range of imported rows and columns'. Below this is the 'Step 2 - Output File Options' section, which includes fields for 'Name:' (set to 'Feedback (2)_2018') and 'Location (library):' (set to 'cas-shared-default/Public'). There are also fields for 'Label:' (placeholder 'Enter label') and 'Format:' (set to 'sas7bdt'). At the bottom, there are radio buttons for 'Cancel import' (selected), 'Replace file', and a checkbox for 'Save as in-memory data only'.

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

- 5 In the **Step 1 - File Import Options** section, use the **File Specifications** tab to specify any options that are needed to accurately copy the contents of the imported file.
- 6 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.
- 7 In the **Step 2 - Output File Options** section, the name of the source was copied to the **Name** field on the right. Accept or change this name for the copy as appropriate.

If you change the name, follow the SAS conventions for table names as described in [“Names for Caslibs, Tables, and Columns” on page 86](#). The target operating system and the data source might have additional constraints on these names. For example, the target table name can be a maximum of 247 characters long. However, a shorter name might be required due to the data source settings for the caslib that is specified in the **Location (library)** 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 46](#).

- 8 A default caslib is provided in the **Location (library)** field. To override the default, select an existing global caslib or add a new one.

If you have the privilege that is required to add a new global caslib, follow the name conventions described in [“Names for Caslibs, Tables, and Columns” on page 86](#). For more information about global caslibs, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).
- 9 Specify what action the import operation should take if the target file name exists in the caslib that was specified in the **Location (library)** field. The options are to cancel the import or to replace the existing file that has the same name.
- 10 Specify a label that helps identify the copied table, if desired. The caslib for the target table must support labels. Caslibs of type PATH and DNFS (SASHDAT only) support labels, for example. Caslibs for DBMS and HDFS do not support labels.
- 11 Use the **Format** selector to select the desired format for the target table. Only SASHDAT and CSV formats are supported.
- 12 When ready, you can right-click one worksheet to be copied and select **Import Item**.

If the import succeeds, a copy of the worksheet is loaded to memory on the CAS server that is specified in the caslib. The copy of the worksheet can be selected from the **In-memory data (available)** list or the **Sources** tab.

If the import fails, see [“Usage Notes for Importing Local Files” on page 45](#). See also [“General Usage Notes” on page 84](#).

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

Usage Notes for Importing Local Files

All Local Data Files

You might experience long wait-times when you import large local files through a web browser. Accordingly, the CAS Management service option `maxFileUploadSize` is set to 25,000 megabytes by default. If the import of a local file fails because the file size is too large, it might be because you have exceeded the `maxFileUploadSize` value for the CAS server. Your administrator can change the value of this option as appropriate for your site.

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

Importing an Excel File with Multiple Worksheets

You might experience long wait times when using the **Multi-Sheet Excel File** option in the Local File Import section of the **Import** tab to import large local files through a web browser. Accordingly, the File services option `maxFileSizeMB` is set to 100 megabytes by default. If the import of a local file fails because the file size is too large, it might be because you have exceeded the `maxFileSize` value for the CAS server.

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

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

Importing Text Files

Here are some key points about importing text files:

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

Importing SAS Data Sets

Here are some key points about importing SAS data sets:

- Importing indexed SAS data sets is not supported.
- Before importing a SAS data set with user-defined formats, ensure that the custom format catalog is available to the CAS server. For more information, see [“Managing User-Defined Formats in SAS Viya” in SAS Viya Administration: Data](#).
- If your character data contains international characters, and the data is truncated on import, you might want to specify a higher value in the **Character multiplier** field. The default value of this field is 1. You can specify a multiplier up to 5 to increase the number of characters that can fit in each cell.

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

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

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

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

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

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

Importing Esri Data for Geoenrichment and Geocoding

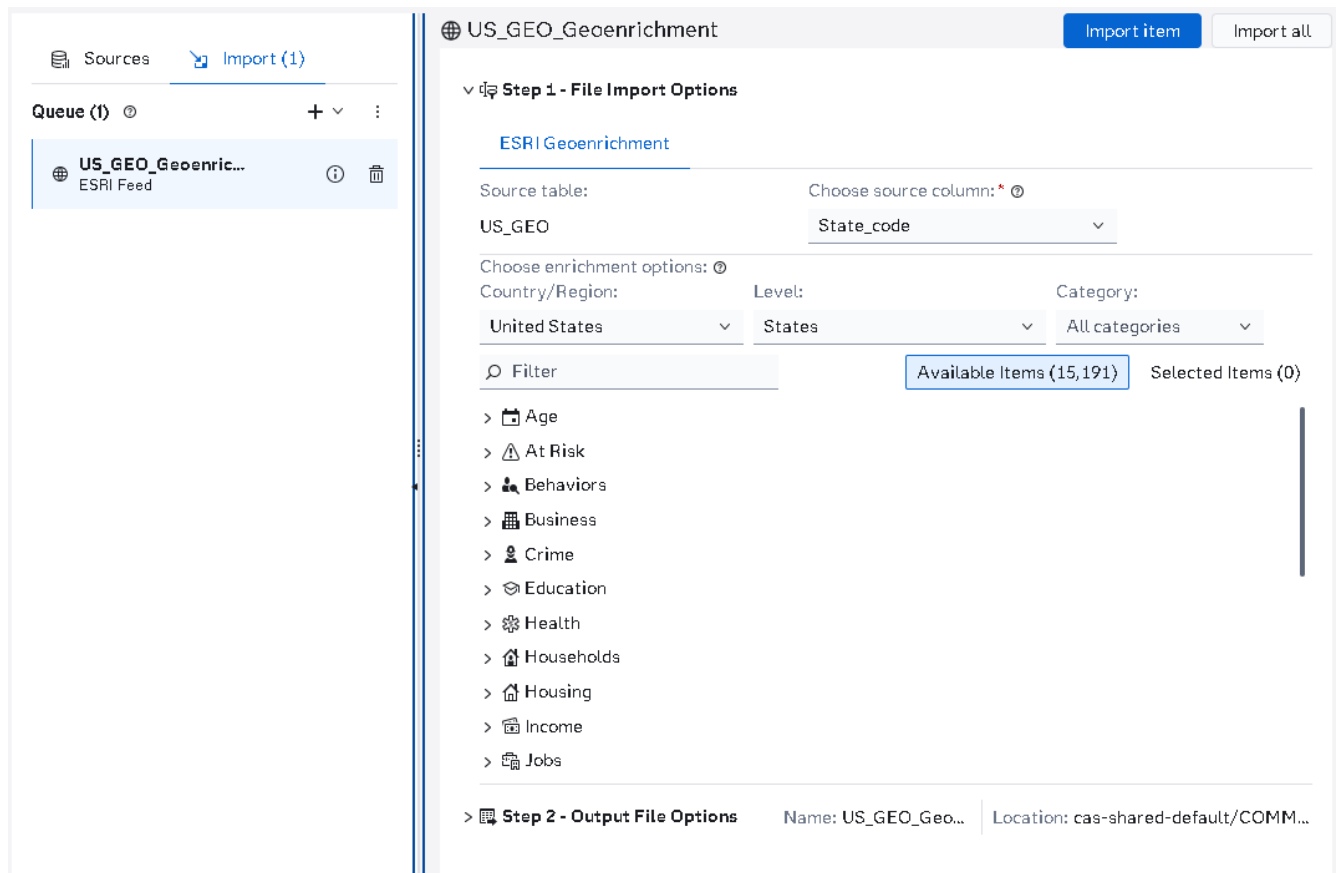
Geoenrichment

Overview of ESRI GeoEnrichment

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

If your login has appropriate privilege, you can select Geoenrichment in the Esri section 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 geoenrichment import item is added to the queue on the left, as shown in the next figure.

Figure 21 Geoenrichment Import Item in the Queue on the Import Tab



Import properties for the Geoenrichment item appear on the right. You specify various options to combine Esri enrichment data with data from the selected source table. Output from the merge is written to the table specified in the **Name** field and loaded to memory. The target table is in the caslib that is specified the **Location (library)** field.

You can use the options under **Step 1 - File Import Options** 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 Feed** 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 **Location (library)** field.

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

Note: Any items in the queue on the **Import** tab are dropped from the queue when you log off. The target tables that you load to memory persist beyond the current session. The tables are associated with a global caslib in the import properties window. For more information about global caslibs, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).

Prerequisites for Geoenrichment

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

- 1 Ask an administrator to add your login to the custom group **Esri Users** in SAS Environment Manager. For information about custom groups, administrators can refer to [“Users and Groups Page” in SAS Environment Manager: User's Guide](#).
- 2 Obtain the user name and password for the Esri ArcGIS Online 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 Hub** section.
- 5 Click **Geographic Mapping**.
If you have been added to the **Esri Users** group in SAS Environment Manager, the **Accept Esri ArcGIS Online Services terms and conditions** check box is displayed.
- 6 Select the **Accept Esri ArcGIS Online Services terms and conditions** check box.
- 7 Click **Sign in to Esri ArcGIS Online account** to access the Esri ArcGIS Online Account Sign In window.
- 8 Specify the user name and password for the Esri ArcGIS Online Service. Click **Sign In** to exit the window.
- 9 Click **Close** to exit the Settings window and return to the selected geoenrichment item.
- 10 Identify the caslib and name of the source table to which you will add Esri data. The source table for a geoenrichment import item must be an in-memory table.
- 11 Identify the caslib and name of the target table that will combine Esri data with data from the selected source table.
- 12 Identify the geographic level that you want to analyze such as by state or ZIP code.
- 13 Identify the Esri enrichment data that will support the analysis that you want to perform on the output table. For example, you might want to analyze demographic data, business data, or landscape data by geographic location. For more information about Esri enrichment data, see [The Esri GeoEnrichment Service](#).

Import Esri Data for Geoenrichment

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

- 1 Select **Geoenrichment** from the Esri section in the **Import** tab.
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, the table is added as an Esri Feed item to the queue on the left. Import properties appear on the right.
- 3 In the **Step 1 - File Import Options** section, enter the geo enrichment options:

Source table

contains the in-memory table that you selected when you clicked Geoenrichment in the **Import** tab. Its data is combined with Esri enrichment data.

Choose source column

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

Country/Region

Select the country or region 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 entities such as **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 entities such as **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, you can click the **Population** category to review items such as for **Gender** and **Language**. Then you can click an item such as **Language** to see more detailed choices.

Click the items that you need to move them from the **Available items** list to the **Selected items** list. Click the **Available items** and **Selected items** buttons to move from one list to the other.

Selected items

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

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

- 4 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.
- 5 In the **Step 2 - Output File Options** section, the default **Name** is the source table name with _Geoenrichment 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 86](#). 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 **Location (library)** field. Contact your data administrator for more information about the permitted length for target table names.

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

If you have the privilege that is required to add a new global caslib, follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 86](#). For more information about global caslibs, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).

- 7 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 and DNFS (SASHDAT only) support labels, for example. Caslibs for DBMS and HDFS do not support labels.
- 8 Specify what action the import operation should take if the target filename exists in the caslib specified in the **Location (library)** field. The options are to cancel the import or to replace the existing item that has the same name.

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

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

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

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

If the import fails, see [“Usage Notes for Importing Local Files” on page 45](#).

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 you can select **Geocoding** in the Esri section 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 Feed** item is added to the queue on the left, as shown in the next figure.

Figure 22 Geocode Import Item in the Queue on the Import Tab

CUSTOMERS_US_Geocode Import item Import all

▼ **Step 1 - File Import Options**

ESRI Geocode

Source table: CUSTOMERS_US

Find All Columns (14) Selected Columns (3) ?

☐	# ↑	Name	Label	Data Type
☐	3	Person	--	varchar
☐	4	Title	--	varchar
☐	5	Firm	--	varchar
☒	6	Address1	--	varchar
☐	7	Address2	--	varchar
☐	8	Address3	--	varchar
☒	9	City	--	varchar
☒	10	Region	--	varchar
☐	11	Postcode	--	double

Test ? Sample address:
ABILENE 2 VILLAGE DRIVE TX

▼ **Step 2 - Output File Options**

Name: * Location (library): * Go to...

Label:

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

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

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

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

Note: Any items in the queue on the **Import** tab are dropped from the queue when you log off. The target tables that you load to memory persist beyond the current session. The tables are associated with a global caslib in the import properties window.

For more information about global caslibs, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).

Prerequisites for Geocoding

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

- 1 Ask an administrator to add your login to the custom group **Esri Users** in SAS Environment Manager. For information about custom groups, administrators can refer to [“Users and Groups Page” in SAS Environment Manager: User's Guide](#).
- 2 Obtain the user name and password for the Esri ArcGIS Online 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 Hub** section.

- 5 Click **Geographic Mapping**.

If you have been added to the **Esri Users** group in SAS Environment Manager, the **Accept Esri ArcGIS Online Services terms and conditions** check box is displayed.

- 6 Select the **Accept Esri ArcGIS Online Services terms and conditions** check box.
- 7 Click **Sign in to Esri ArcGIS Online account** to access the Esri ArcGIS Online Account Sign In window.
- 8 Specify the user name and password for the Esri ArcGIS Online Service. Click **Sign In** to exit the window.
- 9 Click **Close** to exit the Settings window and return to the selected geocoding item.
- 10 Identify the caslib and name of the source table to which you will add Esri data. The source table for a **Esri Feed** item must be an in-memory table.
- 11 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 Select **Geocode** from the Esri section in the **Import** tab.

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, the table is added as an Esri Feed item to the queue on the left. Import properties appear on the right.

3 In the **Step 1 - File Import Options** section, the **Esri Geocode** tab displays a table that lists the columns that are available in the source table. Select one or more columns in the source table that contain address information.

- Use the **Find** field to locate a column buried in a long list.
- Click **All Columns** or **Selected Columns** to switch between a table that displays all of the available columns and a different table that displays only the columns that you have selected.
- Select a check box next to a row to add that column to the selected columns table. Deselect a selected row to remove it from the selected columns table.
- Select the check box in the header row of a table to change the status of all the rows at the same time. When some or all of the columns are not selected, you can click the check box to select them all. When some or all of the columns are selected, you can click the check box to deselect them all.
- Use the arrow in the header row to sort the rows in ascending or descending order. Click  to display or hide the columns in a table.

4 (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.

5 In the **Step 2 - Output File Options** section, the default **Name** is the source table name with **_Geocode** 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 86](#). 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 **Location (library)** field. Contact your data administrator for more information about the permitted length for target table names.

6 A default caslib for the target table is provided in the **Location (library)** 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 86](#). For more information about global caslibs, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).

7 (Optional) Specify a label that will help identify the output table. The caslib for the target table must support labels. Caslibs of type PATH and DNFS (SASHDAT only) support labels, for example. Caslibs for DBMS and HDFS do not support labels.

8 Specify what action the import operation should take if the target filename exists in the caslib that is specified in the **Location (library)** field. The options are to cancel the import or to replace the existing item that has the same name.

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

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

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

Esri geocoding data is combined with the data in the source table and written to the output table. The output table is loaded to memory on the CAS server that is specified in the **Location (library)** field. The output table can be selected from the **In-memory data (available)** list or the **Sources** tab.

If the import fails, see [“Usage Notes for Importing Local Files”](#) on page 45.

Working with Data in CAS

Working with Caslibs and Tables

Working with Caslibs

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

Table 8 Pop-Up Menu Options for a Selected Caslib

Name	Action
Set as default target library	Makes the selected caslib the default location for various operations, including operations on the Import tab. This option is available only for global caslibs.
Import from library ⇒ Delimited files...	Imports multiple delimited files, documents, or images from a directory in the selected caslib. This option is available only for PATH caslibs. For more information about these options, see the following topics: ■ “Importing Delimited Files from a Caslib Directory” on page 68 ■ “Importing Documents from a Caslib Directory” on page 71 ■ “Importing Images from a Caslib Directory” on page 74
View authorization	If available, displays the permissions that each user or group has for the selected caslib. See “CAS Authorization Window” in <i>SAS Viya: CAS Authorization Window</i> .

Name	Action
Edit authorization	If available, enables you to change the permissions that each user or group has for the selected caslib. The ManageAccess permission is required to modify access controls. See “CAS Authorization Window” in SAS Viya: CAS Authorization Window .
Edit connection	If available, enables you to edit the details of the caslibs connection. This option is available if you have AlterCaslib permission for the selected caslib. When you edit a caslib, the loaded contents of that caslib are dropped and a new caslib is created. A warning is displayed.
Delete connection	Removes the selected caslib from memory on a CAS server.
Assume Superuser role	Enables you to assume the Superuser role. Superusers are exempt from all CAS authorization requirements and they can perform all tasks.

Working with Tables

Overview

You can select a table from two lists:

- the **In-memory data (available)** list
- the list of items displayed when you click a caslib in the **Sources** tab

You can click on the name of the table in the **Name** column to open it. You can also click anywhere in its row in one of these lists to select it in the list.

In both cases, two tabs of information about the table are displayed on the right side:

- [“Details Tab” on page 57](#)
- [“Sample Data Tab” on page 58](#)

This Table window contains a toolbar for reviewing table details and sample data:

Table 9 Table Toolbar

Item	Description
Columns	Number of columns
Rows	Number of rows
Size	Table size

Item	Description
Load into memory (⚡) button	When clicked, loads the selected item into memory.
Add to import (➤) button	When clicked, adds the selected item to the current import.
Actions	Enables you to choose an action to perform for the selected item. See “Pop-Up Menu Options for Items on the Data Explorer Page” on page 13.

You can click the **Discover and Govern Data** item in the **Actions** drop-down menu for either the **Details** or **Sample Data** tab to view the table in the Details window in SAS Data Governance. There, you can review information about the table and its columns. See [“About SAS Data Governance”](#) in *SAS Data Governance: User's Guide*.

If an analysis has been run for the table, it is displayed. If not, you can click **Analyze data** to perform an ad hoc analysis. See [“Ad Hoc Analysis”](#) in *SAS Data Governance: User's Guide*.

Details Tab

Displays basic information about the table, such as the number of columns and rows and the data type of columns.

Figure 23 Details Tab

MAILORDER

Columns: 9

Rows: 2.5 K

Size: 177.6 ...

Actions

Details

Sample data

Filter

#	↑	Name	Label	Data Type	Raw Length	Formatted Length	Format	⌵
1		CUSTNUM	--	double	8	12	--	
2		INV	--	double	8	12	--	
3		Date	--	double	8	9	DATE	
4		PCODE	--	char	6	6	\$CHAR	
5		QTY	--	double	8	12	--	

Click  to see the properties pane for the table:

Figure 24 Table Properties

Properties	
Asset type:	In-memory data
Date modified:	Apr 22, 2025 3:54 AM
Modified by:	sasboot
Date created:	Apr 22, 2025 3:54 AM
Created by:	sasboot
Date last accessed:	Apr 22, 2025 2:00 PM
Library:	COMMON_PROD
Encoding:	UTF-8
Label:	--
Source table:	mailorder.sas7bdat
Source table library:	COMMON_PROD

Sample Data Tab

The **Sample Data** tab displays sample data from the selected table. If the sample data includes images, click an image to preview it. Some image formats cannot be displayed by some browsers unless they are decompressed. Other formats might not be displayable unless the browser is configured to launch an external viewer. The overflow menu beside **Sample rows** enables you to select **Show labels** or **Show names** for columns in the table.

Figure 25 Sample Data Tab

CUSTNUM	INV	Date	PCODE	QTY	Qtr
137000	107707	09MAR2000	SP1000	1	1
137000	110633	27MAR2000	GD2200	1	1
137000	111155	30MAR2000	PT1300	3	1
137000	126867	24JUN2000	TY4100	1	2
137000	135872	10AUG2000	TY1200	4	3
137000	135872	10AUG2000	TY2100	1	3

You can specify the number of sample data rows and refresh the data. You can also click to show either names or labels and formatted or unformatted values.

Note: If a table is loaded into memory, the **Sample Data** tab always shows the data that is loaded into memory, unless you select the physical table and the in-memory table has a different name.

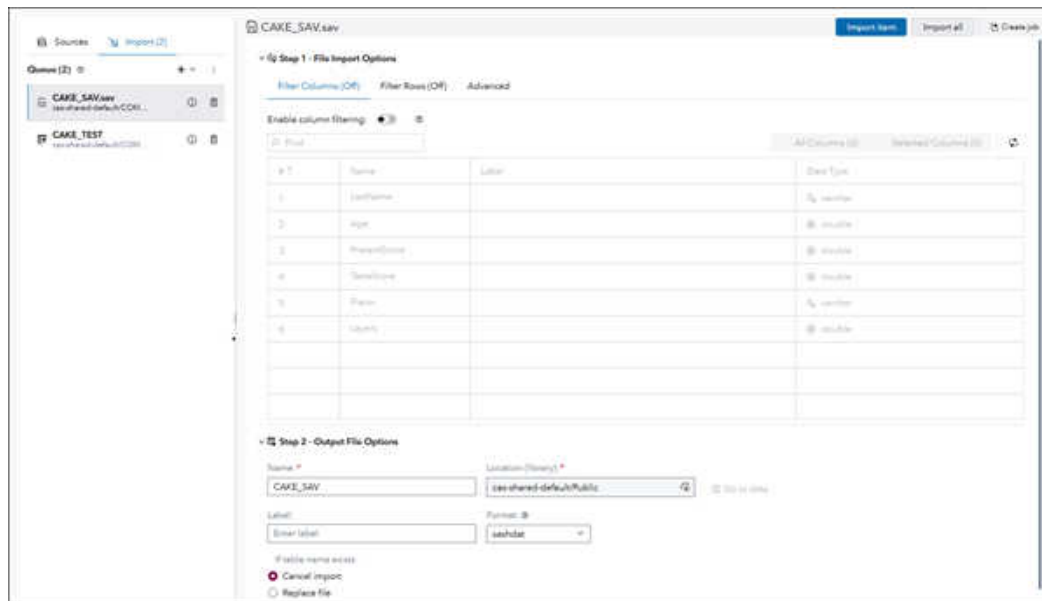
Copying Data from the In-Memory Data Item or the Sources Tab

Overview of the Add to Import Option

The **Add to import** option enables you to copy a table or a file on the **In-memory data (available)** list or the **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 26 Physical Table in the Queue on the Import Tab



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

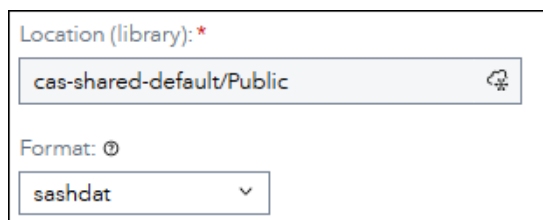
The **Name** field is usually set to the source table name. Accept or change this name as appropriate for the copy.

The name of the default caslib is copied into the **Location (library)** field. If there is no default caslib, the Location (library) remains blank. You can accept or change this caslib as needed.

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

A **Format** selector for the target table is available when certain types of caslibs are selected in the **Location (library)** field.

Figure 27 Format Selector for a Target Table



The **Format** selector enables you to copy the source table to a target table in a different format. For more information, see [“Copying Tables in One Format to Another Format” on page 63](#).

Note the **Filter Columns** tab and **Filter Rows** tabs. These tabs are available for physical tables or files only, not for in-memory tables or files. You can use these tabs to create a copy that contains a subset of columns or rows from the original.

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

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 a 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 66](#).

Note: Any items in the queue on the **Import** tab are dropped from the queue when you log off. The target tables or files that you load to memory persist beyond the current session. The targets are associated with a global caslib in the import properties. For more information about global caslibs, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).

Prerequisites for the Add to Import Option

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

- 1 Identify the source table or file on the **In-memory data (available)** list or the **Sources** tab that you want to copy. Note the size of the item that you want to import. Importing a large table or file can be time-consuming.
- 2 The name of the default caslib, if one exists. This value is copied into the **Location (library)** field. If there is no default caslib, the **Location (library)** remains blank. 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 65](#).
- 3 The name of the source table or file is copied into the **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.

Note: When Windows file semantics are active, then the SASHDAT files that are the source of a CAS table that is currently loaded in-memory are memory-mapped. These files are locked and cannot be overwritten until the in-memory table is unloaded.

- 4 If you want to copy a source table to a target table in another format, identify an appropriate format and target caslib.
- 5 For physical tables or files only, you can create a copy that has a subset of columns or rows from the original. Plan which rows or columns would be appropriate for the copy. For more information, see [“Filter Rows from an Imported Table or File” on page 65](#) or [“Select Columns from an Imported Table or File” on page 66](#).

Add to Import from the Sources Tab or the In-Memory Data List

Use information that you previously gathered to copy a table or a file on the **In-memory data (available)** list or the **Sources** tab:

- 1 .Click  to open the Data Explorer page.
- 2 Click the **In-memory data (available)** list or the **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 **Name** field is usually set to the source table name. Accept or change this name as appropriate for the copy.

If you change the table name, follow the conventions for table names as described in [“Names for Caslibs, Tables, and Columns” on page 86](#). 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 **Location (library)** 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. When Windows file semantics are active, the SASHDAT files that are the source of a CAS table that is currently loaded in-memory are memory-mapped. These files are locked and cannot be overwritten until the in-memory table is unloaded.

- 6 The name of the default caslib was copied into the **Location (library)** field. If there is no default caslib, the **Location (library)** remains blank. 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 65](#).

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 86](#). For more information about global caslibs, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).

- 7 Review the setting for the **Save as in-memory data only** check box. Select this check box to load the imported table to memory without saving a physical copy of the table to the target destination. Only users who have the Promote permission on the target caslib can successfully perform an import using this option.
- 8 Specify what action the import operation should take if the target table name exists in the caslib that is specified in the **Location (library)** 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 field. The options are to cancel the import or to replace the existing table that has the same name.

- 9 Specify a label that helps identify the copy, if desired. The caslib for the target table must support labels. Caslibs of type PATH and DNFS (SASHDAT only) support labels, for example. Caslibs for DBMS and HDFS do not support labels..
- 10 A **Format** selector for the target table is available when certain types of caslibs are selected in the **Location (library)** field. Caslibs of type PATH, DNFS, and HDFS support formats. A caslib for a DBMS does not support formats. To store the copy elsewhere, you can select a different global caslib or add a new one. The selector enables you to copy the source table to a target table in a different format. Select a format that is appropriate for your goals.
- 11 If the item to be copied is a physical item (a table or file with  beside it), the **Filter Rows** tab and **Filter 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 65](#) and [“Select Columns from an Imported Table or File” on page 66](#).

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

- 12 When ready, you can run the import operation. Do one of the following:
 - You can right-click the table or file to be copied and select **Import item**. If the import succeeds, a copy of the table or file is loaded to memory on the CAS server that is specified in the caslib. The copy of the table or file can be selected from the **In-memory data (available)** list or the **Sources** tab. If the import fails, see [“General Usage Notes” on page 84](#).

The **Import item** option is used to run a job that imports the selected data. You can review the status of import jobs on the **Monitoring** tab of the Jobs page in SAS Environment Manager. For more information about using this tab, see [“Monitor Jobs” in SAS Environment Manager: User’s Guide](#).
 - For more information about local file import, see <https://developer.sas.com/rest-apis/compute/updateFileContent>.
 - If you right-click the table or file in the import queue, and the **Create a 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 66](#). 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 **In-memory data (available)** list or the **Sources** tab.

Copying Tables in One Format to Another Format

Overview of the Format Selector

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

Figure 28 Format Selector for a Target Table


Location (library): *

cas-shared-default/Public

Format: ⓘ

sashdat

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

PATH

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

DNFS

sashdat, csv, txt, parquet

HDFS

sashdat, csv, txt

ADLS

orc, csv

S3

sashdat, csv, txt, parquet

Other caslibs

The **Format** selector is hidden.

Usage Notes for the Format Selector

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

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

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

Verify that the target caslib is appropriate for your goals:

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

Copy Data from One Caslib to Another

You can use the **Add to import** option to copy data from one caslib to another. For example, you can use caslibs to copy data from one DBMS to another. To do this, see [“Add to Import from the Sources Tab or the In-Memory Data List” on page 62](#) and specify a caslib in the **Location (library)** field that is different from the source caslib. This creates a physical copy of the data in the target caslib as well as an in-memory copy.

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

Filter Rows from an Imported Table or File

As described in [“Add to Import from the Sources Tab or the In-Memory Data List” on page 62](#), 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” in SAS Language Reference: Concepts](#). The SAS expression is translated into native database syntax and is pushed down to the database, as appropriate.

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

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

- 1 If you have not done so already, perform steps 1–9 as described in [“Add to Import from the Sources Tab or the In-Memory Data List” on page 62](#).
- 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 **Location (library)** field. The copy of the table or file can be selected from the **In-memory data (available)** list or the **Sources** tab.

If the import fails, see [“Usage Notes for Importing Local Files” on page 45](#).

Select Columns from an Imported Table or File

As described in [“Add to Import from the Sources Tab or the In-Memory Data List” on page 62](#), you can right-click a physical table or file and select **Add to import**. The table or file is added to the queue on the **Import** tab. The **Select Columns** tab appears on the right.

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

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

- 1 If you have not done so already, perform steps 1–9 as described in [“Add to Import from the Sources Tab or the In-Memory Data List” on page 62](#).

- 2 Select **Enable column filtering** in the **Filter Columns** tab.

The **Filter Columns** tab displays a table that lists the columns that are available in the source table. Select one or more columns in the source table for inclusion in the import job.

- Use the **Find** field to locate a column buried in a long list.
- Select the check box in the header row of a table to change the status of all the rows at the same time. When some or all of the columns are not selected, you can click the check box to select them all. When some or all of the columns are selected, you can click the check box to deselect them all.
- Click **All Columns** or **Selected Columns** to switch between a table that displays all of the available columns and a different table that displays only the columns that you have selected.
- Select a check box next to a row to add that column to the selected columns table. Deselect a selected row to remove it from the selected columns table.

- 3 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 **Location (library)** 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 **In-memory data (available)** list or the **Sources** tab.

Create Import Job Requests

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

The **Create job** option is not valid for local files, social media feeds, or Esri data. It is valid only for tables or files that have been copied from the **In-memory data (available)** list or the **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 Sources Tab and Import Tab” on page 84](#).

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

- 1 If you have not done so already, perform steps 1–9 as described in [“Add to Import from the Sources Tab or the In-Memory Data List” on page 62.](#)
- 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 **Name** field cannot be the same as the original unless you have selected the **Replace file** option.

Note: When Windows file semantics are active, then the SASHDAT files that are the source of a CAS table that is currently loaded in-memory are memory-mapped. These files are locked and cannot be overwritten until the in-memory table is unloaded.

- 4 When ready, right-click the table or file and select the **Create job** option. The Create Job window is displayed.
- 5 Specify a name and description for the job and click **OK**. The name cannot be longer than 100 characters. A status message is displayed, stating whether the job request was created successfully.

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

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

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 **In-memory data (available)** list or the **Sources** tab.

Import a Snapshot of an ESP Project

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

To import a snapshot of an ESP project:

- 1 Add a user-defined connection to a project on an ESP server. See [“Example: User-Defined Connection to an ESP Project” on page 35.](#)

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

- 2 Import one of these tables and load it to memory. See [“Add to Import from the Sources Tab or the In-Memory Data List” on page 62.](#)

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

- 3 If the import is successful, you can right-click the loaded table and run tasks with this snapshot of streamed data. See [“Working with Tables” on page 56.](#)

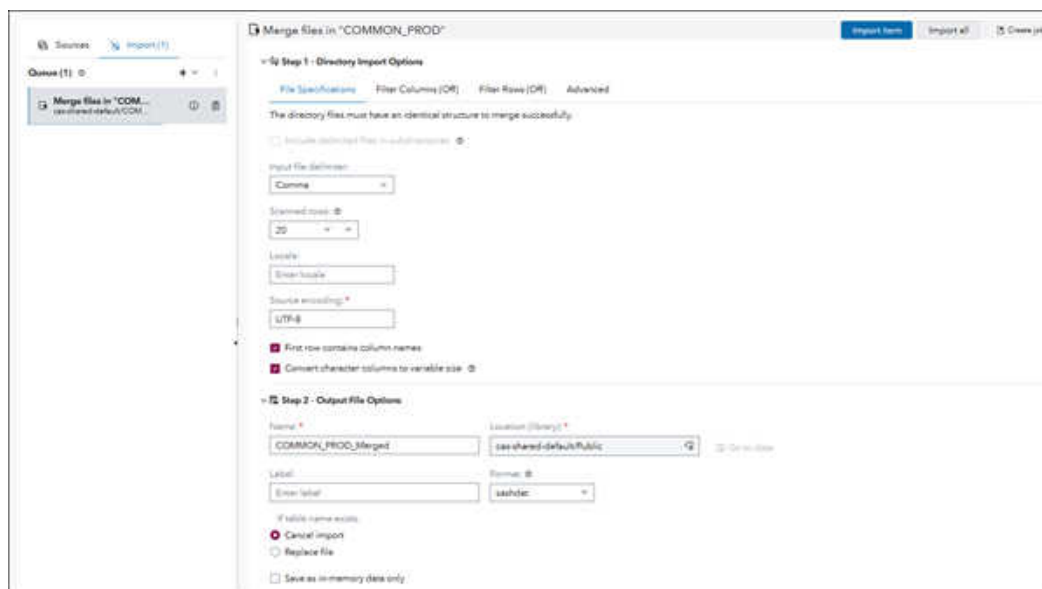
Importing Delimited Files from a Caslib Directory

Overview of Importing Delimited Files from a Directory

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

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

Figure 29 Merge Files Item in the Queue on the Import Tab



Import properties for the **Merge files** item appear on the right. You specify various options and import the files that are specified in the item. The contents of the files are copied into the table in

the **Name** field and loaded to memory. The target table is in the caslib that is specified in the **Location (library)** field.

Prerequisites for Importing Delimited Files from a Directory

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

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

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

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

Select the **Include delimited files subdirectories** check box in the **File Specifications** tab when you add the caslib. This setting is required in order to support the Choose a Directory window on the **Import** tab. This option is enabled only when the selected caslib contains subdirectories.

For more information about PATH and DNFS caslibs, see [“Connecting to Remote File Systems” on page 28](#).

- 4 Add or identify the target caslib.

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

Import Delimited Files from a Directory

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

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

- a Click the **Sources** tab.
- b Right-click a path-based caslib that contains the delimited files that you want to analyze. Select **Import from library** ⇒ **Delimited files**.

A **Merge files** item is added to the import queue, and import properties appear on the right. Go to Step 2.

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

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

Note: The filter discovers matching items in the nodes that you have previously expanded. You must manually expand the tree hierarchy to the depth that you need to search.

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

- 2 In the **Step 1 - Directory Import Options** section, use the options on the **File Specifications** tab as needed. Click ⓘ for information about these options.
- 3 Use the **Filter Rows** tab and **Filter Columns** tab to subset columns or rows from the original file. For more information about these tabs, see [“Filter Rows from an Imported Table or File” on page 65](#) and [“Select Columns from an Imported Table or File” on page 66](#).
- 4 Use the options on the **Advanced** tab as needed. Click ⓘ for information about these options. To add options, select **Create additional input options**. Options are entered as name/value pairs. Click **Add Option** to add the option to the target caslib.

- 5 In the **Step 2 - Output File Options** section, the **Name** is the name of the caslib or the selected directory with the suffix **_Merged**. This name specifies the output table that contains the extracted information. Accept or change this name as appropriate.

If you change the name, follow the conventions for table names as described in [“Names for Caslibs, Tables, and Columns” on page 86](#). The target operating system and data source might have additional constraints on these names. For example, the target table name can be a maximum of 247 characters long. However, a shorter name might be required due to the data source settings for the caslib that is specified in the **Location (library)** field. Contact your data administrator for more information about the permitted length for target table names.

- 6 In the **Location (library)** field, select an existing global caslib or add a new one.

The caslib must be a global caslib on the same CAS server as the source caslib. If you have the privilege that is required to add a new global caslib, follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 86](#). For more information about global caslibs, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).

- 7 Specify a label that helps identify the output table, if desired. The caslib for the target table must support labels. Caslibs of type PATH and DNFS (SASHDAT only) support labels, for example. Caslibs for DBMS and HDFS do not support labels.
- 8 Specify what action the import operation should take if the target table name exists in the caslib that is specified in the **Location (library)** field. The options are to cancel the import or to replace the existing table that has the same name.
- 9 A **Format** selector for the target table is available when certain types of caslibs are selected in the **Location (library)** field. The selector enables you to copy the source table to a target table in a different format. Select a format that is appropriate for your goals. For more information, see [“Copying Tables in One Format to Another Format” on page 63](#).

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

If the import succeeds, contents from files in the source caslib are copied into a table in the target caslib. The output table is loaded to memory on the CAS server that is specified in the **Location (library)** field.

To verify the output, select the output table on the **In-memory data (available)** list or the **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 87](#).

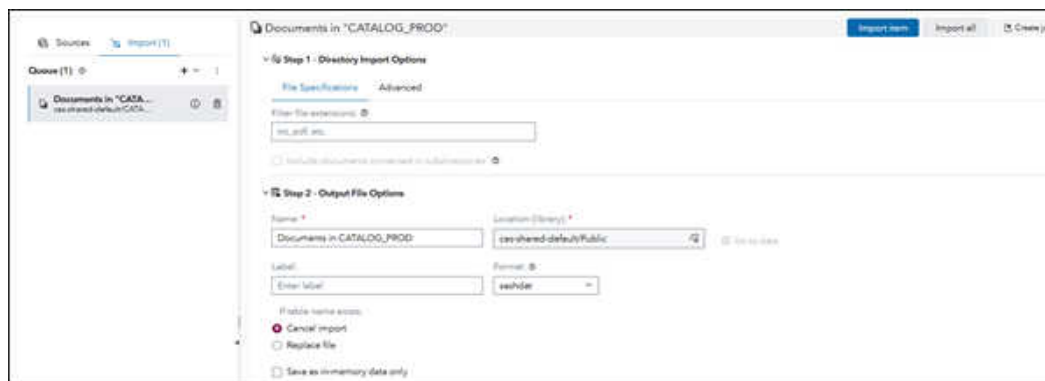
Importing Documents from a Caslib Directory

Overview of Importing Documents

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

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

Figure 30 Document Item in the Queue on the Import Tab



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

Prerequisites for Importing Documents

Perform the following tasks before importing a directory of documents:

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

Table 10 Document Formats That Can Be Converted

Format	File Extensions
HTML	.html
XML	.xml

Format	File Extensions
MS Office	.doc, .docx, .xls, .xlsx, .ppt, .pptx
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 caslib is a caslib of type PATH whose root directory can include documents and subdirectories of documents that you want to analyze.

Select the **Include documents included in subdirectories** check box in the **File Specifications** tab when you add the caslib. This setting is required in order to support the Choose a Directory window on the **Import** tab. This option is enabled only when the selected caslib contains subdirectories. For more information about adding PATH caslibs, see [“Connecting to Remote File Systems” on page 28](#).

4 Add or identify the target caslib.

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

Import Documents from a Directory

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

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

a Click the **Sources** tab.

b Right-click a path-based caslib that contains the documents that you want to analyze. Select **Import from library** ⇒ **Documents**.

A **Documents** item is added to the import queue, and import properties appear on the right. Go to Step 2.

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

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

Note: The filter discovers matching items in the nodes that you have previously expanded. You must manually expand the tree hierarchy to the depth that you need to search.

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

- 2 In the **Step 1 - Directory Input Options** section, use the options on the **File Specifications** tab as needed. Click ⓘ for information about these options.
- 3 Use the options on the **Advanced** tab as needed. Click ⓘ for information about these options. To add options, select **Create additional input options**. Options are entered as name/value pairs. Click **Add Option** to add the option to the target caslib.
- 4 In the **Step 2 - Output File Options** section, the default **Name** is *Documents in <Source Name>*. This is the name of the output table that contains the extracted text and metadata. Accept or change this name as appropriate. Note that you must also enter a value in the **Short name** field.

If you change the name, follow the conventions for table names as described in [“Names for Caslibs, Tables, and Columns” on page 86](#). The target operating system and data source might have additional constraints on these names. For example, the target table name can be a maximum of 247 characters long. However, a shorter name might be required due to the data source settings for the caslib that is specified in the **Target location** field. Contact your data administrator for more information about the permitted length for target table names.
- 5 In the **Location (library)** field, select an existing global caslib or add a new one.

The caslib must be a global caslib on the same CAS server as the source caslib. If you have the privilege that is required to add a new global caslib, follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 86](#). For more information about global caslibs, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).
- 6 Specify a label that helps identify the output table, if desired. The caslib for the target table must support labels. Caslibs of type PATH and DNFS (SASHDAT only) support labels, for example. Caslibs for DBMS and HDFS do not support labels.
- 7 A **Format** selector for the target table is available when certain types of caslibs are selected in the **Location (library)** field. The selector enables you to copy the source table to a target table in a different format. Select a format that is appropriate for your goals. For more information, see [“Copying Tables in One Format to Another Format” on page 63](#).
- 8 Specify what action the import operation should take if the target table name exists in the caslib that is specified in the **Location (library)** field. The options are to cancel the import or to replace the existing table that has the same name.

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

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

To verify the output, select the output table on the **In-memory data (available)** list or the **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 87](#).

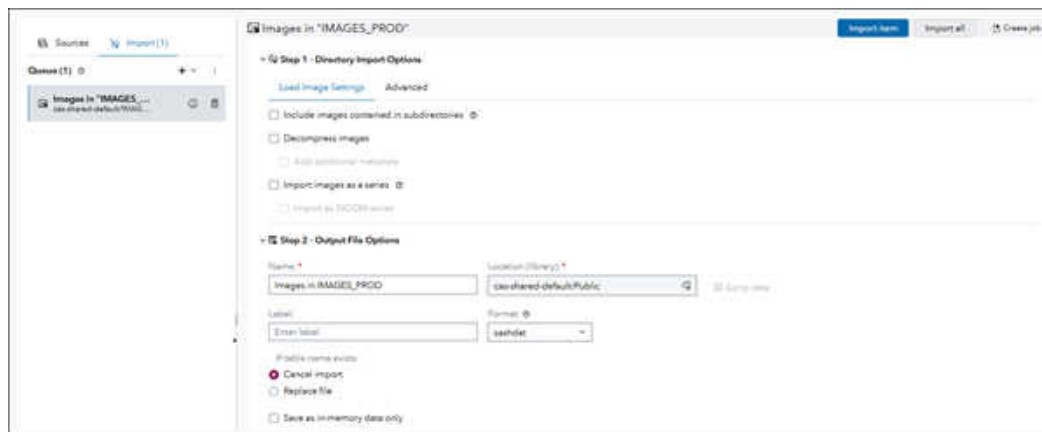
Importing Images from a Caslib Directory

Overview of Importing Images

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

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

Figure 31 Images Item in the Queue on the Import Tab



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

Prerequisites for Importing Images

Perform the following tasks before importing a directory of images:

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

Table 11 *Image Formats That Can Be Imported*

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

Table 12 *Image Types that Can Be Viewed in SAS Data Hub*

Image Format	File Extensions
Windows bitmap	.bmp
Joint Photographic Experts Group (JPEG)	.jpeg, .jpg
Portable Network Graphics	.png
Graphics Interchange Format	.gif

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

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

Select the **Include images included in subdirectories** check box in the **Load Image Settings** tab when you add the caslib. This setting is required in order to support the Choose a Directory window on the **Import** tab. This option is enabled only when the selected caslib contains subdirectories. For more information about adding PATH caslibs, see [“Connecting to Remote File Systems” on page 28](#).

- 4 Add or identify the target caslib.

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

Import Images from a Directory

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

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

- a Click the **Sources** tab.
- b Right-click a path-based caslib that contains the images that you want to analyze. Select **Import from library** ⇒ **Image files**.

An **Images** item is added to the import queue, and import properties appear on the right. Go to Step 2.

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

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

Note: The filter discovers matching items in the nodes that you have previously expanded. You must manually expand the tree hierarchy to the depth that you need to search.

An **Images** item is added to the import queue, and import properties appear on the right.

- 2 In the **Step 1 - Directory Import Options** section, use the options on the **Load Image Settings** tab as needed.

Include images contained in subdirectories.

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

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

Import images as a series.

- To combine multiple images that are not in DICOM format into a single 3-D image, select **Import images as a series**. Make sure that **Import as a DICOM series** is deselected. The import operation reads all images from the specified directories and attempts to layer them into one 3-D image. The images must be photographic images with the same width and height. The images are layered in the alphabetical order of their file names. Only those images that constitute a 3-D image are fully supported by this option.
 - To combine multiple images in DICOM format into a series, select both **Import images as a series** and **Import as a DICOM series**. The import operation reads all images from the specified directories and attempts to assemble them into a DICOM series. Each image in the series is assigned a unique path that is based on the DICOM Unique Identifier (UID) of the series.
- 3 Use the options on the **Advanced** tab as needed. Click ⓘ for information about these options. To add options, select **Create additional input options**. Options are entered as name/value pairs. Click **Add Option** to add the option to the target caslib.
 - 4 In the **Step 2 - Output File Options** section, the default **Name** is the name of the selected directory. This name specifies the output table that contains the extracted information. Accept or change this name as appropriate.

If you change the name, follow the conventions for table names as described in [“Names for Caslibs, Tables, and Columns” on page 86](#). The target operating system and data source might have additional constraints on these names. For example, the target table name can be a maximum of 247 characters long. However, a shorter name might be required due to the data source settings for the caslib that is specified in the **Target location** field. Contact your data administrator for more information about the permitted length for target table names.
 - 5 In the **Location (library)** field, select an existing global caslib or add a new one.

The caslib must be a global caslib on the same CAS server as the source caslib. If you have the privilege that is required to add a new global caslib, follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 86](#). For more information about global caslibs, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).
 - 6 Specify a label that helps identify the output table, if desired. The caslib for the target table must support labels. Caslibs of type PATH and DNFS (SASHDAT only) support labels, for example. Caslibs for DBMS and HDFS do not support labels.
 - 7 A **Format** selector for the target table is available when certain types of caslibs are selected in the **Location (library)** field. The selector enables you to copy the source table to a target table in a different format. Select a format that is appropriate for your goals. For more information, see [“Copying Tables in One Format to Another Format” on page 63](#).
 - 8 Specify what action the import operation should take if the target table name exists in the caslib that is specified in the **Location (library)** field. The options are to cancel the import or to replace the existing table that has the same name.

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

If the import succeeds, text and metadata from images in the source caslib are written to a table in the target caslib. The output table is loaded to memory on the CAS server that is specified in the **Location (library)** field.

To verify the output, select the output table on the **In-memory data (available)** list or the **Sources** tab. Click the **Sample Data** tab to see a sample of the output. Click an image to preview it. Some image formats cannot be displayed by some browsers unless they are decompressed. Other formats might not be displayable unless the browser is configured to launch an external viewer.

If the import fails, see [“If Data Access Fails” on page 87](#).

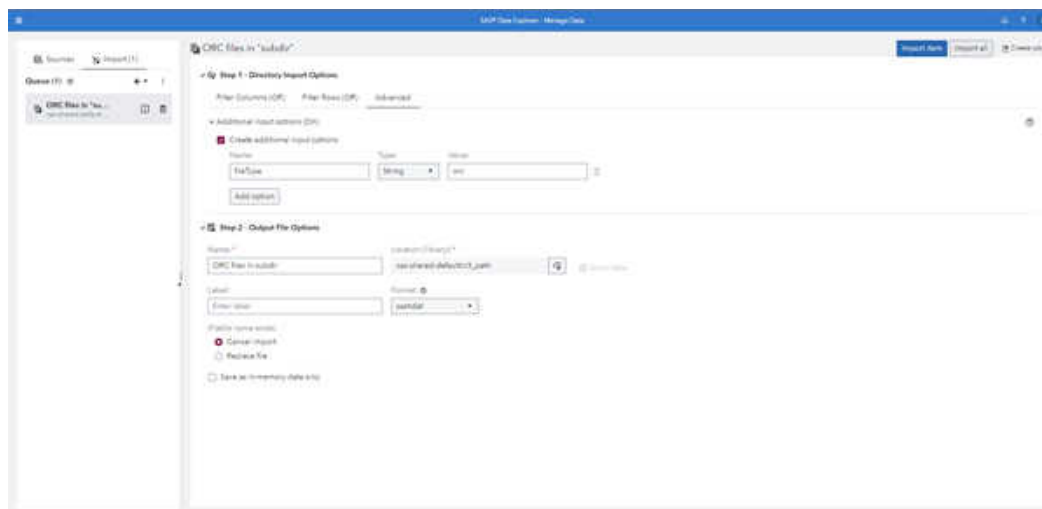
Importing ORC Files from a Caslib Directory

Overview of Importing ORC Files

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

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

Figure 32 ORC Files Item in the Queue on the Import Tab



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

Prerequisites for Importing ORC Files

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

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

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

- 3 Add the source caslib.

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

Select the **Include ORC files subdirectories** check box in the in the **CAS** section of the **Advanced** tab when you add the caslib that contains the ORC files. This setting is required in order to support the Choose a Directory window on the **Import** tab. This option is enabled only when the selected caslib contains subdirectories.

For more information about adding path-based caslibs, see [“Connecting to Remote File Systems” on page 28](#).

- 4 Add or identify the target caslib.

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

Import ORC Files from a Directory

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

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

Note: The filter discovers matching items in the nodes that you have previously expanded. You must manually expand the tree hierarchy to the depth that you need to search.

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

- 4 In the **Step 1 - Directory Import Options** section, use the **Filter Rows** tab and **Filter Columns** tab to subset columns or rows from the original files. For more information about these tabs, see [“Filter Rows from an Imported Table or File” on page 65](#) and [“Select Columns from an Imported Table or File” on page 66](#).

- 5 Use the **Advanced** tab with the **Additional Input Options** control as needed. A **fileType** option is provided by default. No additional options are available for ORC files.
- 6 In the **Step 2 - Output File Options** section, the default **Name** is the name of the selected directory. This name specifies the output table that contains the extracted information. Accept or change this name as appropriate. Note that you must also enter a value in the **Short name** field.

If you change the name, follow the conventions for table names as described in [“Names for Caslibs, Tables, and Columns” on page 86](#). The target operating system and data source might have additional constraints on these names. For example, the target table name can be a maximum of 247 characters long. However, a shorter name might be required due to the data source settings for the caslib that is specified in the **Location (library)** field. Contact your data administrator for more information about the permitted length for target table names.

- 7 In the **Location (library)** field, select an existing global caslib or add a new one.
The caslib must be a global caslib on the same CAS server as the source caslib. If you have the privilege that is required to add a new global caslib, follow the conventions for caslib names as described in [“Names for Caslibs, Tables, and Columns” on page 86](#). For more information about global caslibs, see [“Caslibs on the Sources Tab and Import Tab” on page 84](#).
- 8 Specify a label that helps identify the output table, if desired. The caslib for the target table must support labels. Caslibs of type PATH and DNFS (SASHDAT only) support labels, for example. A Caslibs DBMS and HDFS do not support labels.
- 9 A **Format** selector for the target table is available when certain types of caslibs are selected in the **Location (library)** field. The selector enables you to copy the source table to a target table in a different format. Select a format that is appropriate for your goals. For more information, see [“Copying Tables in One Format to Another Format” on page 63](#).
- 10 Specify what action the import operation should take if the target table name exists in the caslib that is specified in the **Location (library)** field. The options are to cancel the import or to replace the existing table that has the same name.
- 11 When ready, you can right-click the **ORC files** item in the queue and select **Import item**.

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

To verify the output, select the output table on the **In-memory data (available)** list or the **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 87](#).

Managing the Access Policy for a Connection

You can manage the authorization for data connections by editing the access policy for the connection. You can modify the access policy by using the **Access Policy** tab in the Connection Details window, by right-clicking a connection and selecting **Access policy**, or by selecting the **Access Policy** tab in the right pane of the Manage Connections window and clicking **Edit**.

To add users or groups, click **Users and Groups** and then select the users or groups that you want to add.

For each user and group, click **Access Type** to select the type of access that is granted. You can override individual permissions by clicking the icons in the table.

Here are the default settings for each access type selection:

Table 13 Access Type Values

Access Type	Description	Rule or Action	Permissions
Read-only	Grants permission to view and use the connection.	grant	View 
Modify	Grants permission to view, use, and edit the connection.	grant	View  , edit  , and delete 
Full access	Grants permission to view, use, edit, delete, and secure the connection.	grant	View  , edit  , delete  , and secure 
No access	Prohibits all access to the connection.	prohibit	None
Inherit	Inherits the permissions of any parent groups.	various	Various

Denying the View permission effectively prevents that user or group from making any use of the connection or its data.

The edit, delete, and secure permissions on a connection affect only the connection definition. Access to the data within a connection is dependent on table-level and row-level authorization rules. For more information, see [“Secure Access to Library Connections” in SAS Viya Platform: General Authorization](#).

Maintaining SAS Data Hub

Administrative Tasks for SAS Data Hub 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 Hub application unless stated otherwise.

Controlling Access to Features

Limit Access to the Download Table Option

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

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

Limit Access to the Import Tab

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

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

Import tab

`/casManagement_capabilities/importData`

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

Limit Access to Job Definitions

The ability to run most import actions in SAS Data Hub can be restricted to a group that includes administrators and other users who retain READ permission to the folder containing the job definitions used for import. When this option is implemented, the only data that is not restricted for import within SAS Data Hub in this way is local files, multi-sheet Excel files, and the content in Viya folders.

To limit access to job definitions, an administrator adds the `/folders/folders/9cb58ee6-fde2-403c-bd06-cfc4d763b3a6` rule in SAS Environment Manager. Enter the following values in the New Rule window:

- Object URI: `/folders/folders/9cb58ee6-fde2-403c-bd06-cfc4d763b3a6`
- Container URI: `/folders/folders/9cb58ee6-fde2-403c-bd06-cfc4d763b3a6`
- Principal Type: **User, Group, or Custom Group**
- Principal: an appropriate name for the user or group affected by the new rule
- Rule Type: *Prohibit*
- Permissions: *Read*

When a user or a member of a group affected by this restriction attempts to import data, an error message is displayed. For information about updating rules, administrators can refer to the topic for the [“Rules Page” in SAS Environment Manager: User’s Guide](#).

CAS Management Service Options

Administrators can change how some features work in SAS Data Hub updating service options. Service options are updated from the Configuration page in SAS Environment Manager. For more information, see [“Configuration Page” in SAS Environment Manager: User’s Guide](#).

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

SAS Data Hub Options

Modify SAS Data Hub Settings

To modify settings for the SAS Data Hub web application:

- 1 In the application bar, click your name, and then click **Settings**. The Settings window is displayed.
- 2 The following settings are available:
 - Global** settings. These settings are applied to all SAS web applications. Click ⓘ for details about each setting.

Data Explorer settings. These settings are saved on a per-user basis. All of your settings persist between sessions.

General settings.

There are two user preferences in this section.

Default target location

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

If a default target location is not set in this field, the default CAS server and caslib for the CAS Management Service is used.

Geographic Mapping settings.

Use the options in this section to accept the terms and conditions for Esri ArcGIS Online Services. For more information about these options, see [“Importing Esri Data for Geoenrichment and Geocoding” on page 47](#).

General Usage Notes

Caslibs on the Sources Tab and Import Tab

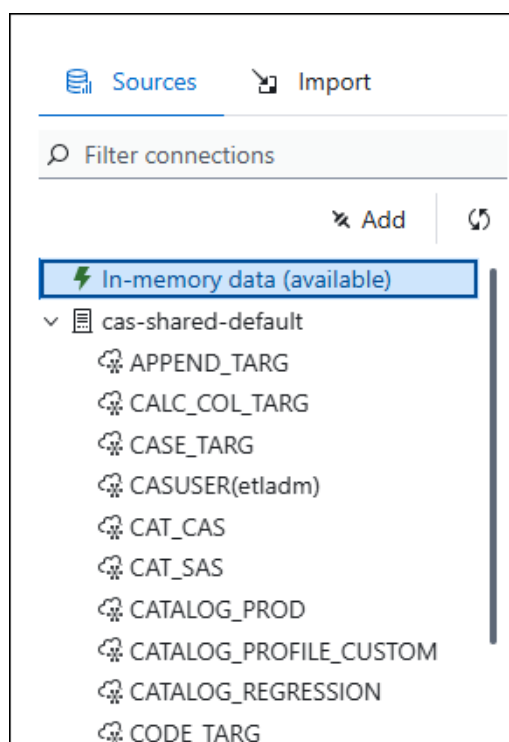
Overview

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

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

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

Figure 33 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 **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 into memory.** This option is used to load a table or file to memory on a CAS server.
- **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 In-Memory Data Item or the Sources Tab” on page 59.](#)
- Only global caslibs can be selected in the **Location (library)** 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, Predefined, and Manually Added Caslibs” in SAS Cloud Analytic Services: Fundamentals.](#)

Caslibs and the Sources Tab

The **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 **Sources** tab, the **Select a Data Source** window is displayed. This dialog box enables you to add a new caslib that connects to a data source. When you select a data source, a Connection Details window is displayed for that data type.

The **Connection Details** window has an option for setting the connection scope. If you select **Session** and specify a connection, the caslib that is added is a session-based caslib. If you select **Global (Shared)** and specify a connection, the caslib that is added is a global caslib. Your login must have the privilege that is required to add a global caslib, or the connection fails.

You can check to see what options are available for the tables or files in the new caslib. If the saved connection uses a global caslib, the following options are available when you right-click a table in the caslib: **Load into memory**, **Delete**, and **Add to import**. For information about these options, see [“Working with Tables” on page 56](#).

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 into memory** option, which is used to load a table or file to memory on a CAS server.

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 into memory** option. For information about the **Add to import** option, see [“Copying Data from the In-Memory Data Item or the Sources Tab” on page 59](#).

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” in SAS Cloud Analytic Services: Fundamentals](#). The target operating system and data source might have additional constraints on these names. For example, a colon cannot be used in a Windows file name.

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

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

Specifying a Default Target Location

You can select a caslib where target tables is stored by default. This caslib serves as the default for various operations, including operations on the **Import** tab. To make this selection, right-click a global caslib on the **Sources** tab and select **Set as default target library**. See the **Default target location** setting in [“Modify SAS Data Hub Settings” on page 83](#).

Enabling Other Users to Access a Global Caslib

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

If Data Access Fails

If you fail to access data from the **Sources** tab or the **Import** tab, an error message is displayed. For connections, you can click **View** in the **Log** column on the **Manage Connections** page to see details.

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 **In-memory data (available)** list and the **Sources** tab 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 Hub supports the databases, browsers, and other third-party software that are supported by the SAS Viya platform. For more information, see [Third-Party Software Requirements for Use with SAS Viya](#).

Performance Options for the In-Memory Data List

On large systems, the **In-memory data (available)** list might be slow to refresh. An administrator can set the following SAS Data Hub service options for the **In-memory data (available)** list.

filterAvailableTab

Specifies whether the **In-memory data (available)** list should display only loaded tables from the default caslib. You can use the **Sources** tab to work with any tables that are subsequently loaded from other caslibs.

filterAvailableTabByCaslibList

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

filterAvailableTabCasliblist

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

useBatchQueryForAvailableTab

Enhances the performance of the **In-memory data (available)** list. On by default. Available as of 2021.2.1.

To set these options, see [“CAS Management Service Options” on page 83](#).

No Tables Are Displayed on In-Memory Data List

If the **In-memory data (available)** list 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 **In-memory data (available)** list might not display any tables. The **Sources** tab might not display the list of initial CAS servers.

The preferred way to address this issue is to edit the SAS Data Hub 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 83](#).

Changing the maxTablePageSize Option

You can change the size of the SAS Data HUB **maxTablePageSize** option in SAS Environment Manager. This option limits the number of tables that are returned when you open a caslib. Lowering this value can prevent overwhelming the browser. When the browser is overwhelmed, out of memory exceptions can occur.

- 1 Log in to SAS Environment Manager as an Administrator.
- 2 Click  to access the Configuration pane.
- 3 Select **Services** from the drop-down list in the **View** field.
- 4 Search for *SAS Data Hub* in the list of search results.
- 5 Click  to access the Edit window for the configuration.
- 6 Scroll down to the **maxTablePageSize** item in the list of configuration instances.
- 7 Enter a number greater than 1 in the **maxTablePageSize** field. The default value is 10000.
- 8 Click **Save** to save the change to the option.