SAS Information Catalog

workshop exercises

Patric Hamilton

Cecily Hoffritz

Data Engineering Advisory

Nordics

Contents

[1. Investigating the SAS Information Catalog user interface. 3](#_Toc94163873)

[2. Creating a discovery agent for Public CAS library 6](#_Toc94163874)

[3. Creating a Discovery Agent for SASHELP with inclusion rules 8](#_Toc94163875)

[4. Discovering assets in SAS Information Catalog part 1 11](#_Toc94163876)

[5. Discovering assets in SAS Information Catalog part 2 13](#_Toc94163877)

[6. Discovering assets in SAS Information Catalog part 3 15](#_Toc94163878)

[Next Steps 17](#_Toc94163879)

[7. Preparing CAS libraries for information assets 18](#_Toc94163880)

[8. Preparing SAS Compute libraries for information assets 21](#_Toc94163881)

[9. Creating discovery agents for WORKSHOP and CASHOME libraries 23](#_Toc94163882)

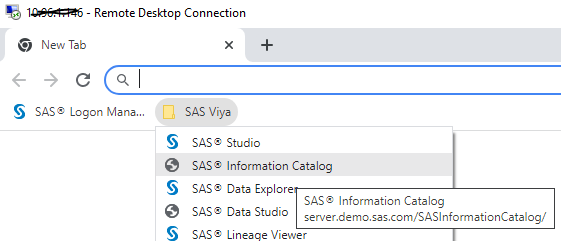
[10. Monitoring and scheduling discovery agent jobs 24](#_Toc94163883)

[11. Case - Discovering assets in SAS Information Catalog 28](#_Toc94163884)

# Investigating the SAS Information Catalog user interface.

**The purpose of this exercise is to familiarize yourself with the user interface of SAS Information Catalog.**

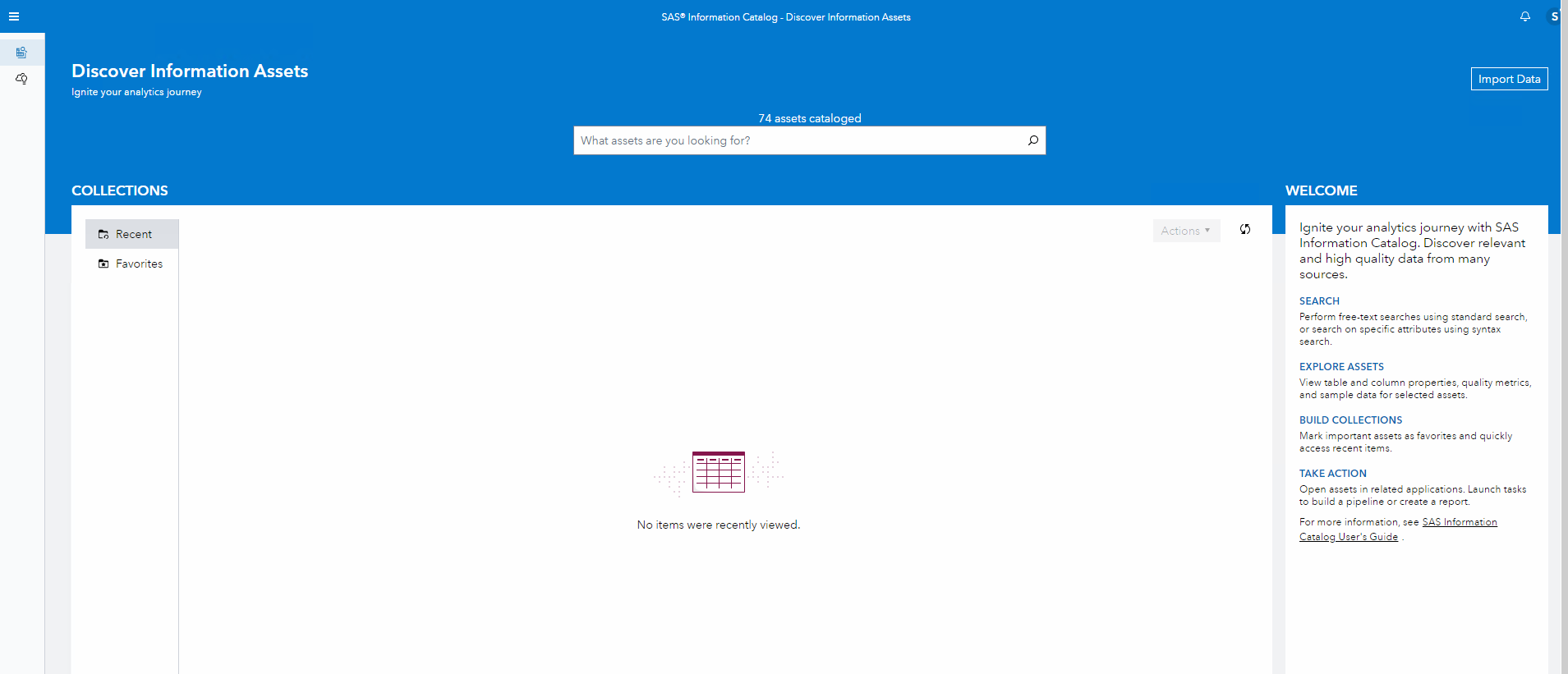
1. Make a remote desktop connection as specified in your reservation RACE mail.
2. Select Google Chrome on the remote desktop.
3. In the SAS Viya shortcut menu, select SAS Information Catalog and sign in as **sasdemo** with the password **Orion123**.



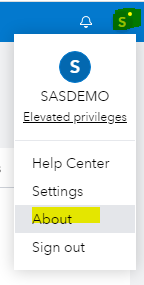
1. Do the following in SAS Information Catalog home page:
   1. Notice, no items were recently viewed.
   2. Expand the left pane by clicking on >> (bottom of left pane). You see that the current menu is Catalog.
   3. Make a note of the number of assets that are currently catalogued for this environment (number is seen just above the search field).
   4. Read the WELCOME box to the right to familiarize yourself with tasks that you can perform in this application.
   5. Click Import Data button (upper right).

What application is started if you were to continue with the import (which you won’t be doing at this point)? Click Cancel to exit Import Data.

* 1. Notice that assets can be favorites. Your favorites list will be empty if you haven’t been working with favorites before. You will be adding your favorites later.



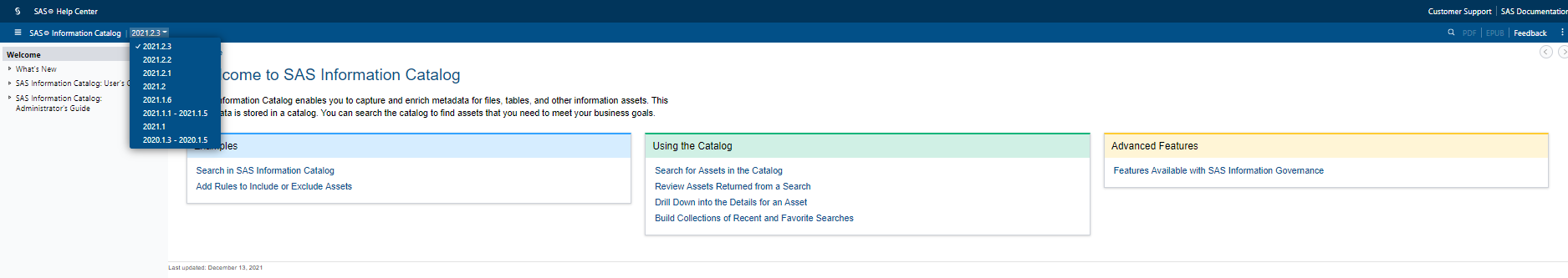
1. Click the sign in icon (upper right) and select About.



1. Make a note of the SAS Viya version.
2. Click on the link to access the documentation for SAS Information Catalog.

[SAS Help Center: Welcome to SAS Information Catalog](https://go.documentation.sas.com/doc/en/infocatcdc/v_008/infocatwlcm/titlepage.htm?homeOnFail)

1. Click the drop down list for the versions. What is the latest stable release in the documentation and how does this compare to your current RACE image?

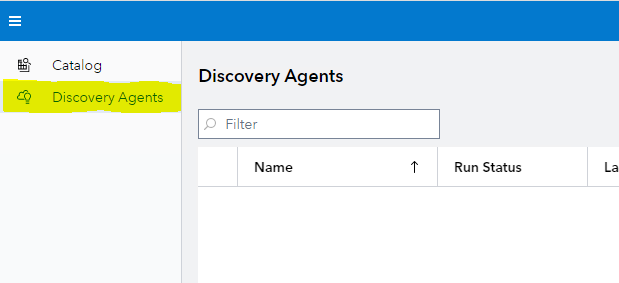


1. Click on the very latest release in the drop down menu and select What’s newàStable Releases where you can see new features for each new release.



What are the new features in the January release?

1. In SAS Information Catalog, click the **Discovery Agents** menu in the left pane.

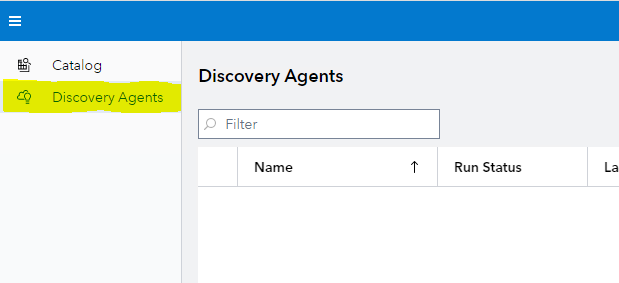


Are there any existing discovery agents?

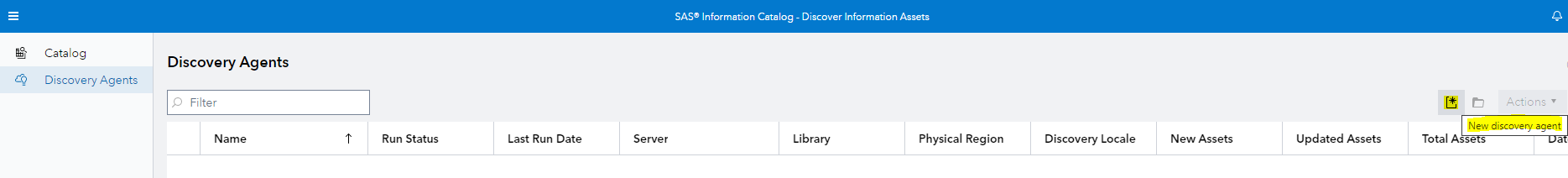
# ThCreating a discovery agent for Public CAS library

**The purpose of this exercise is to create a new SAS Information Catalog Discovery Agent that crawls and analyzes content in the Public CAS library.**

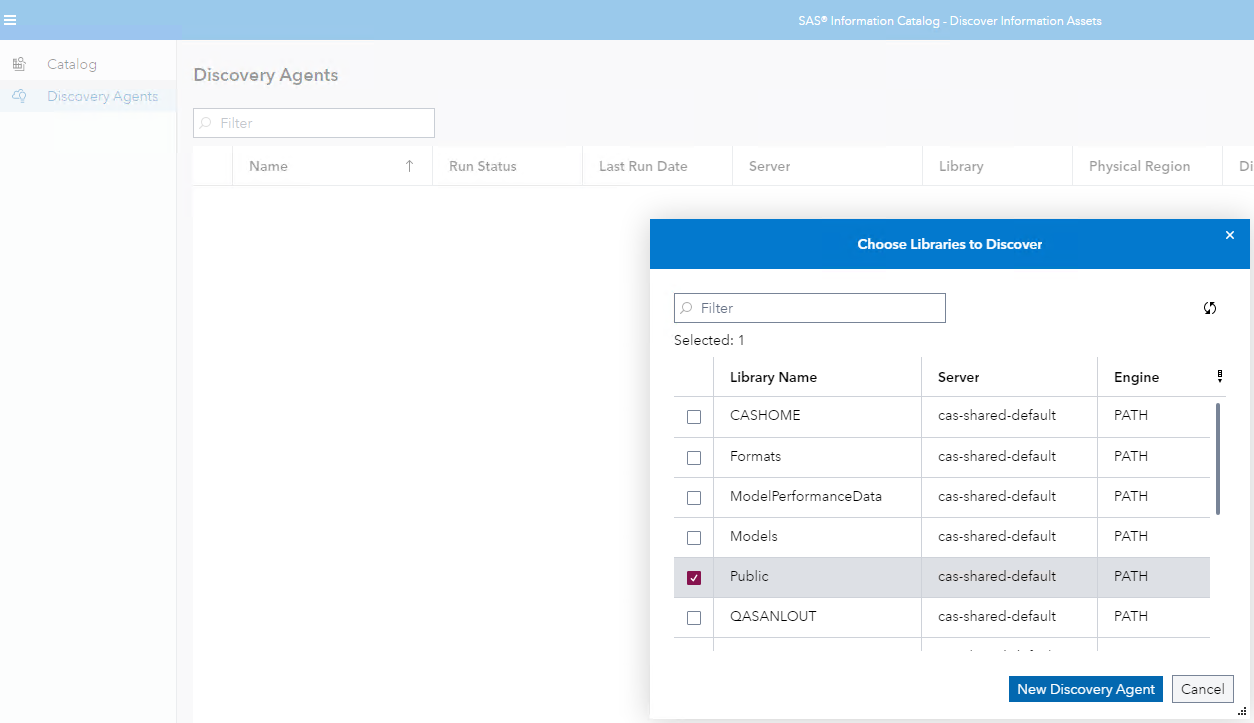
1. In SAS Information Catalog, click the **Discovery Agents** menu in the left pane.



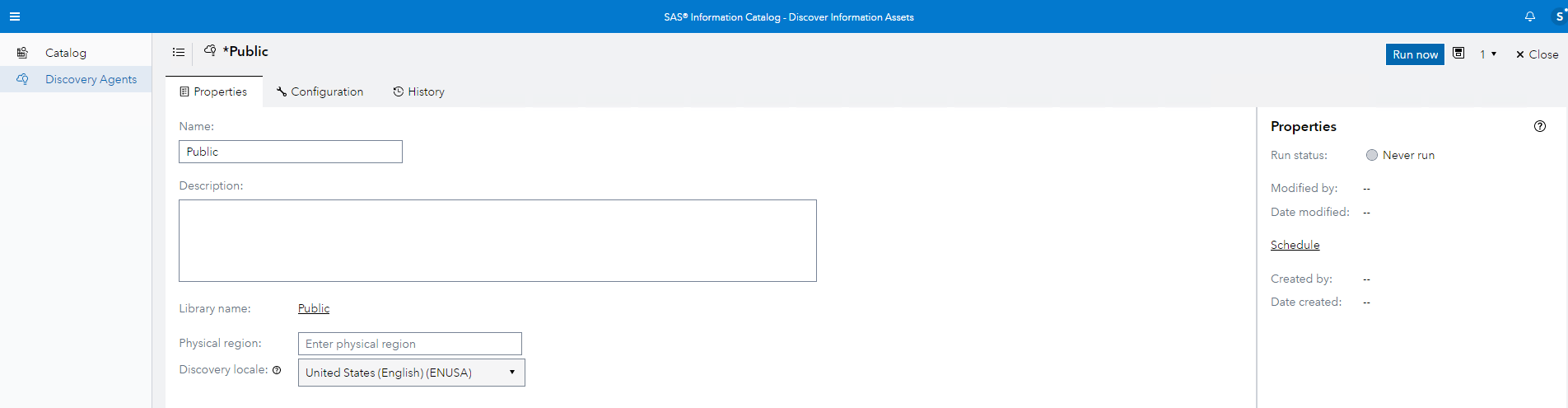
1. Click the New discovery agent icon.



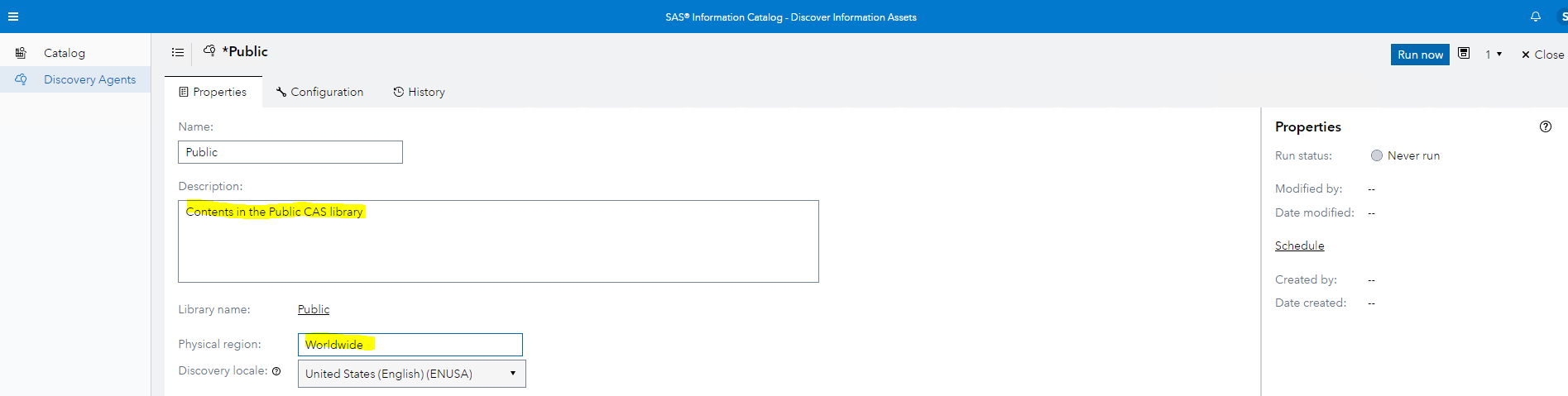
1. Select the Public library



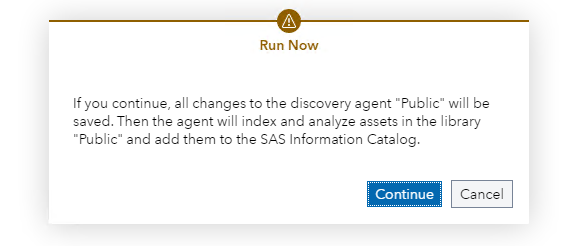
1. Click New Discovery Agent button. You land in the Properties tab.



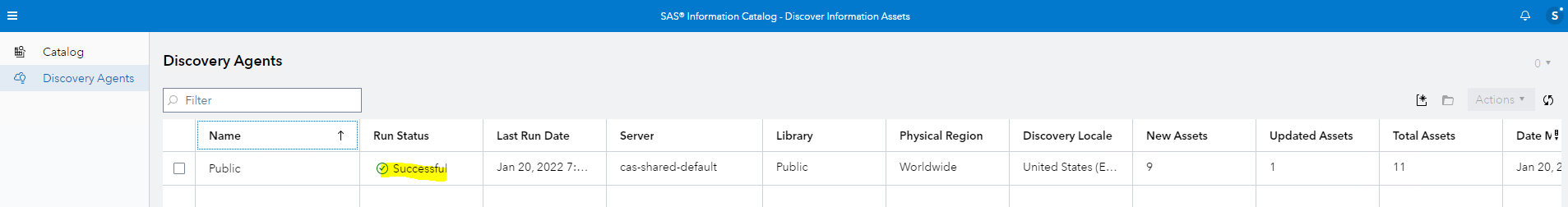
1. Do the following in the Properties tab.
   1. Add a Description: Stable 2021.2.1 Stable 2021.2.1 Stable 2021.2.1
   2. Add a physical region: Worldwide.



1. Click Run now. You get the message below that your discovery agent will be saved.



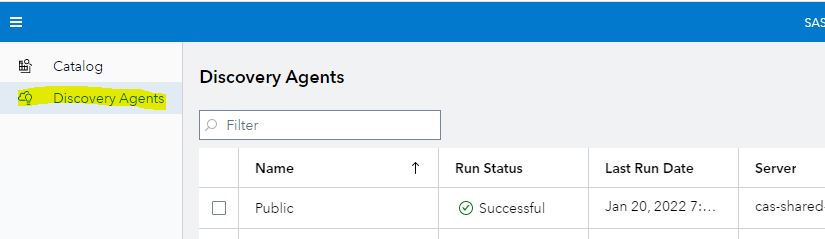
1. Click Continue. Your discovery agent is saved and it starts to run. It can take a while on the RACE image to complete.



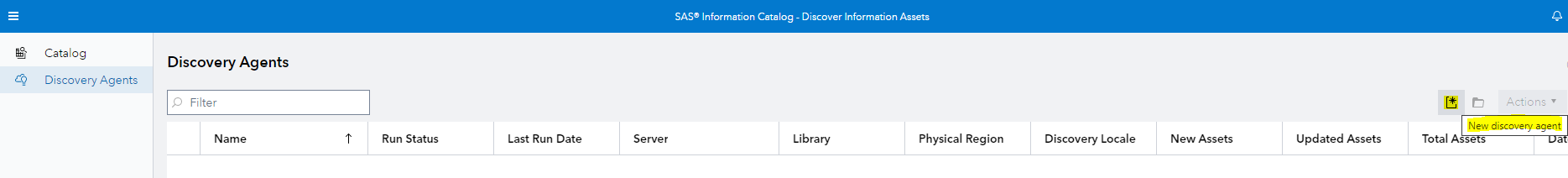
# Creating a Discovery Agent for SASHELP with inclusion rules

**The purpose of this exercise is to create a discovery agent on a SAS Compute library and apply inclusion rules.**

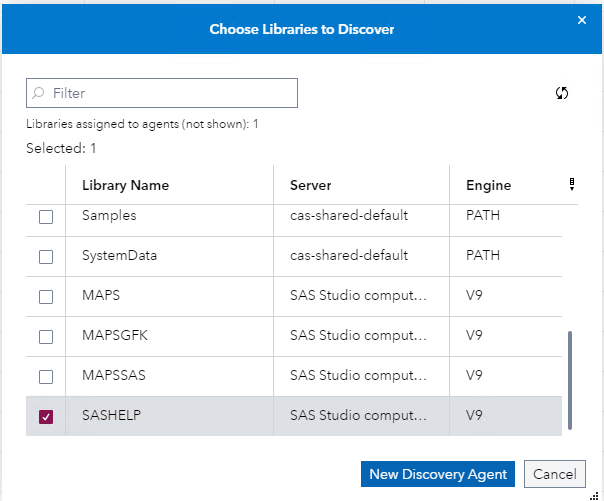
1. In SAS Information Catalog, click the **Discovery Agents** menu in the left pane if you are currently not in this menu.



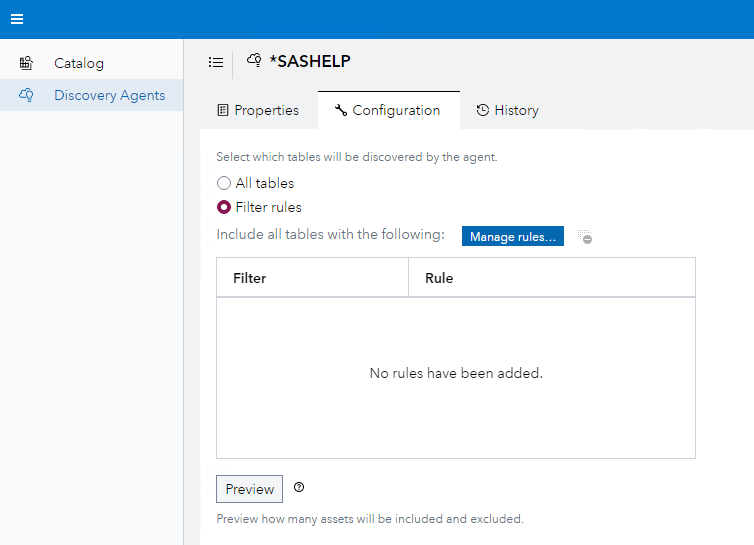
1. Click the New discovery agent icon.



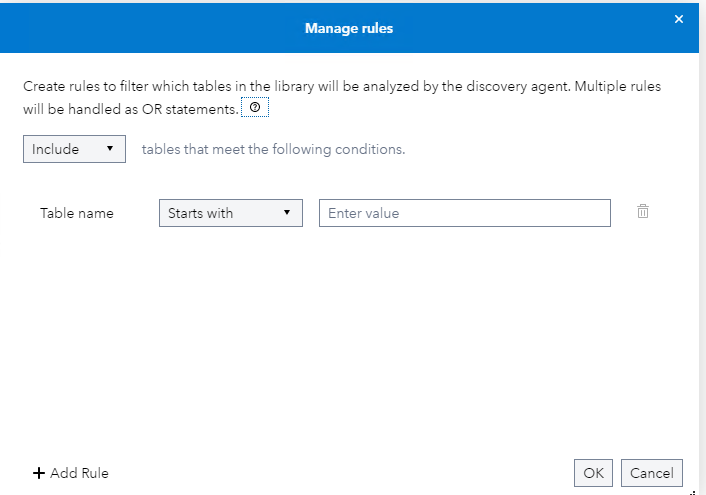
1. Select the SASHELP library.



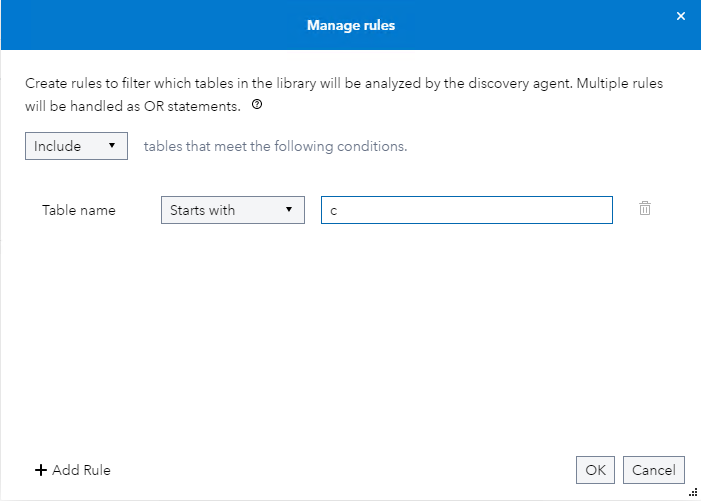
1. Click the New Discovery Agent button. You land back into the Properties tab.
2. Add a description and physical region. For example your description could be **SASHELP sample data** and your region, **United States**.
3. Select the Configuration tab and click Filter rules and then Manage rules…



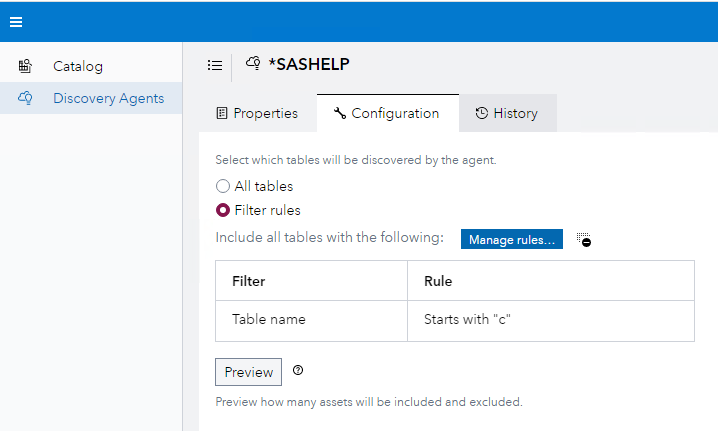
You land in the Manage Rules page.



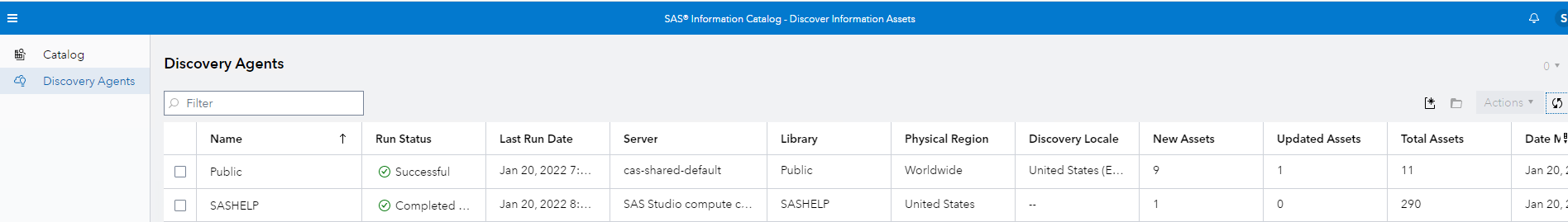
1. Set up a rule where you include all tables starting with c.



1. Select OK and your new rule appears.



1. Click Run now and Continue to save and execute the discovery agent for SASHELP.
2. Compare your total assets for SASHELP with the total assets below. Add your number to the Teams chat so that others can compare their results to yours.



# Discovering assets in SAS Information Catalog part 1

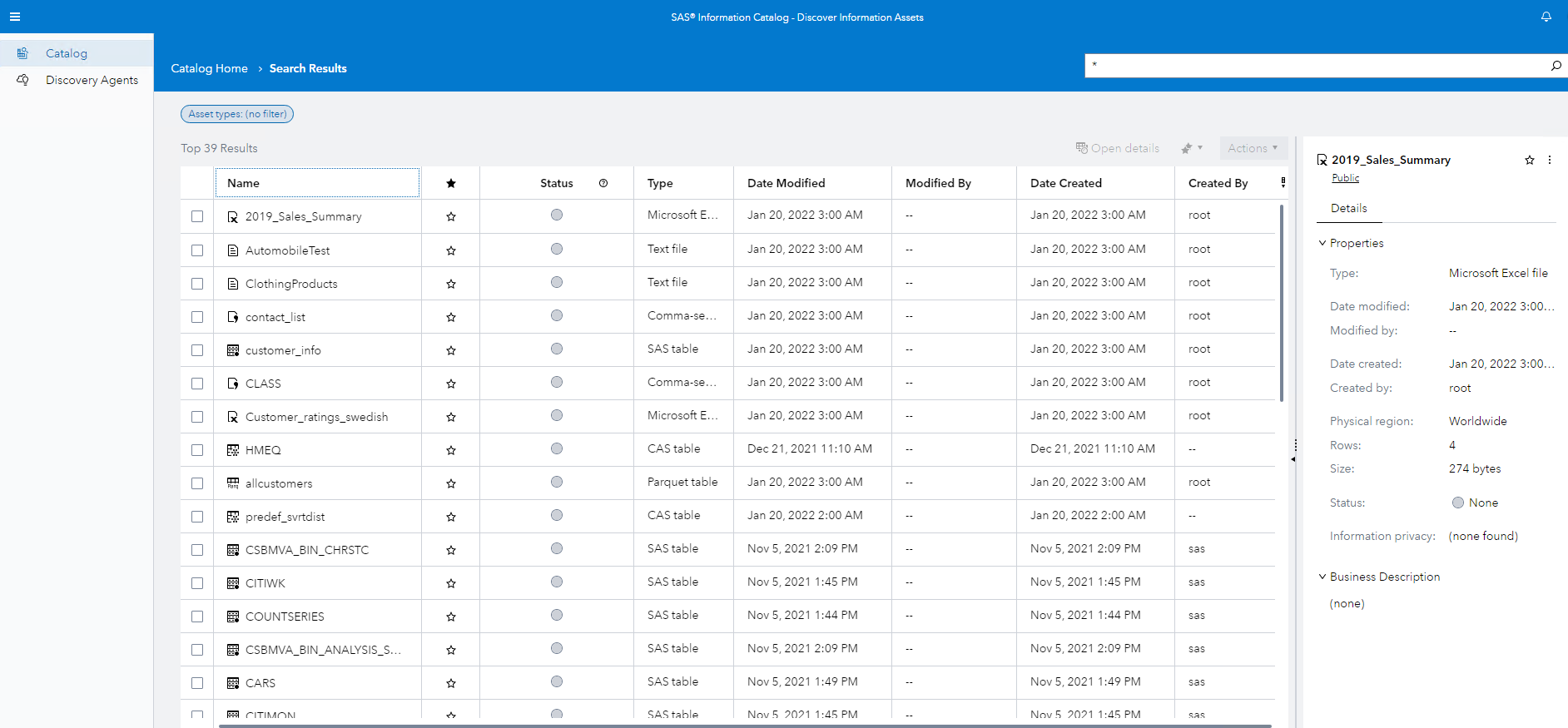
**The purpose of this exercise is taking the first step to understand your options to search assets.**

**If not already signed in, sign into SAS Information Catalog as sasdemo with the password Orion123.**

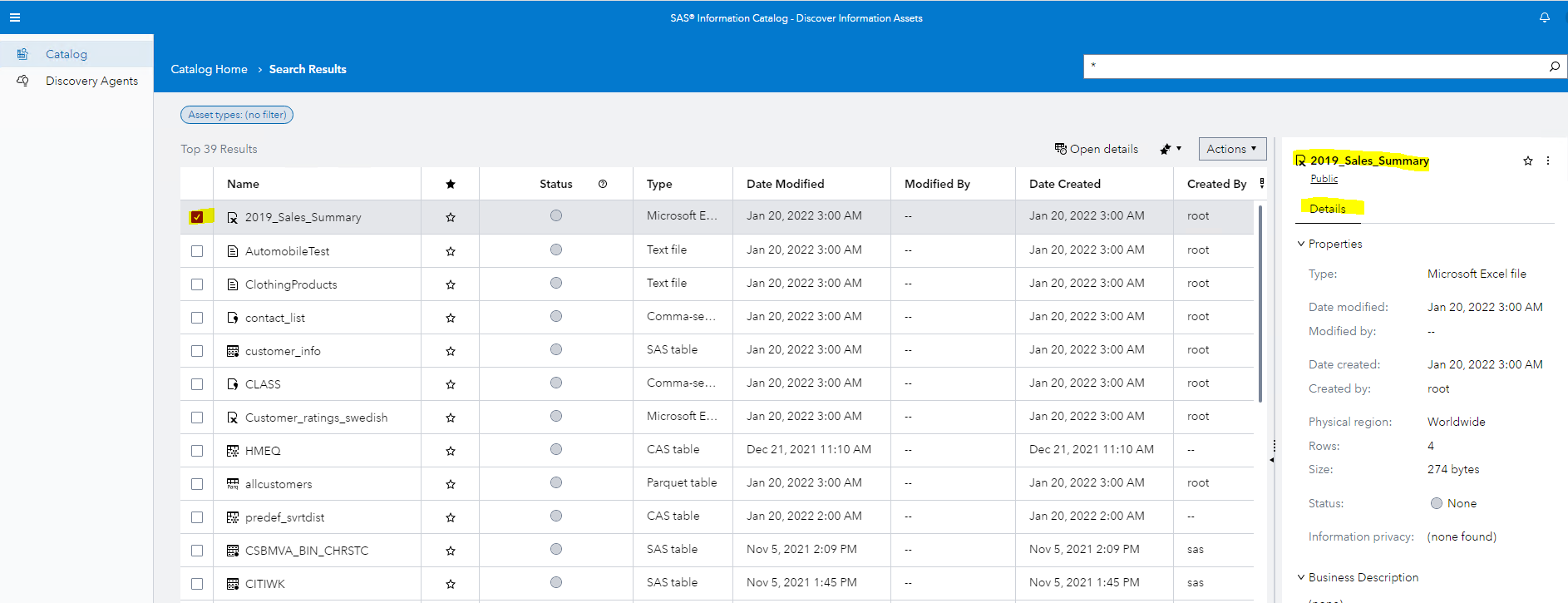
1. In SAS Information Catalog switch to the Catalog menu.



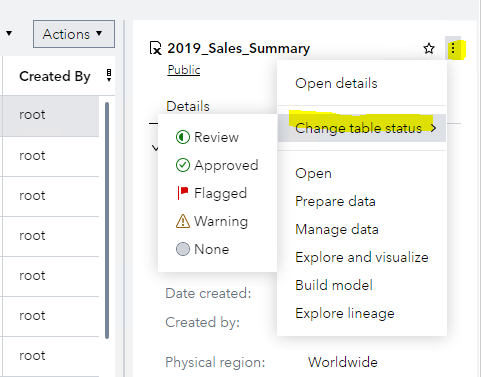
1. Add an asterix \* in the Search field and press Enter. You get a long list of assets.



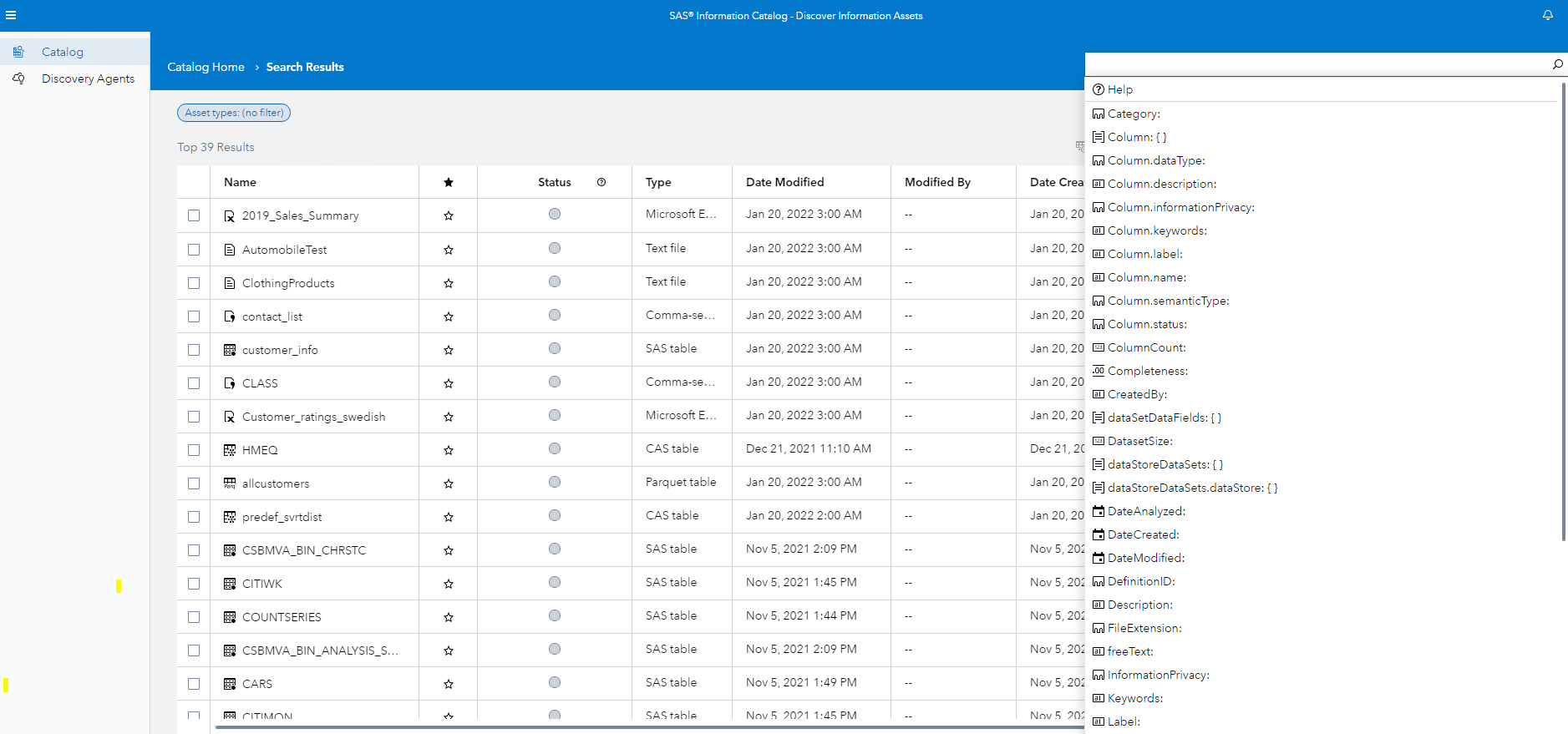
1. Notice, that not all assets are data assets.
2. Scroll down in the list and notice that previously defined SAS Visual Analytics reports, models etc. have been automatically cataloged, without having to create a discovery agent. Assets that are indexed automatically are not analyzed by discovery agents.
3. Notice, that if you click the tick mark for an asset in the left pane, then its details are visible in the right pane.



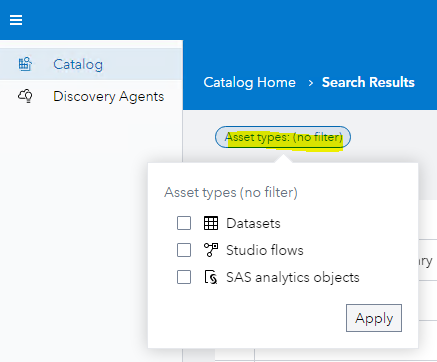
1. Notice, that you can click the snowman (3 dots on top of each other) in Details to open a menu. Without having to open a table first, you can change the table status. This is handy when you have many tables.



1. Remove the asterix in the Search field and press Spacebar. You get a list of facets to search. Facets are extremely helpful and make your searches more precise.



1. Click the Asset types menu to see your options of asset types to filter.

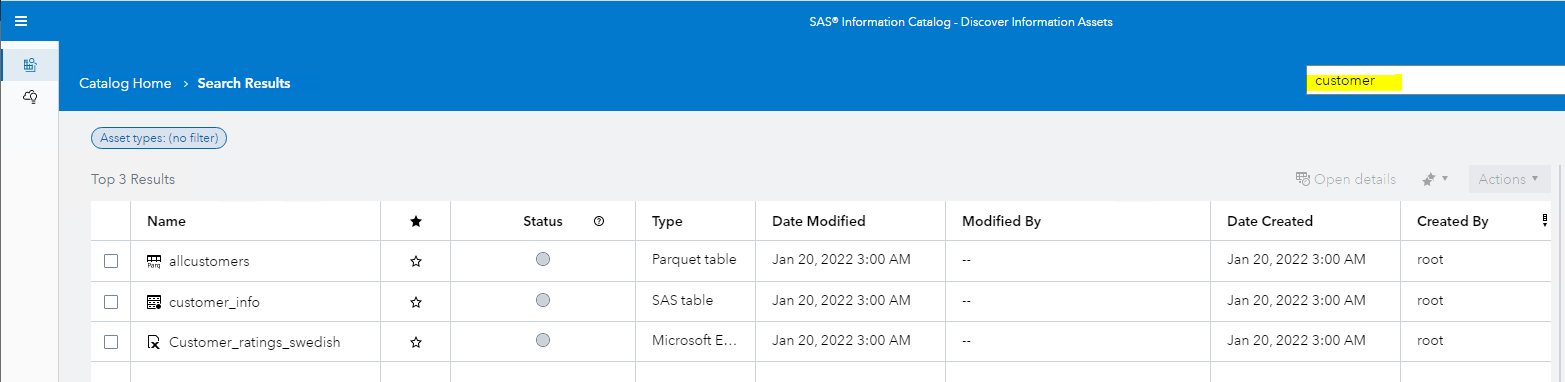


# Discovering assets in SAS Information Catalog part 2

**The purpose of this exercise is digging deeper to understand your options to search assets.**

**If not already signed in, sign into SAS Information Catalog as sasdemo with the password Orion123.**

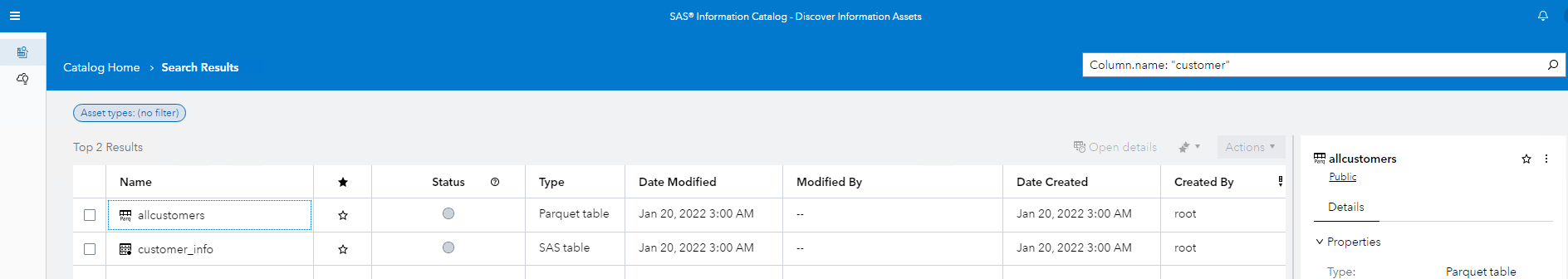
1. Perform a free-text search where you search assets with **customer**.



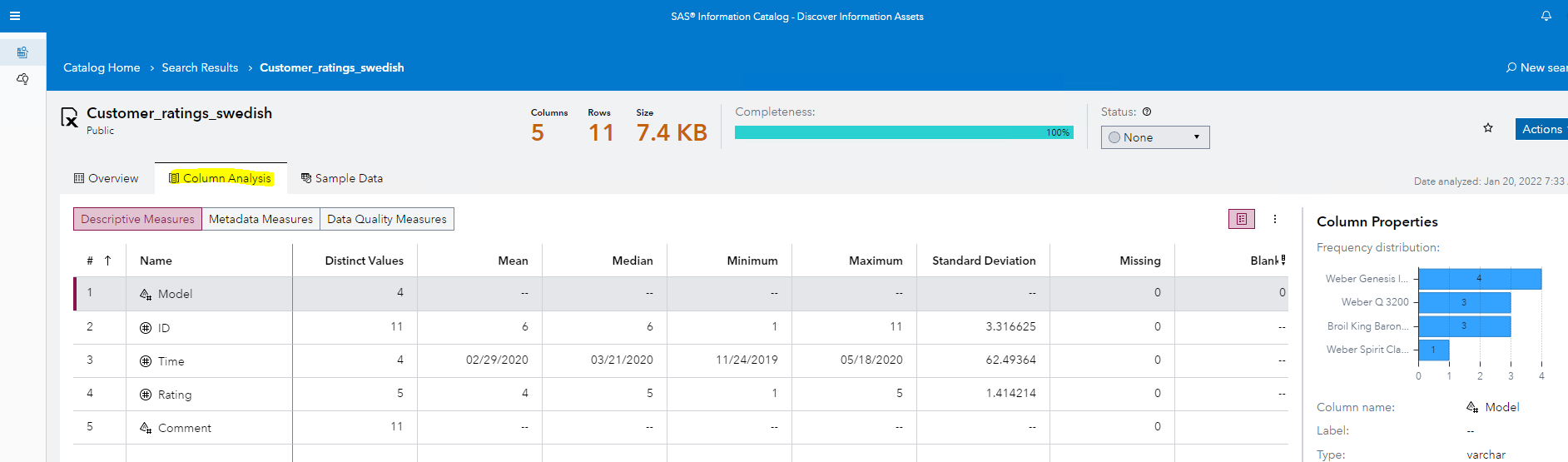
1. Click the star to the right of customer\_info to make the table a favorite. The star becomes black.



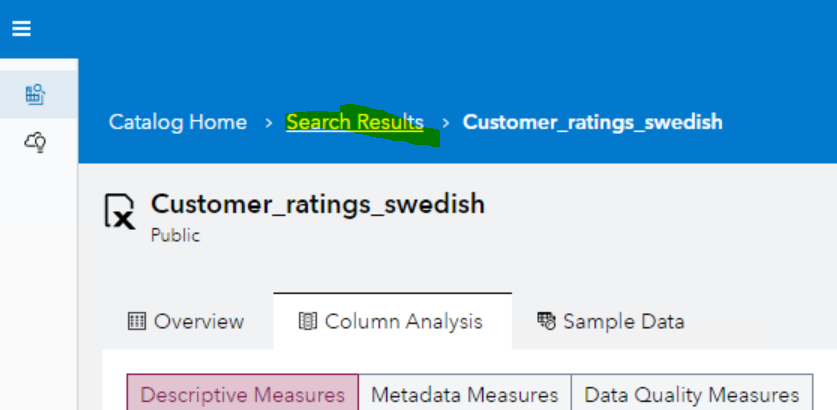
1. Perform a facet search where you search assets with customer as part of **column name**. Remember, you can press Enter to get the Facet list where you find the Column name facet. Notice that the Excel file with Swedish customer ratings disappears when you use the facet.



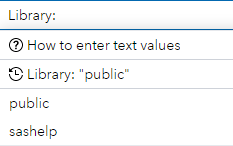
1. Revert to the free-text search with **customer** so that the Excel file with Swedish customer ratings returns to the list
2. Double click to open the Excel file with customer ratings, and in the Column Analysis tab, verify that there is no column with the text customer in its name.



1. Go back to your search results by clicking Search Results in the blue area above.



1. Do the following:
   1. Find assets that are comma separated files in the Public library. Use the **Library** facet.

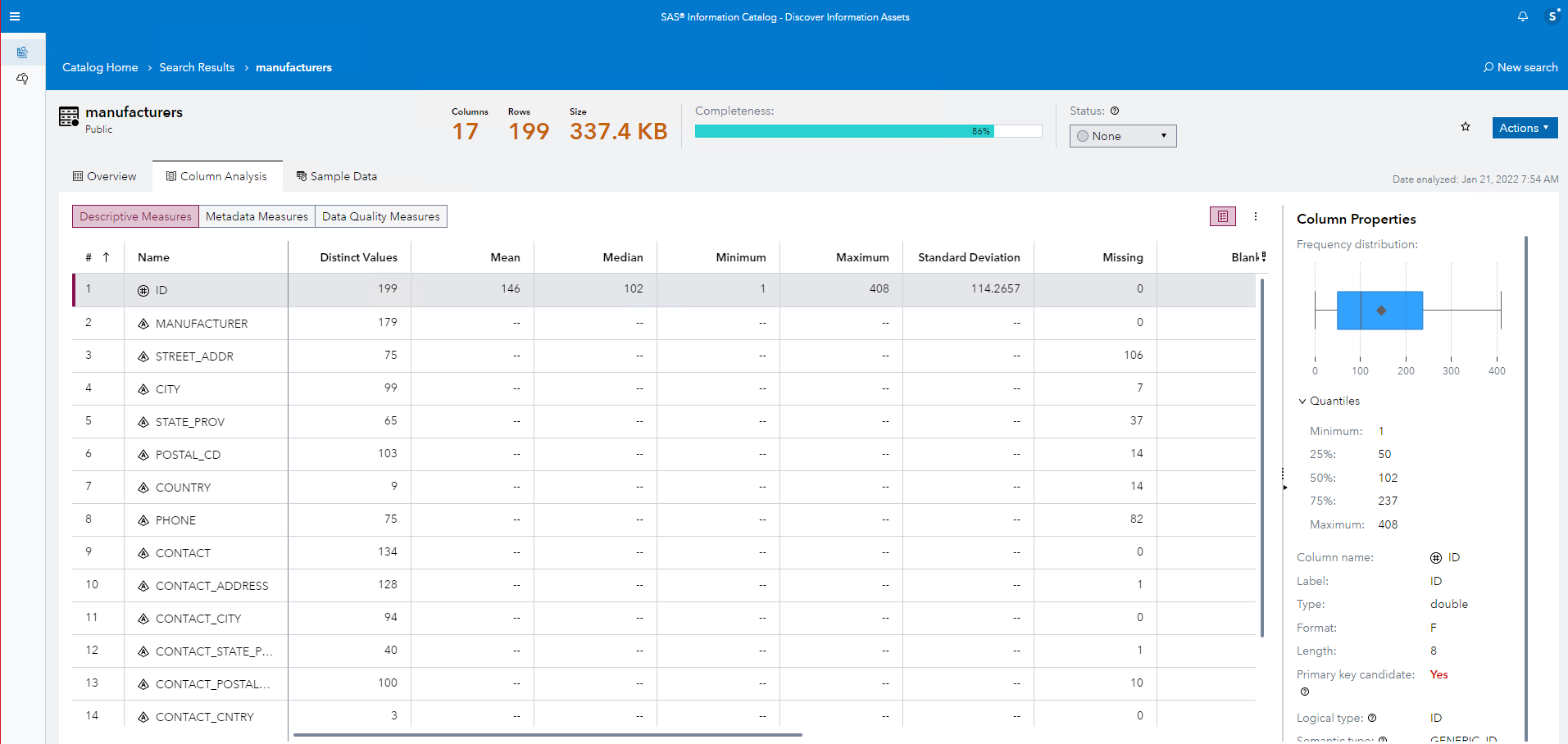


* 1. Find assets with privacy data. Use the **Information Privacy** facet.
  2. Find assets that are Excel files. Use the **sourceSystem** facet.
  3. Are there any tables that have been approved? Use the **Status** facet.
  4. Find assets in the region **worldwide.** Use the **Region** facet.
     1. Open **allcustomers.parquet**
     2. In the Overview tab, add the description **This data is used in Snowflake**.
     3. Write **snowflake** in the search field. Your table should appear.
     4. Use the **Description** facet to find the table that you previously added a description to.

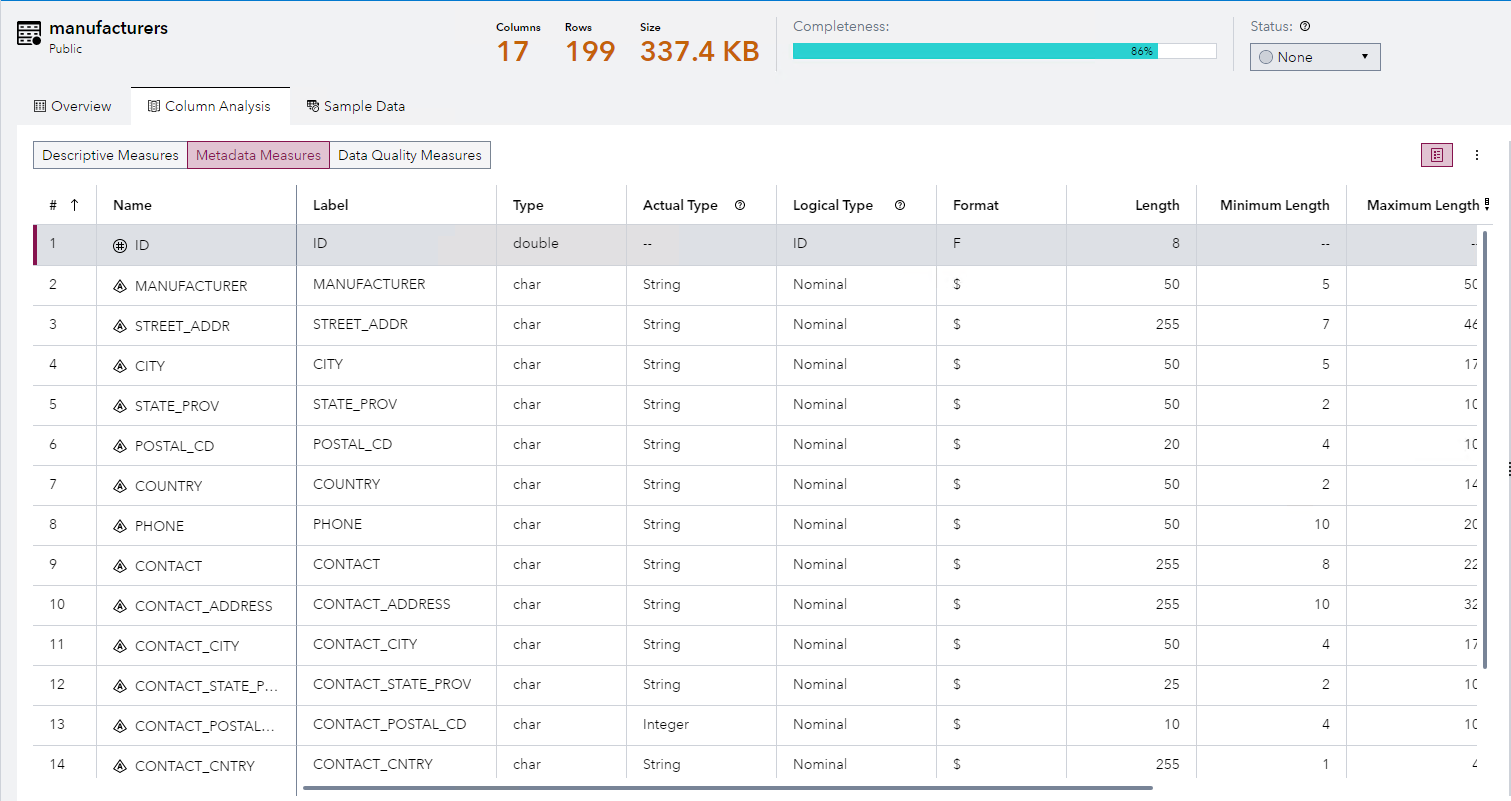
# Discovering assets in SAS Information Catalog part 3

**The purpose of this exercise is to analyze tables and columns to ascertain that they are fit for purpose for analytics. You will also investigate tables that contain privacy data.**

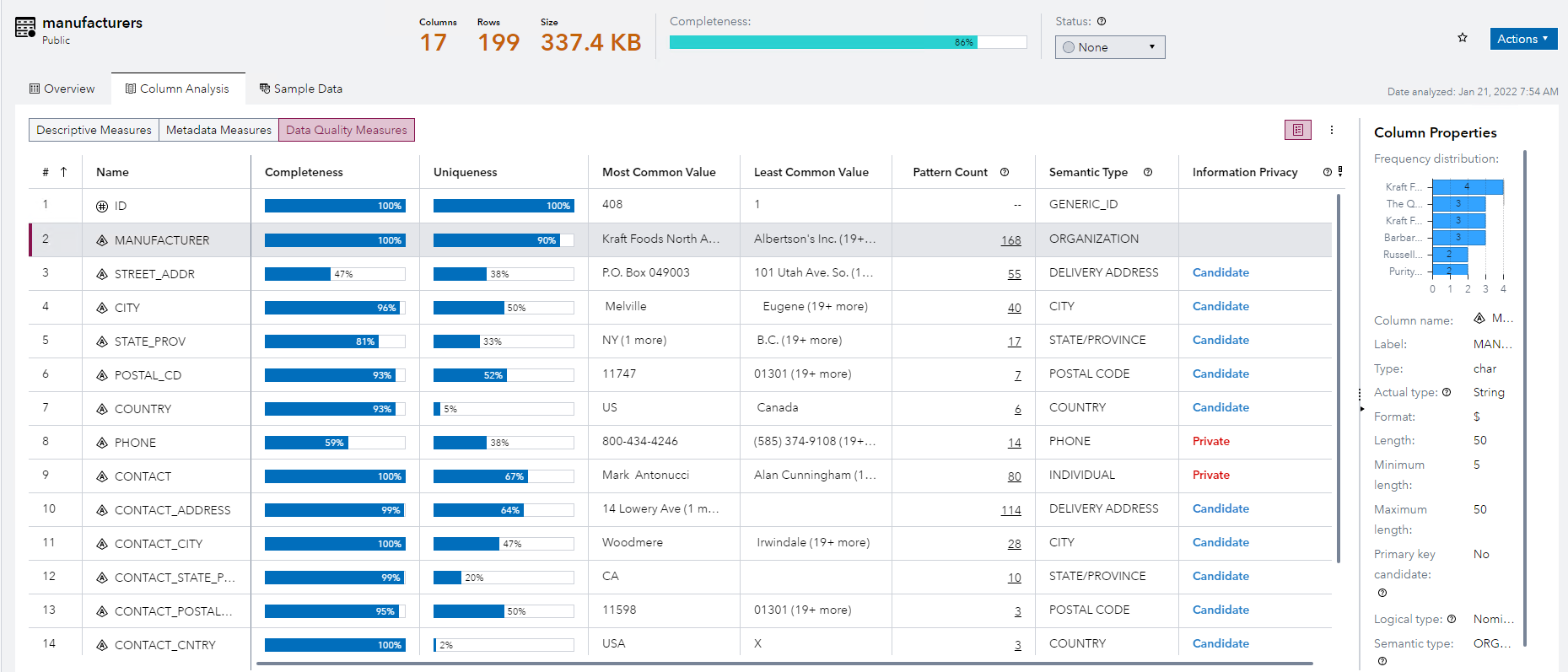
1. Use the **Name** facet to find the SAS table **manufacturers** and open it.
2. Select the **Column Analysis** tab. Notice, that
   1. There are three types of measures that you can investigate: Descriptive, Metadata, Data Quality.
   2. You are currently viewing content in **Descriptive Measures**.
   3. Descriptive measures are about Missing, Blanks, Outliers, Skewness, and other statistics.
   4. If you click a text column or a numeric column, **Column Properties** will show different statistics.



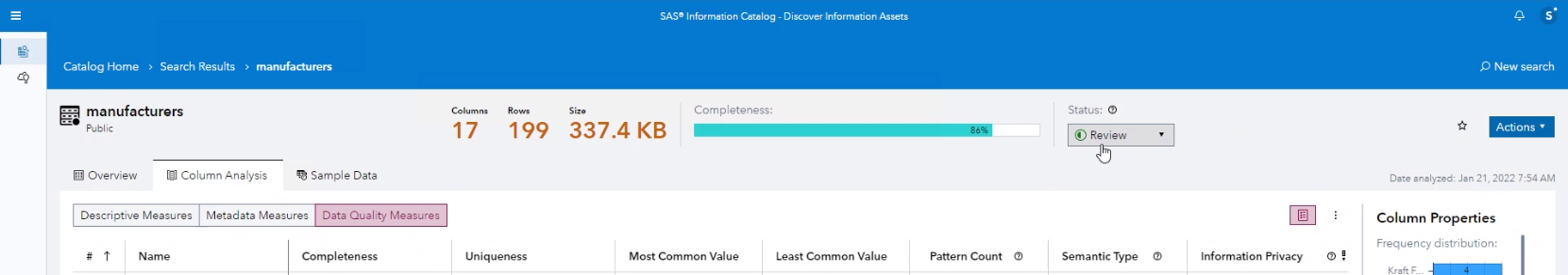
1. Click **Metadata Measures**. Notice, that
   1. Information is about column attributes such as Name, label, format length.
   2. There is Type, Actual Type and Logical Type where the last two have a question mark beside them to explain their meaning.



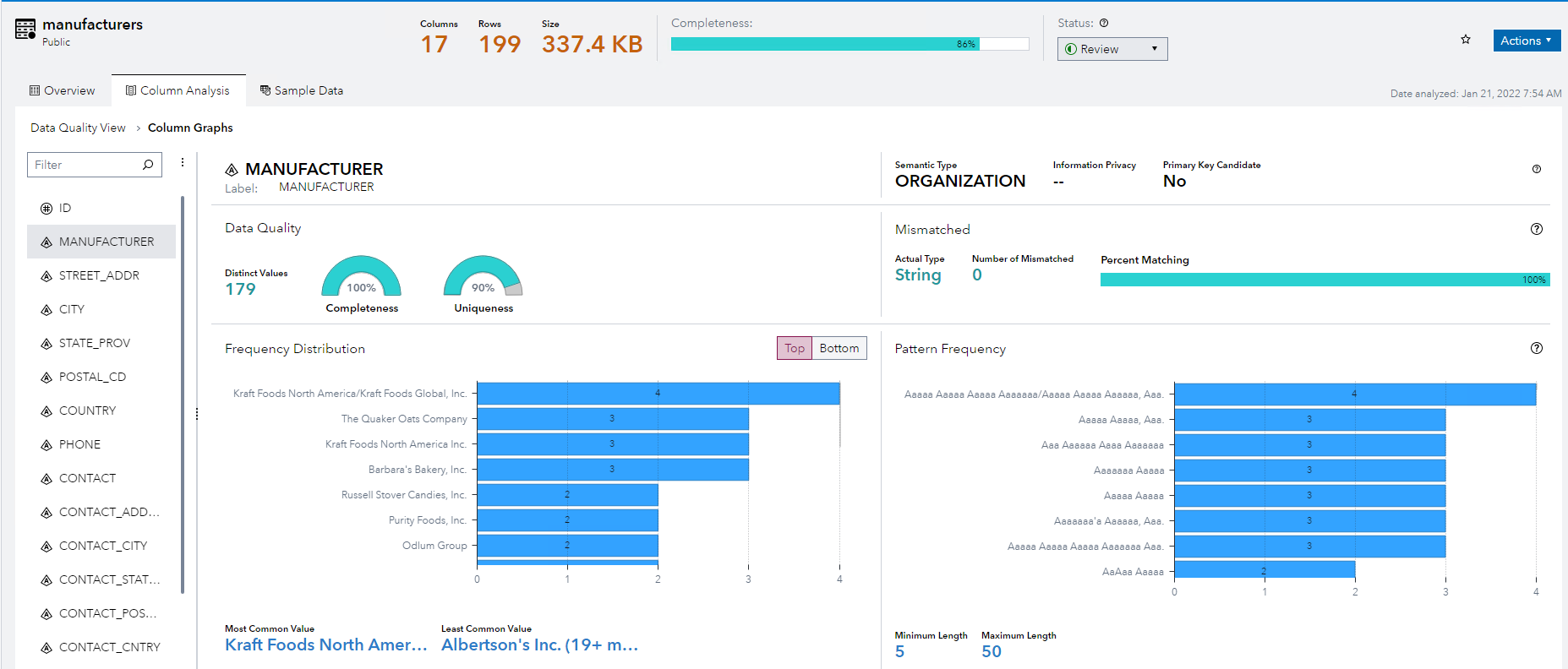
1. Click **Data Quality Measures**. Notice, that
   1. Information is about completeness, uniqueness, most common and least common value.
   2. There is also information about the privacy of the data. Click the question mark beside Information Privacy to understand more about this information.
   3. There is information about Semantic Type. Click the queston mark beside Semantic Type to understand what semantic type is about.



1. Answer these questions. Click this documentation [link](https://go.documentation.sas.com/doc/en/infocatcdc/v_008/infocatag/p09m7zz3ui9j80n1ek8m7ld05fvt.htm) to find answers.
   1. Is Information Privacy a basic or advanced feature of SAS Information Catalog?
   2. What SAS license do you need for advanced features in SAS Information Catalog?
   3. Does Semantic Type use the QKB?
   4. What procedure/CAS action is used behind the scenes when data sampling of big or wide tables is done?
2. Change **status** of manufacturers to **Review**.



1. In **Data Quality Measures**, investigate these columns.
   1. What is the completeness and uniqueness of the column MANUFACTURER?
   2. What is the most common name in the CONTACT column?
   3. What is the actual maximum length of CONTACT compared to the defined length of the column?
   4. What data quality issues does COUNTRY have. Add your findings to the Teams chat.

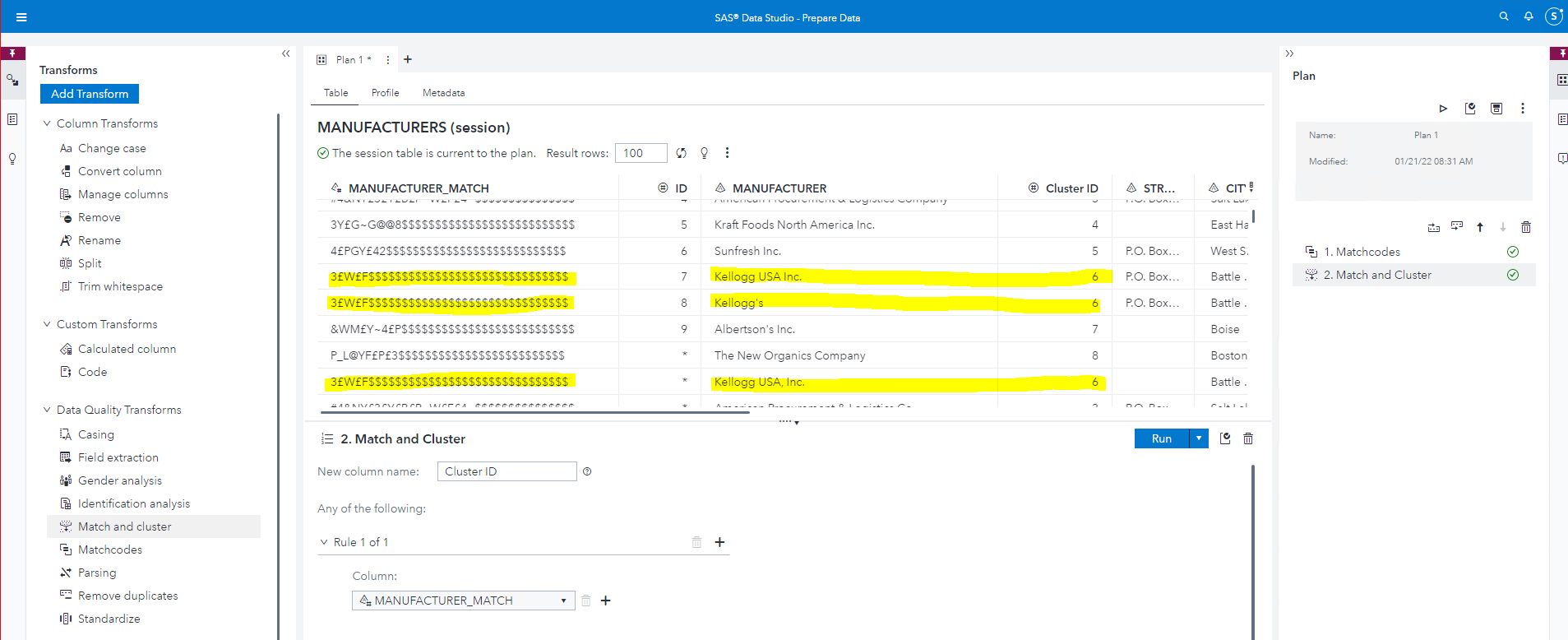


1. Use a facet search to find tables that are under review.
2. Search assets that are favorites. You made one a favorite in a previous exercise.

## Next Steps

You will not be completing next step actions due to workshop time constraints. There is a nice integration feature where you can perform an action on an asset such as opening a table asset in SAS Studio (where you can create flows or code), preparing a table asset in SAS Data Studio or opening a model in SAS Model Studio.

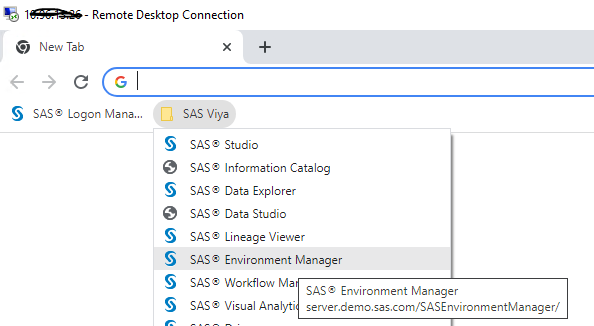
You have investigated issues on data quality in some of your table assets. For example, the table MANUFACTURERS has a problem with a manufacturer not being registered the same way. Using SAS Data Studio, you can do further investigation by creating match codes and clusters to group each manufacturer with a later intent of entity resolution.



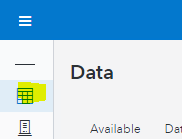
# Preparing CAS libraries for information assets

**The purpose of this exercise is to create a global CAS library, CASHOME, so that you can create a SAS Information Catalog Discovery Agent to crawl and analyze the contents.**

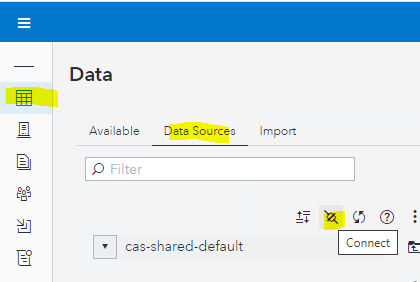
1. In Google Chrome on your remote Windows desktop select SAS Environment Manager from the SAS Viya shortcut menu and sign in as sasdemo with the password Orion123.



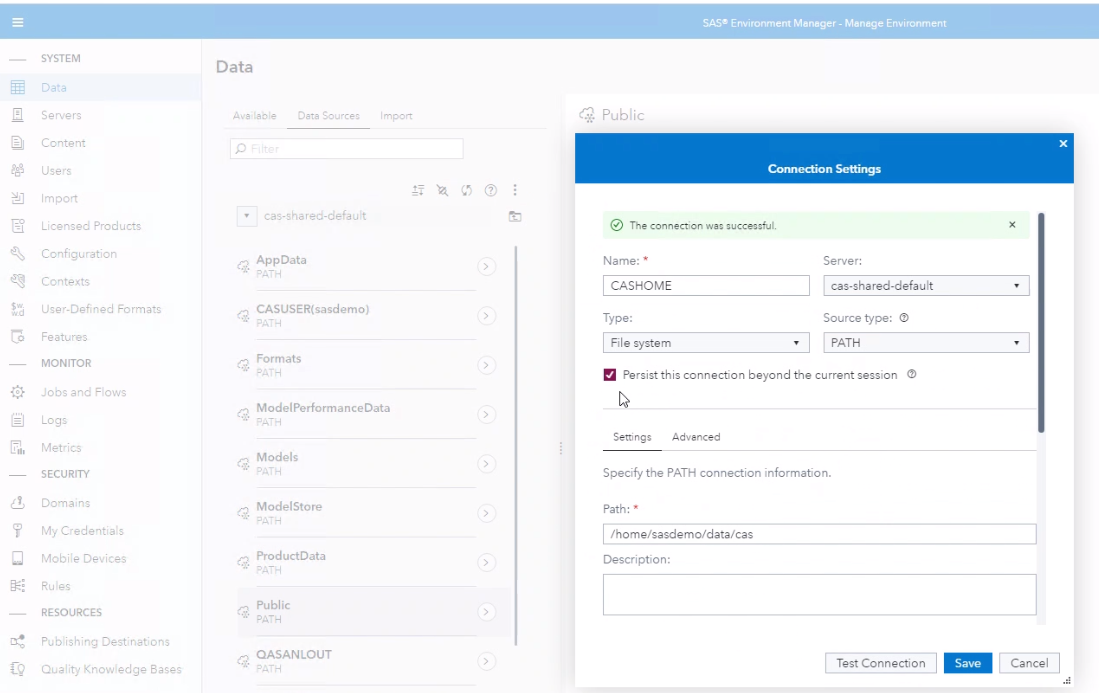
1. Select the Data menu in the left pane.



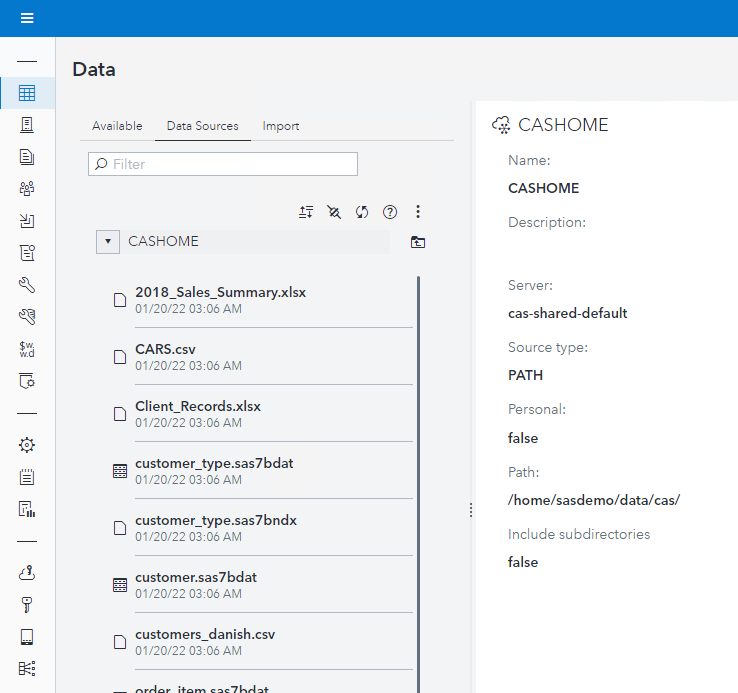
1. Select Data Sources and then the Connect icon.



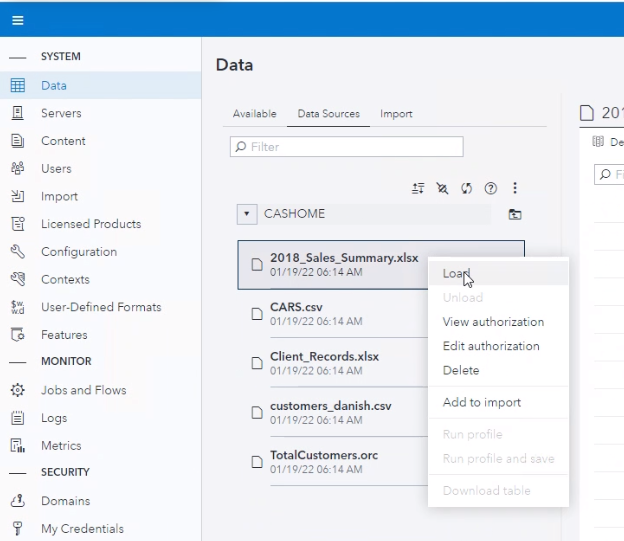
1. Provide this information in Connection Settings, test the connection and save it.



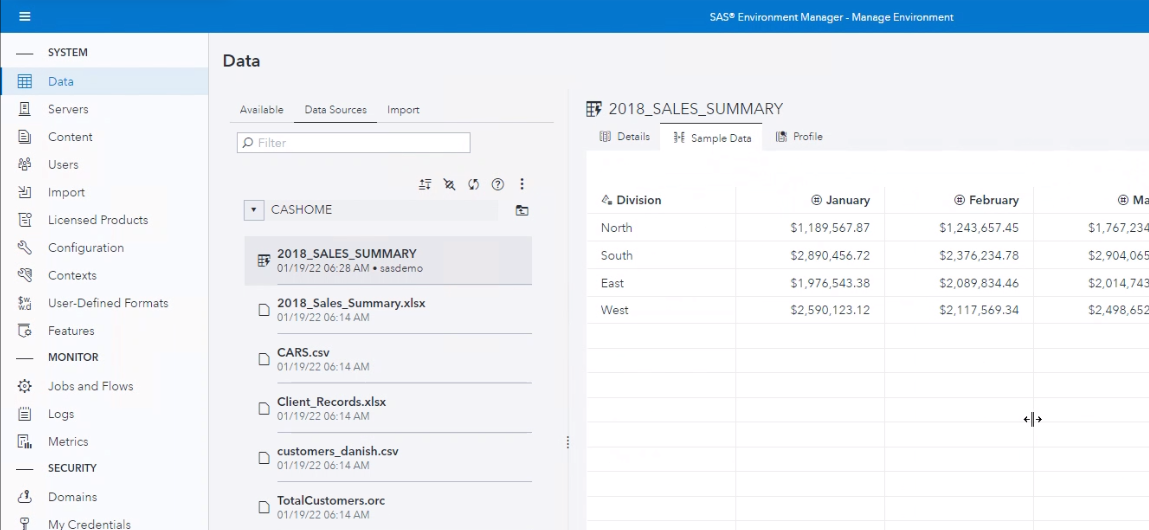
1. Expand CASHOME. You see all the physical files that you previously uploaded.



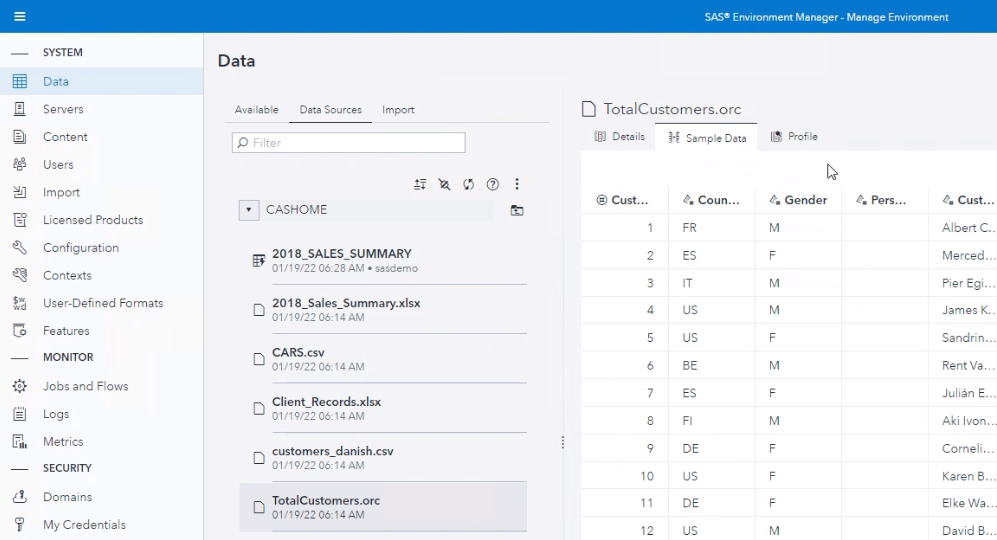
1. Load the file below to CAS.



1. Select the Sample Data tab to open the loaded CAS table.



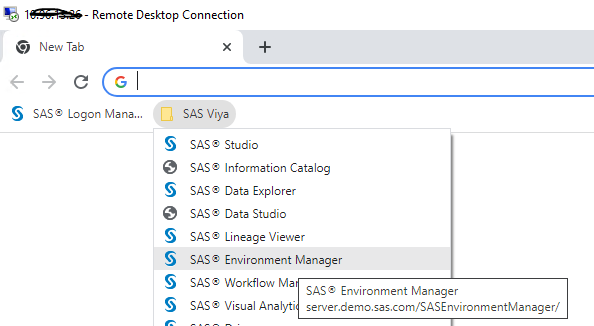
1. Confirm that you can see the data of the .ORC file.



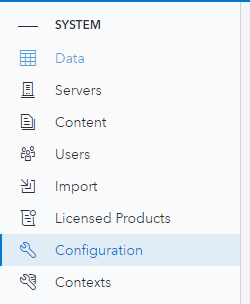
# Preparing SAS Compute libraries for information assets

**The purpose of this exercise is to create a SAS Compute library, so that you can create a SAS Information Catalog Discovery Agent to crawl and analyze the contents. (SAS compute libraries are only viable as an information asset when SAS Information Governance is licensed).**

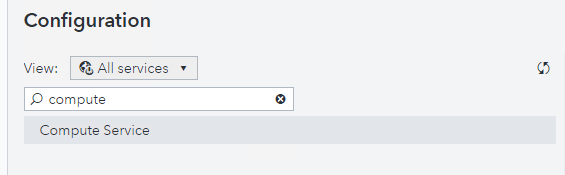
1. In Google Chrome on your remote Windows desktop select SAS Environment Manager from the SAS Viya shortcut menu and sign in as sasdemo with the password Orion123.



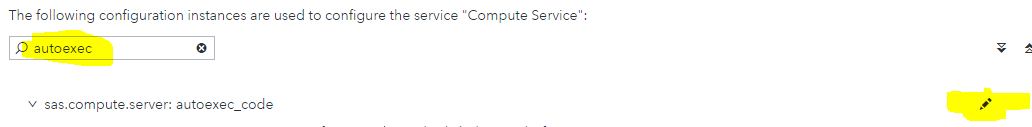
1. Select the Configuration menu in the left pane.



1. Type in *compute* in the search window and select Compute Service

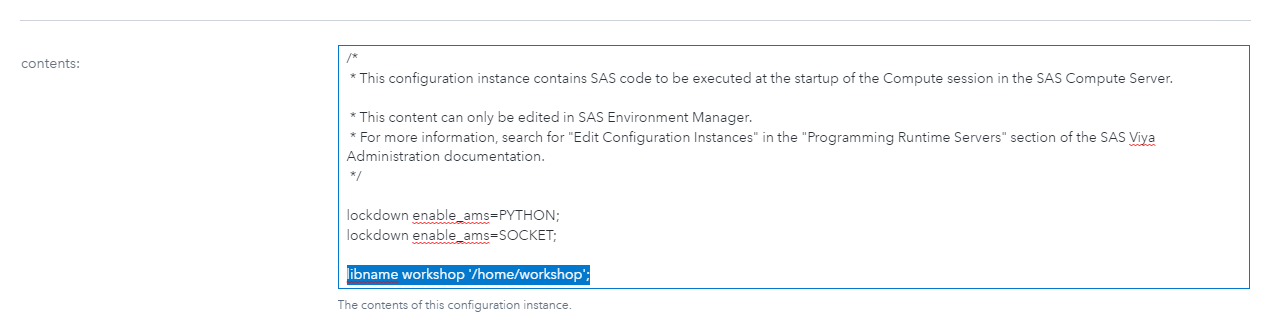


1. Type *autoexec* in the filter window and click the pencil to edit.



1. Add the following libname statement to the contents window and click Save.

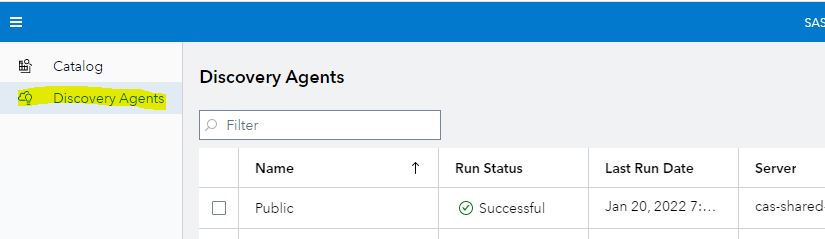
libname workshop '/home/workshop';



# Creating discovery agents for WORKSHOP and CASHOME libraries

**This is a case exercise without much guidance, but with tips. You can also review earlier exercises for step by step guidance for creating discovery agents.**

1. In SAS Information Catalog, click the **Discovery Agents** menu in the left pane if you are currently not in this menu.

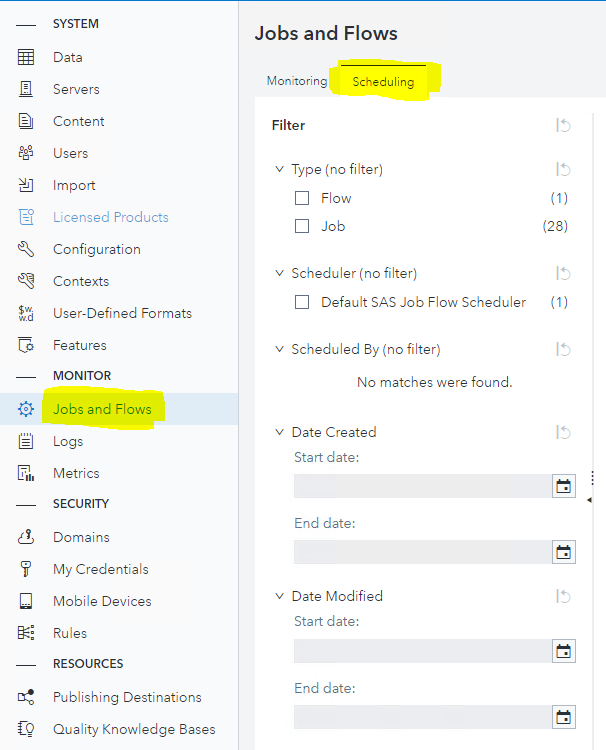


1. Add a new Discovery agent for the compute library workshop. (if workshop doesn’t show up in the list of libraries, make sure to hit the refresh button) 
2. Add a new Discovery agent for the cas library CASHOME.
3. Don’t forget to run the Discovery agents!

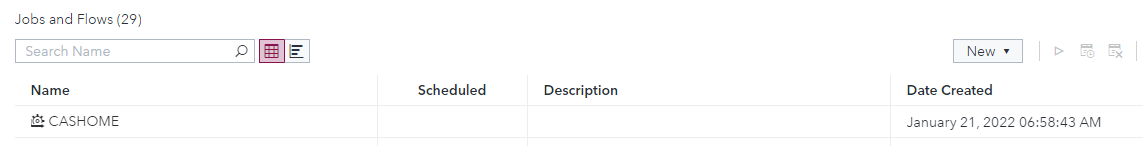
# Monitoring and scheduling discovery agent jobs

**In this exercise we will schedule the CASHOME discovery agent to run on a daily basis. In a real case the discovery agent could be scheduled to run after the datasource it’s conneted to has been updated.**

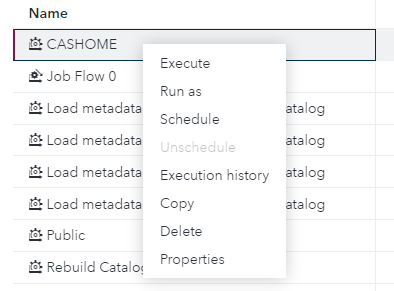
1. Open SAS Environment Manager (Manage environment), select Jobs and Flows and click the Scheduling tab.

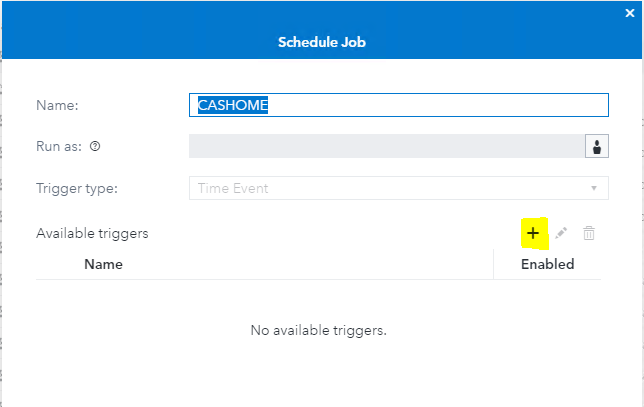


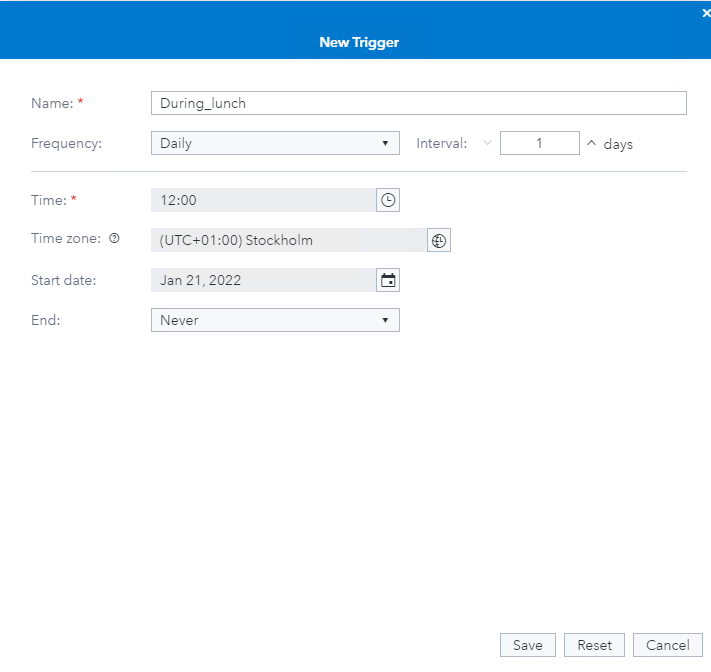
1. Notice, that to the right, you see a list of all jobs and flows. Find the job, CASHOME.



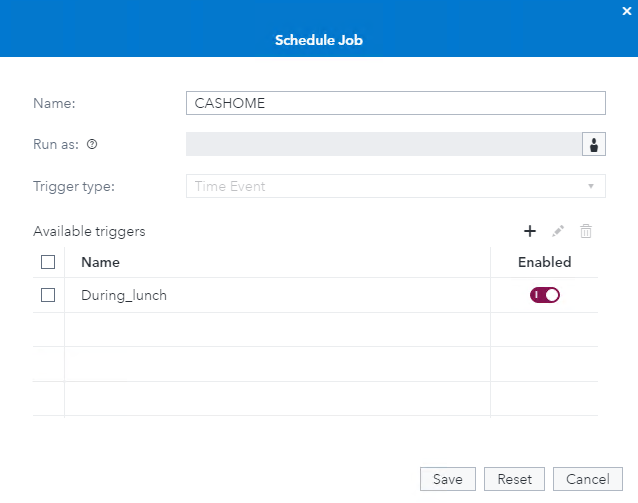
1. Do the following.
   1. Right click the CASHOME job to see the list of actions. Notice, that you can execute the Discovery agent immediately, or schedule it to run later.
   2. Click Schedule.



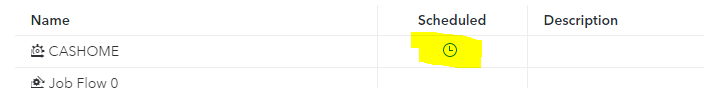
* 1. Notice that the trigger type is Time Event (in a later release there are other events).
  2. Add a Time event trigger by clicking the + sign.
  3. In the New Trigger window, you can configure at which time the scheduled job should run. Provide your trigger with a name and start time and ensure that you have selected the correct Time zone. Click Save when you are done.



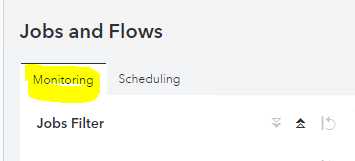
* 1. Enable your trigger and click Save. Your job is now scheduled.



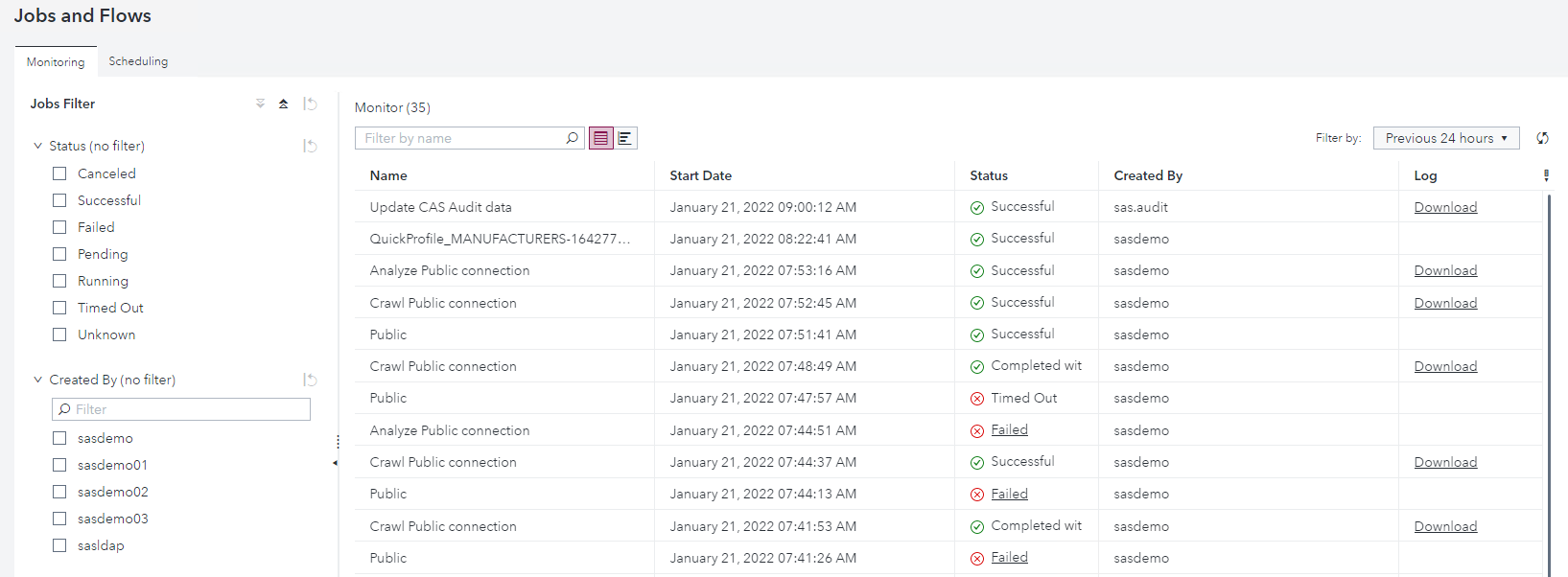
* 1. When a job is scheduled, the scheduled icon will show up in the list of Jobs and flows.



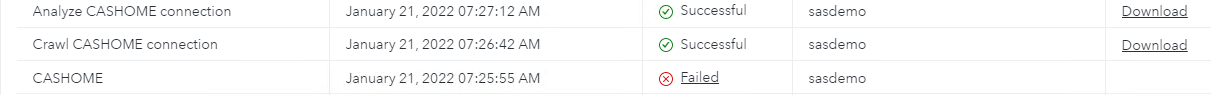
1. Click the Monitoring tab.



You get information about all jobs and flows that have been executed in the environment. You are also able to download the logs from the executed jobs.



1. Notice, that the Discovery agent job appears in the list once it is executed.

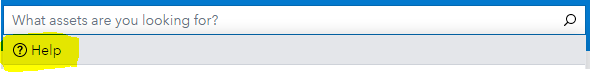


Each Discovery agent consists of a main job which has the same name as the discovery agent (CASHOME). When the main job is executed it will start two other jobs, one Crawl job and one Analyze job. The crawl job will search the library and list all assets in it. The analyze job then analyzes the data which results in all the information that you see in Information Catalog.

# Case - Discovering assets in SAS Information Catalog

**This is a case exercise where you discover assets to answer questions, using the functionality that you worked with in previous exercises that contain step by step guidance.**

1. Answer the following questions in any order you prefer using SAS Information Catalog. Don’t forget that you can use the Help button for information around searching for assets.



* 1. Have any Model Studio Projects been automatically catalogued?
  2. Which table has the lowest Completeness (tables that lack completeness don’t count)?
     1. Add Flagged as a status.
  3. Do you retrieve assets that contain a delivery address?
     1. Are any of them missing?
  4. Make a note of an asset that contains either a Given Name or Family Name.
  5. How much wood would a woodchuck chuck if a woodchuck could chuck wood (it is actually specified in the data)?
  6. What is the difference between the Semantic Type Individual and Given Name?
  7. Can you search for text in a Business Description?
  8. Are there assets that are a Visual Analytics report?
     1. Investigate a report and make a note of the table(s) that the report is using in lineage.
  9. Paste the address http://server.demo.sas.com/SASInformationCatalog/?q=Name:”customer” into your web browser.
     1. Which assets are returned?
     2. Try and change the facet in the URL to go directly to other information.

1. http://server.demo.sas.com/SASInformationCatalog/?q=Name:”customer”