

SAS[®] Information Catalog: Administrator's Guide

2020.1.3 - 2020.1.5

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.

Deploy SAS Information Catalog

Deployment administrators typically install and update SAS Information Catalog and other SAS Viya software. Deployment instructions are included with your customer order. Instructions can also be found in the SAS Viya: Deployment Guide.

If Elasticsearch is not working, verify that the Elasticsearch customization was applied during deployment. See "Configure Open Distro for Elasticsearch" in SAS Viya: Deployment Guide.

Basic and Advanced Features

SAS Information Catalog supports the basic and advanced features that are described in the following table. You must license SAS Information Governance to use the advanced features.

	SAS Information Catalog	SAS Information Governance
User Interfaces	 Home page. Search results. Table properties. Column properties. Discovery agent interface. 	Same interfaces.
Information Assets	Tables and columns can be indexed in the catalog.	Same assets.
Discovery Agents	A discovery agent can crawl assets in a CAS library (caslib) only.	A discovery agent can crawl assets in a SAS compute library or a caslib.
Metrics	 Data profiling. Wide table handling by subsetting columns. 	 The following additional metrics are supported: Semantic type, based on Identifiable Data Analysis of field names and field content, using SAS Quality Knowledge Base. See "Understanding Content Analysis" on page 13. Information privacy detection. Data sampling of big or wide tables is done with the srs CAS action or the SURVEYSELECT procedure.
Visualizations	 Visualize assets by physical region. Visualize assets by data type, cardinality, and other metrics. 	 The following additional visualizations are supported: Drill-in metrics for semantic type. Drill-in metrics for information privacy.
Architecture	 Customized Elasticsearch search engine. PostgreSQL for catalog storage. 	Same architecture.
Lineage	Tables and columns from caslibs can be viewed in SAS Lineage Viewer.	Tables and columns from SAS compute libraries and caslibs can be viewed in SAS Lineage Viewer.

Review Status	Not available.	Administrators can use the Status property to indicate that an asset is under review, is approved, has been flagged for attention, or has a warning.
Deployment	 Cloud-enabled. Validated on Red Hat OpenStack and Microsoft Azure. 	Same deployment.
	 Lineage relationships are automatically migrated from SAS Viya 3.5 and later. 	

Main Tasks for Administrators

By default, members of the SAS Administrator group can create discovery agents that capture and enrich metadata for information assets. This metadata is stored in a catalog. Users can search the catalog, explore details about assets, take actions on assets, and build collections of searches.

Administrators perform the following tasks to load metadata into the catalog:

- Identify databases, files, or other information assets that should be indexed in the catalog.
- Identify or create libraries for assets. See "Prepare Libraries" on page 4.
- Add a discovery agent for each library. See "Add and Run Discovery Agents" on page 10.

Administrators perform tasks such as the following to maintain the catalog:

- Monitor discovery agent jobs in SAS Environment Manager. See "Monitor Agent Jobs" on page 12.
- Respond to notifications, such as a notification to rebuild the catalog index. See "Respond to Notifications" on page 16.
- Assign a review status to an asset, such as Approved. See "Assign a Review Status to an Asset" on page 16.

Prepare Libraries

Overview

Each SAS Information Catalog discovery agent crawls one library and ingests metadata from the assets in that library. The library must be configured as a global library and must be visible to everyone.

If you have basic SAS Information Catalog, assets must be in a SAS Cloud Analytic Services (CAS) library. If you license SAS Information Governance, assets can be in a SAS compute library or a CAS library.

Perform the following tasks before you create a discovery agent:

- Identify databases, files, or other information assets that should be indexed in SAS Information Catalog.
- Identify an existing CAS library or SAS compute library that can be used to access these assets.
- Create a library for any assets that do not have a library.

Note: After a library is associated with an agent, it cannot be associated with any other agent. This rule helps to prevent duplicate information from being loaded into the catalog.

.....

General Usage Notes for Libraries and Tables

Some Libraries Do Not Support Some Metrics

Some libraries do not support some SAS Information Catalog metrics. For example, most database libraries do not return the last updated date or who created or updated an asset.

Some Metrics Might Not Be Returned from Big and Wide Tables

Statistical sampling is used in some cases to process big and wide tables. CAS libraries, which can use CAS data connectors, might be more efficient than SAS compute libraries for processing very large data.

CAS Libraries and SAS Compute Libraries Support Different Actions

When users run a search in SAS Information Catalog, the results are displayed in a Search Results window. If you select an asset in this window, the **Actions** menu at the top right becomes active. The available actions are different for different types of assets.

For example, if the selected asset is a table in a CAS library and you have licensed SAS Visual Analytics, you can select **Explore and visualize** to open the CAS table in SAS Visual Analytics. If the selected asset is a table in a SAS compute library, you have fewer options.

Preparing CAS Libraries (Caslibs)

Overview

SAS Cloud Analytic Services (CAS) libraries are referred to as caslibs. A caslib is an in-memory space to hold tables, and access control lists and data source information. Caslibs can use CAS data connectors. A caslib can have one of two scopes: session scope or global scope.

If you have basic SAS Information Catalog, information assets must be in a global caslib that is visible to everyone. If an appropriate caslib does not exist, you can create one.

Data Sources Supported by CAS Libraries

SAS Information Catalog discovery agents can crawl CAS libraries for the following data sources:

- Remote file systems such as DNFS. See "Platform Data Sources" in SAS Cloud Analytic Services: User's Guide.
- Platform files such as delimited files. See "Platform File Types" in SAS Cloud Analytic Services: User's Guide.
- DBMSs such as Oracle. See "Data Connectors" in SAS Cloud Analytic Services: User's Guide.

Known exceptions include the following:

DB2

Use SAS Environment Manager to Create CAS Libraries

One way to create CAS libraries is as follows:

- Click Import Data in the main SAS Information Catalog window.
 You are prompted to switch to SAS Environment Manager to import data.
- 2 Click Continue.

The Data page in SAS Environment Manager is displayed.

3 Click the Data Sources tab.

For instructions on using this tab, click ③

Preparing SAS Compute Libraries

Overview

A SAS library is a collection of one or more SAS files that are recognized by SAS and that are referenced and stored as a unit. You reference a SAS library by a logical name called a libref. The SAS Compute Server provides access to Base SAS in a SAS Viya environment. A SAS compute library is a SAS library (libref) that is managed by a SAS Compute Server. SAS compute libraries can use SAS/ACCESS engines.

If you license SAS Information Governance, discovery agents can crawl SAS compute libraries. The compute library must be configured as a global library and must be visible to everyone. Discovery agents cannot crawl personal libraries such as WORK and SASUSER. If an appropriate library does not exist, you can create one.

Data Sources Supported by SAS Compute Libraries

SAS Information Catalog discovery agents can crawl SAS compute libraries for most of the following data sources:

- SAS data sets. See SAS V9 LIBNAME Engine: Reference.
- PC files. See SAS/ACCESS Interface to PC Files: Reference.
- ORC files. See SAS Viya Engine for ORC: Reference.
- DBMSs such as Oracle. See "DBMS-Specific Reference" in SAS/ACCESS for Relational Databases: Reference.
- Nonrelational DBMSs such as MongoDB. See SAS/ACCESS for Nonrelational Databases: Reference.

Known exceptions include the following:

- Greenplum
- SAP R/3
- DB2

Note: A few libraries require special handling, as noted later in this section.

Use SAS Studio to Create a SAS Compute Library

The easiest way to create a SAS compute library is to use the library creation window in SAS Studio. If you use this window, the **Assign and connect to data sources at startup** option and the **Make source available to all users** option must be enabled for each compute library. See "Working with Libraries" in *SAS Studio: User's Guide*.

Pre-assign a SAS Compute Library

As an alternative to using SAS Studio to create a SAS compute library, you can add a LIBNAME statement in the **sas.compute.server: autoexec_code** configuration instance of the **Compute service**. You can then select this library to be crawled by a discovery agent. For an example of how this can be done, see "Compute Libraries Using PCFILES Engine" on page 8.

Compute Libraries Using SAS/ACCESS to Hadoop

A discovery agent cannot crawl a compute library that uses SAS/ACCESS to Hadoop unless you set some path options in the **sas.compute.server: autoexec_code** configuration instance of the **Compute service**. You can then select this library to be crawled by a discovery agent.

- Identify or create a compute library that uses SAS/ACCESS to Hadoop. See "Preparing SAS Compute Libraries" on page 6.
- 2 Note the Hadoop JAR path and Hadoop config path that is required for the library.
- 3 Select Manage Environment from the application menu at the top left.

SAS Environment Manager is displayed.

4 Click 🖏.

The **Configuration** page is displayed.

5 Type compute into the service filter.

The **Compute service** becomes available.

6 Click Compute service.

Configuration instances of the **Compute service** become available.

7 Type autoexec into the configuration instance filter.

The sas.compute.server: autoexec_code configuration instance becomes available.

- 8 Click to edit the sas.compute.server: autoexec_code configuration instance.
- 9 Add the following statements to the contents area for sas.compute.server: autoexec_code:

```
%let HADOOPJARPATH="_path_";
%let HADOOPCONFIGPATH="_path_";
options set=SAS_HADOOP_JAR_PATH=&HADOOPJARPATH.;
options set=SAS_HADOOP_CONFIG_PATH=&HADOOPCONFIGPATH.;
```

10 Click Save to save your changes.

Compute Libraries for ORC Files in Microsoft Azure Data Lake

You cannot use the library creation window in SAS Studio to create a SAS compute library that will access ORC files in Microsoft Azure Data Lake. You can add an ORC LIBNAME statement in the **sas.compute.server: autoexec_code** configuration instance of the **Compute service**. You can then select this library to be crawled by a discovery agent.

- 1 Identify the path and other attributes that you will need to specify an ORC LIBNAME statement for files in Microsoft Azure Data Lake. See "Azure Requirements" in SAS Viya Engine for ORC: Reference.
- 2 Select Manage Environment from the application menu at the top left.

SAS Environment Manager is displayed.

3 Click 🕄.

The **Configuration** page is displayed.

4 Type compute into the service filter.

The **Compute service** becomes available.

5 Click Compute service.

Configuration instances of the **Compute service** become available.

6 Type autoexec into the configuration instance filter.

The sas.compute.server: autoexec_code configuration instance becomes available.

- 7 Click r to edit the **sas.compute.server: autoexec_code** configuration instance.
- 8 Add the ORC LIBNAME statement to the contents area for sas.compute.server: autoexec_code. See "Example: Read an ORC Table" in SAS Viya Engine for ORC: Reference.
- 9 Click Save to save your changes.

Compute Libraries Using PCFILES Engine

A discovery agent cannot crawl a compute library that uses the PCFILES engine unless you add a LIBNAME statement for that library in the **sas.compute.server: autoexec_code** configuration instance of the **Compute service**. You can then select this library to be crawled by a discovery agent.

- 1 Identify or create a compute library that uses the PCFILES engine. See "Preparing SAS Compute Libraries" on page 6.
- 2 Note the LIBNAME, SERVER, PATH, and other attributes specified in the library.
- 3 Select Manage Environment from the application menu at the top left.

SAS Environment Manager is displayed.

4 Click 🔍.

The Configuration page is displayed.

5 Type compute into the service filter.

The Compute service becomes available.

6 Click Compute service.

Configuration instances of the **Compute service** become available.

7 Type autoexec into the configuration instance filter.

The **sas.compute.server: autoexec_code** configuration instance becomes available.

- 8 Click r to edit the **sas.compute.server: autoexec_code** configuration instance.
- 9 Add statements such as the following to the contents area for sas.compute.server: autoexec_code:

LIBNAME myxlsx PCFILES Server="D2323" PATH="c:\demo.xlsx" PORT=9621;

10 Click **Save** to save your changes.

Compute Libraries for SAP ASE

You cannot use the library creation window in SAS Studio to create a SAS compute library that will access SAP Adaptive Server Enterprise (SAP ASE). You can add an SAPASE LIBNAME statement in the **sas.compute.server: autoexec_code** configuration instance of the **Compute service**. You can then select this library to be crawled by a discovery agent.

- 1 Identify the path and other attributes that you will need to specify an SAPASE LIBNAME statement. See "SAS/ACCESS Interface to SAP ASE" in SAS/ACCESS for Relational Databases: Reference.
- 2 Select Manage Environment from the application menu at the top left.

SAS Environment Manager is displayed.

3 Click 𝕄.

The Configuration page is displayed.

4 Type compute into the service filter.

The **Compute service** becomes available.

5 Click Compute service.

Configuration instances of the **Compute service** become available.

6 Type autoexec into the configuration instance filter.

The **sas.compute.server: autoexec_code** configuration instance becomes available.

- 7 Click Click to edit the sas.compute.server: autoexec_code configuration instance.
- 8 Add statements such as the following to the **contents** area for **sas.compute.server: autoexec_code**:

libname aselib sapase user="dmtest" password=XXXXXX server="server_name"
database="db name";

9 Click Save to save your changes.

Encoding Options for SAS Compute Libraries

If a discovery agent fails to crawl a compute library due to encoding errors, you can set the INENCODING=ANY and OUTENCODING=ANY options on the library. If these options are set, the agent ignores encoding errors and does the best it can for cross encoding issues. For more information about encoding options for compute libraries, see "INENCODING=, OUTENCODING= Options Statements" in SAS National Language Support (NLS): Reference Guide.

The library creation window in SAS Studio includes a **Transcoding Fail Behavior** option. You can set this option to SILENT.

Maintain Discovery Agents

Add and Run Discovery Agents

Each discovery agent crawls one library. The agent ingests, analyzes, and enhances metadata from the assets in that library. To add and run discovery agents, perform the following tasks:

- Prepare a library for the assets that you want to include in the catalog. See "Prepare Libraries" on page 4.
- 2 Log in to SAS Viya as an administrator. Opt in to all assumable groups.

SAS Drive is displayed.

3 Select Discover Information Assets from the applications menu at the top left. Alternatively, click the Discover Information Assets tab. Then click the Discover Information Assets button.

SAS Information Catalog is displayed.

4 Click Cick <l

The **Discovery Agents** window is displayed.

If you do not see $\widehat{c_{2}}$ on the left, then you are not logged in as an administrator.

5 Click [*.

The Choose Libraries to Discover window is displayed. Only global libraries are shown. Global libraries that are already assigned to agents are not shown.

.....

Note: The library selection window filters out those libraries that are already in use so that you cannot select them a second time. This rule helps to prevent duplicate information from being loaded into the catalog.

6 Select the library for the assets that you want to index in the catalog. Click **New Discovery Agent**.

A window opens for the agent. The name of the library that you selected appears in the **Name** field and in the **Library name** field.

- 7 Enter a name for the agent in the Name field or accept the default name.
- 8 Enter a description of the library that is crawled by the agent.

This text is optional, but it can help you understand which library is being crawled by the agent.

9 Enter a physical region for the library this is crawled by the agent.

This value is optional, but it can be useful. **Physical region** is a searchable field. A search of the **Physical region** field could return all assets in a particular region, for example.

If you select an asset that is evaluated from a search and you open the properties for the asset, then the library for that asset is displayed on the right. If you click the name of the library, then the server, engine, and physical region are displayed.

- 10 If you have licensed SAS Information Governance, select the Discovery locale whose country and language is appropriate for the assets that are crawled by this agent. For example, if the assets contain names and addresses from the United States in English, select the United States (English) locale. For more information about this feature, see "Understanding Content Analysis" on page 13.
- 11 Click 🔳 to save your changes.
- 12 In most cases, agents should be run periodically on a schedule. See "Schedule Agent Jobs" on page 11.

To run the agent immediately, click Run now and click Continue when prompted.

A main job is created for the agent and starts running. The main job runs several other jobs that crawl the library, capture metadata about assets in that library, analyze the metadata, and generate statistics about the assets. All of this information is added to the catalog.

Sometimes a single run will not crawl all tables. Running the agent again picks up any tables that were not collected in the previous run.

13 The Job Status field for an agent indicates whether the job is running or is idle. To determine whether the job succeeded or failed, you must use SAS Environment Manager. See "Monitor Agent Jobs" on page 12.

Schedule Agent Jobs

In most cases, agents should be run periodically on a schedule. Use SAS Environment Manager to schedule a discovery agent.

1 Select Manage Environment from the applications menu at the top left.

SAS Environment Manager is displayed.

2 Click o on the left.

The Jobs and Flows page is displayed. The Monitoring tab is selected.

- **3** To schedule the main job for an agent, click the **Scheduling** tab of the **Jobs and Flows** window. The job for the agent has the same name as the agent.
- 4 Right-click the job for an agent and select Schedule.
- 5 Follow the steps in "Schedule Jobs" in SAS Environment Manager: User's Guide.

6 To determine whether the scheduled job succeeded or failed, you must use SAS Environment Manager. See "Monitor Agent Jobs" on page 12.

Monitor Agent Jobs

The **Job Status** field for an agent indicates that the job is running or is idle. To determine whether the job succeeded or failed, you must use SAS Environment Manager. The agent must have been run at least once.

1 Select Manage Environment from the applications menu at the top left.

SAS Environment Manager is displayed.

2 Click of on the left.

The Jobs and Flows page is displayed. The Monitoring tab is selected.

The main job for an agent has the same name as the agent. The main job runs other jobs that crawl the library, capture metadata about assets in that library, analyze the metadata, and generate statistics about the assets. The following figure shows the jobs for an agent that is named COMMON_TARG.

J	Jobs and Flows			
Monitor (45)				
Filter by name 🔎 📃	E			
Name	Start Date	Status		
Analyze COMMON_TARG connect	January 19, 2	⊘ Successful		
Crawl COMMON_TARG connection	January 19, 2	⊘ Successful		
COMMON_TARG	January 19, 2	⊘ Successful		

COMMON_TARG is the main job for the agent. **Crawl COMMON_TARG connection** is the job that crawls the library and captures metadata about assets in that library. **Analyze COMMON_TARG connection** is the job that analyzes metadata and generates statistics for assets in the library. If there are many assets in a library, the main job launches multiple **Analyze** jobs.

You can check the job logs to see the status of runs. The agents use SAS code to index the information, and the log represents information about the SAS code that is used in the agents.

For general information about monitoring jobs, see "Monitor Jobs" in SAS Environment Manager: User's Guide.

Usage Notes for Discovery Agents

Running Multiple Discovery Agents Simultaneously

For best performance, do not run many discovery agents simultaneously unless you have a highavailability execution environment.

Understanding Content Analysis

Overview

If you have licensed SAS Information Governance, discovery agents can analyze the names and content of columns in assets. If the analysis can determine what type of content is in a column, then a semantic type is added to the metadata for that column in the catalog. A column that contains last names might get a semantic type of **Family Name**, for example. Semantic types can help you determine whether a table contains useful information for your purposes.

You can view semantic types for the columns in an asset by viewing the asset's details, and then selecting **Column Analysis** ⇒ **Data Quality Measures**. The next figure shows semantic types for columns in a table.

🛙 Overview 🛛 🗐 Colum	nn Analysis	🕏 Sampi	e Data	
Descriptive Measures Meta		Measures	Data Quality Measures	
Name		Semantic Type		
♣ Company		ORGANIZATION		
≜ _# Contact		INDIVIDUAL		
♠ Address		DELIVERY ADDRESS		
€ ∎ Address2		ADDRESS		

Figure 1 Semantic Types from Content Analysis

Discovery Agents and Discovery Locales

If you have licensed SAS Information Governance, you can select the **Discovery locale** whose country and language is appropriate for the assets that are crawled by the agent. The agent uses the discovery locale to perform content analysis on the names and content of columns in assets. Select the discovery locale whose country and language is appropriate for the assets that are crawled by this agent. For example, if the assets contain names and addresses from Canada in French, select the **Canada (French)** locale.

When you log in to SAS Information Catalog, you connect to a CAS server, and that server has a default locale. You can select a different locale as appropriate for the assets that are crawled by this agent.

Semantic Types and SAS Quality Knowledge Base (QKB)

The locale that you specify for a discovery agent is a SAS Quality Knowledge Base (QKB) locale. A locale is the combination of a language and a country or region. Definitions in the QKB are designed to process data rendered in different languages from different geographical regions.

Semantic types in SAS Information Catalog are derived from the Field Content and Field Name Identification Analysis definitions for a QKB locale. Not all locales support both field name and field content analysis, as shown in the next table.

Country	Language	QKB Locale	Field Content Analysis	Field Name Analysis
Argentina	Spanish	ESARG		Field name
Australia	English	ENAUS		for all locales.
Belgium	Dutch	NLBEL	Field content Semantic Types for the Dutch, Belgium locale.	
Belgium	French	FRBEL	Same semantic types as the Dutch, Belgium locale.	
Canada	English	ENCAN	Field content Semantic Types for the English, Canada locale.	
Canada	French	FRCAN	Same semantic types as the	

Table 2Semantic Types by Locale

Country	Language	QKB Locale	Field Content Analysis	Field Name Analysis
			English, Canada locale.	
China	Chinese	ZHCHN		
Denmark	Danish	DADNK	Field content Semantic Types for the Danish, Denmark locale.	
France	French	FRFRA	Field content Semantic Types for the French, France locale.	
Germany	German	DEDEU	Field content Semantic Types for the German, Germany locale.	
Hong Kong	English	ENHKG		
India	English	ENIND		
Israel	Hebrew	HEISR		
Italy	Italian	ΙΤΙΤΑ	Field content Semantic Types for the Italian, Italy locale.	
Japan	Japanese	JAJPN		
Mexico	Spanish	ESMEX		
New Zealand	English	ENNZL		
Netherlands	Dutch	NLNLD		
Norway	Norwegian	NONOR	Field content Semantic Types for the Norwegian, Norway locale.	
Philippines	English	ENPHL		
Singapore	English	ENSGP		
South Africa	English	ENZAF		
Spain	Spanish	ESESP		
United Kingdom	English	ENGBR	Field content Semantic Types	

Country	Language	QKB Locale	Field Content Analysis	Field Name Analysis
			for the English, United Kingdom locale.	
United States	English	ENUSA	Field content Semantic Types for the English, United States locale.	
Sweden	Swedish	SVSWE	Field content Semantic Types for the Swedish, Sweden locale.	

A QKB is typically installed with the rest of your SAS software. A QKB can also be installed later, as described in "Set the Default QKB and the Default Locale" in SAS Viya: QKB Management. You can customize a QKB by adding or modifying its definitions and rules. For more information, see "Customize a QKB" in SAS Viya: QKB Management.

Assign a Review Status to an Asset

If you license SAS Information Governance, an administrator can use the **Status** property to indicate that an asset is under review, is approved, has been flagged for attention, or has a warning. If the **Status** property has been consistently applied, you can use syntax search to return all assets with a given review status. See "Syntax Search" in SAS Information Catalog: User's Guide.

To assign a review status to an asset, perform the following steps:

- 1 Log in to SAS Viya as an administrator and display SAS Information Catalog.
- 2 Search for the asset to which you will assign a review status.
- **3** Open the properties for that asset.
- 4 Access the menu for the **Status** field and select an appropriate review status.

Respond to Notifications

If you log in to SAS Information Catalog as an administrator, you will receive push notifications of errors and other conditions. These conditions require a response.

Check for notifications in the user profile area at the top right. If you have any notifications, a number is displayed next to your user name. Click the number to display notifications. The next table describes notifications and responses that you can take.

Notification	Re	sponse	Details
The search index has not been updated with the most recent catalog metadata. Run the job Rebuild catalog index "datasets" from SAS Environment Manager to fix this problem.	1	Log in to SAS Viya as an administrator.	Elasticsearch can get out of sync with SAS Information Catalog if the container goes down or
	2	Select Manage Environment from the application menu.	was updated in some way. SAS Information Catalog automatically detects this problem and alerts the administrator through this notification. Take the actions at the left in the Response column to ensure that you can search catalog
	3	Select Jobs and Flows.	content correctly.
	4	Click the Scheduling tab.	
	5	Search for "Rebuild." The Rebuild catalog index "datasets" job becomes available.	
	6	Run the Rebuild catalog index "datasets" job.	

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies. Copyright © 2021, SAS Institute Inc., Cary, NC, USA. All Rights Reserved. April 2021 v_001-P1:infocatag

