



Baza wiedzy – SAS Data Catalog

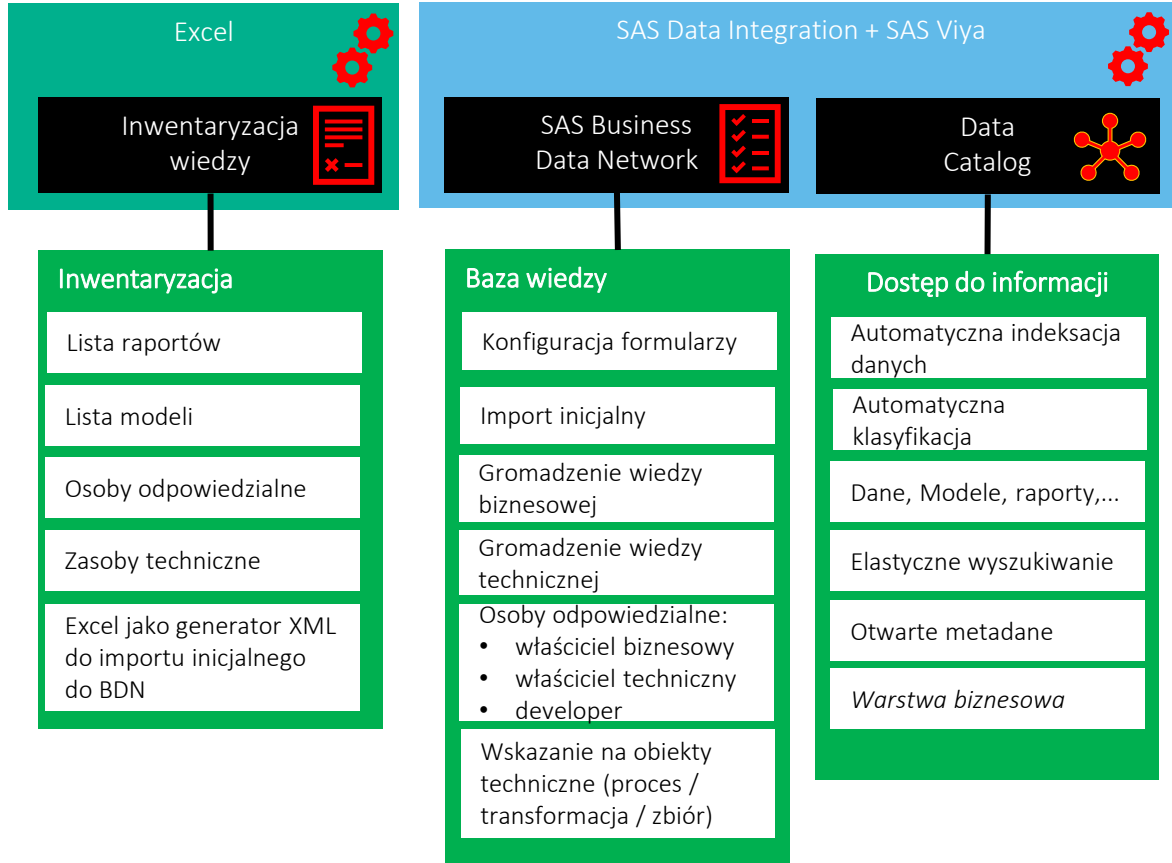
Łukasz Leszewski

Mariusz Kuśmierczyk

Mariusz Zychowicz

Warszawa, 24.02.2022

Wizja



Baza wiedzy - SAS Business Data Network



SAS Business Data Network

File Help Sign Out

Projektory wydania 2.0 2.1 2.3 B.001 T.001 B.002 B.003 2.4 2.5 CBP

New Term... Open Tags... Lineage Total count: 6

Term Name	Description	Type	Importance	Status	Workflow	Last Modified
Projekt	Centralna Baza Kontrahentów	projekt	☆☆☆☆	Not Specified		03/05/19 05:29:43 PM
wydania	lista wydań systemu	wydania	☆☆☆☆	Not Specified		
2.0	Projekt optymalizacji zasileń CBx	wydanie	☆☆☆☆	Production		
B.001	Test biznesowy deduplikacji	test	☆☆☆☆	Production		03/06/19 11:56:44 AM
B.002	Test biznesowy deduplikacji	test	☆☆☆☆	Production		03/06/19 11:48:42 AM
B.003	Test biznesowy deduplikacji	test	☆☆☆☆	Production		03/06/19 11:48:52 AM
T.001	Test techniczny czasu zasileń	test	☆☆☆☆	Production		03/06/19 11:33:32 AM
2.1	Rotacja danych w archiwum	wydanie	☆☆☆☆	On Hold		03/06/19 11:30:42 AM
2.3	Proces wykonujący eksport na potrz...	wydanie	☆☆☆☆	Production		03/06/19 11:30:51 AM
2.4	Proces weryfikacji spójności pomied...	wydanie	☆☆☆☆	Under Review		03/06/19 11:52:11 AM

B.001

Identification Hierarchy Associated Items Notes and Contacts History

Description: Test biznesowy deduplikacji

Requirements

Attributes

numer wersji: 1
opis: test biznesowy
URL:
opis długi: **Założenia testu:**
Oczekiwany wynik: Deduplikacja osób fizycznych, deduplikacja firm

Details

Type: test
Status: Production
Importance: Medium

Links: GitLab, SVN

Tags

Related Terms: 2.1 (wydanie), 2.5 (wydanie)

Lista zasobów, składowe zasobu (np. model, raport i zakładki)

Status, ważność zasobu

Wyzwolenie workflow

Osoby kontaktowe
Historia zmian

Dla danego typu zasobu - dedykowane formularze do gromadzenia danych

Linki zewnętrzne np. do GIT, SVN, JIRA...

Linki wewnętrzne w bazie wiedzy

BDN - Połączenie świata biznesowego i technicznego



SAS® Business Data Network

File Help Sign Out

Projektory x wydania x 2.0 x 2.1 x 2.3 x B.001 x T.001 x B.002 x B.003 x 2.4 x 2.5 x CBP x

New Term... Open Tags... Lineage Total count: 6

Term Name	Description	Type	Importance	Status	Workflow	Last Modified
Projekt		Default	☆☆☆☆☆	Not Specified		03/05/19 05:29:43 PM
CBK	Centralna Baza Kontrahentów	projekt	☆☆☆☆☆	Production		03/05/19 05:35:49 PM
wydania	lista wydań systemu	wydania	☆☆☆☆☆	Not Specified		03/06/19 11:30:28 AM
2.0	Projekt optymalizacji zasileń CBx	wydanie	☆☆☆☆☆	Production		03/06/19 11:28:48 AM
B.001	Test biznesowy deduplikacji	test	☆☆☆☆☆	Production		03/06/19 11:56:44 AM
B.002	Test biznesowy deduplikacji	test	☆☆☆☆☆	Production		03/06/19 11:48:42 AM
B.003	Test biznesowy deduplikacji	test	☆☆☆☆☆	Production		03/06/19 11:48:52 AM
T.001	Test techniczny czasu u zasileń	test	☆☆☆☆☆	Production		03/06/19 11:33:32 AM
2.1	Rotacja danych w archiwum	wydanie	☆☆☆☆☆	On Hold		03/06/19 11:30:42 AM
2.3	Proces wykonujący ekstrakt na potrz...	wydanie	☆☆☆☆☆	Production		03/06/19 11:30:51 AM
2.4	Proces weryfikacji spójności pomied...	wydanie	☆☆☆☆☆	Under Review		03/06/19 11:52:11 AM

B.001

Identification Hierarchy **Associated Items** Notes and Contacts History

Lineage

Name	ID	Description	Note	Type	Path
__BID_D0D_DECU1 - Or...	A55UOD6S.BK00000Z/Phy...			Table	/User Folders/SAS Demo Us...

Połączenie informacji biznesowych i technicznych - wskazanie na obiekty metadanych (zbiory / procesy / transformacje)

BDN – dedykowane formularze



The screenshot shows a settings form for BDN. It includes sections for 'Type Identification' (Name: test, Description:), 'Workflows (optional)' (Create, Edit, Delete dropdowns), 'Specify status values and select the default' (Production, Editing, On Hold, Under Review, Not Specified), 'Specify importance labels and select the default' (Critical, High, Medium, Low, Very Low, Not Specified), and 'Business roles' (Business Owner, Interested Party, Technical Owner).

Możliwość definiowania workflow opartego o akcje create / edit / delete

Statusy (np. w trakcie budowy, testów, używane produkcyjnie)

Priorytety (jak ważny jest dany zasób)

Role (właściciel, developer...)

Atrybuty gromadzone dla danego zasobu

Label	Instructions	Default Value	Type	Required	Hidden
numer wersji			Single Line Text Entry	Yes	No
opis			Multi-Line Text Entry	Yes	No
URL			URL	No	No
opis dlugi			Rich Text Entry	No	No

BDN – dedykowane formularze



Settings Attributes

New Attribute... Edit... Preview

Label	Instructions	Default Value	Type	Required	Hidden
numer wersji			Single Line Text Entry	Yes	No
opis			Multi-Line Text Entry	Yes	No
URL			URL	No	No
opis dlugi			Rich Text Entry	No	No

New Attribute

* Label:
New Attribute

Instructions:

Type:
Single Line Text Entry |
Single Line Text Entry
Multi-Line Text Entry
Rich Text Entry
Single Selection
Multiple Selection
Boolean
URL
Date

OK Cancel

Typ atrybutu

BDN– definiowanie formularzy

Możliwość dokumentacji ABT

Atrybut: * aktualność

Algorytm wyliczania:

Kolejność wyliczenia:

Sprawdzenie: DATE >= & MV_BusinessDate.

Filtr danych wchodzących do reguły:

Słownik reguły - nazwa:

Słownik reguły - kolumna:

Słownik reguły - filtr:

Dopuszczalne puste:

Funkcja weryfikująca: * brak

Funkcja weryfikująca argumenty:

Kolumna grupująca: Jeśli algorytm wyliczania zakłada wyliczenie statystyk w grupach

Słownik wyjątków - nazwa:

Słownik wyjątków - kolumna:

Słownik wyjątków - wskaźnik:

Słownik wyjątków - filtr:

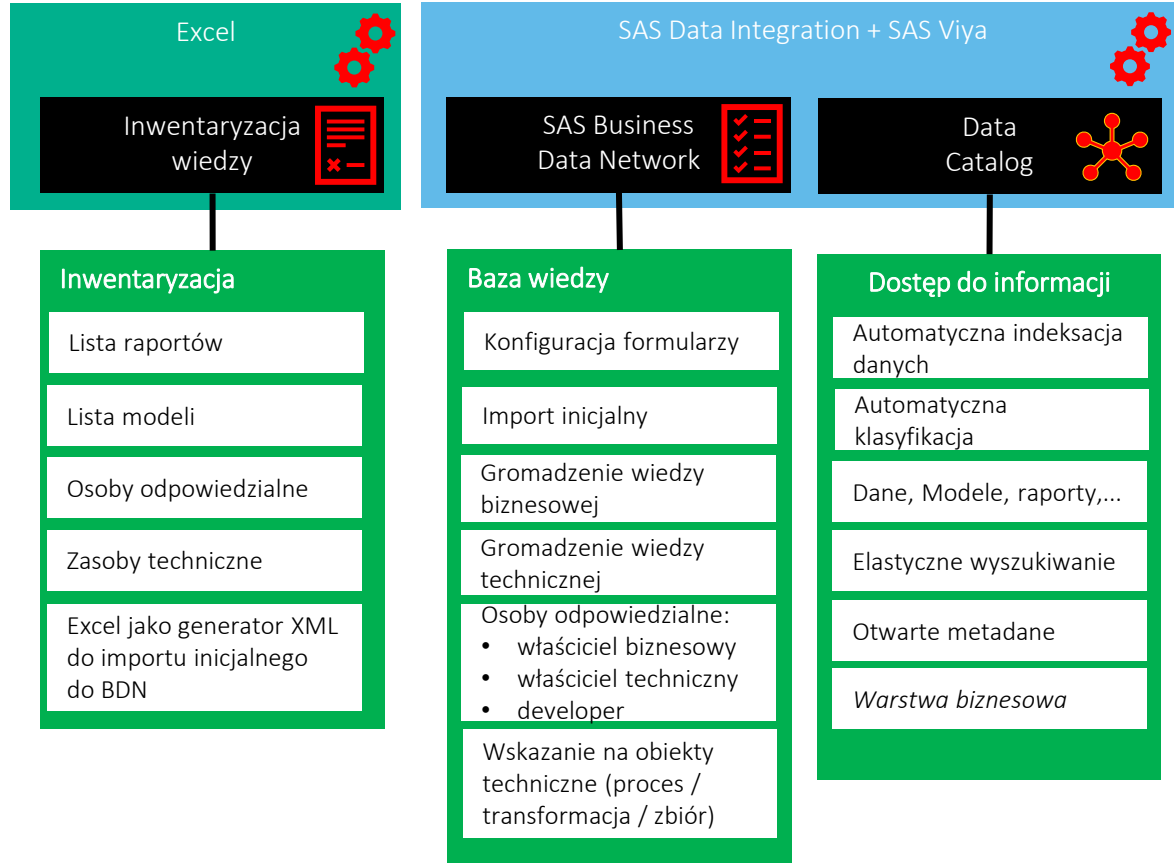
Obszary danych		Default	★★★★☆	● Produkcja
01. Dane klientów indywidualny		Default	★★★★☆	● Produkcja
CLASS		Obszar danych - tabela	★★★★☆	● Produkcja
DATE AKTUALNOŚĆ	Aktualność danych	Obszar danych - kolumna	★★★★☆	● Weryfikacja
NAME FORMAT	Sprawdzenie formatu	Obszar danych - kolumna	★★★★☆	● Produkcja
NAME UNIKALNOŚĆ	Sprawdzenie unikalności kolumny NAME	Obszar danych - kolumna	★★★★☆	● Produkcja
NAME_ILE ZLICZ UNIK	Zmienna zliczająca ilość unikalnych wartości kolu.	Obszar danych - kolumna	★★★★☆	● Produkcja
SEX BRAK	Wykrycie braków danych	Obszar danych - kolumna	★★★★☆	● Produkcja
SEX FORMAT	Sprawdzenie formatu	Obszar danych - kolumna	★★★★☆	● Produkcja
SEX SPÓJNOŚĆ	Spójność ze słownikiem	Obszar danych - kolumna	★★★★☆	● Produkcja
02. Dane klientów biznesowych		Default	★★★★☆	

Przykład formularza dla reguł walidacji danych

- Reguły definiowane poprzez prosty formularz
- Reguła **sprawdzająca**: dowolne wyrażenie w język SAS 4G np. `length(PESEL) = 11`
 - Reguła **słownikowa** – sprawdza, czy wartość znajduje się w słowniku
 - Reguła używająca **predefiniowanych funkcji** (np. weryfikacja PESEL, REGON)
 - Reguła **wyjątków** – sprawdza, czy wartość znajduje się w słowniku wyjątków
 - Reguła **kolumna wyliczana** – służy do wyliczenia pomocniczych wartości
 - Reguła **agregująca** – umożliwia wyliczenie wartości zagregowanych

Możliwość definiowania formularzy do gromadzenia danych, tu przykład do gromadzenia informacji o tabelach lub kolumnach tabel

Wizja



Zasoby informacyjne

Zarządzanie

SAS Information Catalog - Discover Information Assets

Name	Run Status	Last Run Date	Server	Library	Physical Region	Discovery Locale	New Assets	Updated Assets	Total Assets	Date
Public	Successful	Jan 20, 2022 7:...	cas-shared-default	Public	Worldwide	United States (E...	9	1	11	Jan 20, 2

Administrator tworzy i zarządza agentami do przeszukiwania danych

SASHELP

Properties Configuration History

Select which tables will be discovered by the agent.

All tables

Filter rules

Include all tables with the following: [Manage rules](#)

Filter	Rule
No rules have been added.	

[Preview](#)

Preview how many assets will be included and excluded.

Manage rules

Create rules to filter which tables in the library will be analyzed by the discovery agent. Multiple rules will be handled as OR statements.

tables that meet the following conditions.

Table name:

[+ Add Rule](#)

<input type="checkbox"/>	Water Consumption and M...	Report
<input type="checkbox"/>	Customer Segmentation tSNE	Model Studio project
<input type="checkbox"/>	Write_Dashboard_Tables.flw	Studio flow
<input type="checkbox"/>	Mathias_demo	Model project
<input type="checkbox"/>	RETAIL ANALYSIS DATA	CAS table
<input type="checkbox"/>	AIR	SAS table

To nie tylko katalog danych.
Dostępne są również inne zasoby

Searching for assets

SAS Information Catalog - Discover Information Assets

Catalog Home > Search Results

Asset types: (no filter)

Top 39 Results

Name	Status	Type	Date Modified	Modified By	Date Created	Created By
2019_Sales_Summary	☆	Microsoft E...	Jan			
AutomobileTest	☆	Text file	Jan			
ClothingProducts	☆	Text file	Jan			
contact_list	☆	Comma-se...	Jan			
customer_info	☆	SAS table	Jan			
CLASS	☆	Comma-se...	Jan			
Customer_ratings_swedish	☆	Microsoft E...	Jan			
HMEQ	☆	CAS table	Dec			
allcustomers	☆	Parquet table	Jan			
predel_surtlist	☆	CAS table	Jan			
CSBMVA_BIN_CHRSTC	☆	SAS table	Nov			
CITWIK	☆	SAS table	Nov			
COUNTSERIES	☆	SAS table	Nov			
CSBMVA_BIN_ANALYSIS_S...	☆	SAS table	Nov			
CARS	☆	SAS table	Nov			
CYTMANU	☆	SAS table	Nov			

Use a * to get a list of assets

Use Space bar to get a list of facets

SAS Information Catalog - Discover Information Assets

Catalog Home > Search Results

Asset types: (no filter)

Top 3 Results

Name	Status	Type	Date Modified	Modified By	Date Created	Created By
allcustomers	☆	Parquet table	Jan 20, 2022 3:00 AM	--	Jan 20, 2022 3:00 AM	root
customer_info	☆	SAS table	Jan 20, 2022 3:00 AM	--	Jan 20, 2022 3:00 AM	root
Customer_ratings_swedish	☆	Microsoft E...	Jan 20, 2022 3:00 AM	--	Jan 20, 2022 3:00 AM	root

Add free text.

SAS Information Catalog - Discover Information Assets

Help

- Category:
- Column: {}
- Column.dataType:
- Column.description:
- Column.informationPrivacy:
- Column.keywords:
- Column.label:
- Column.name:
- Column.semanticType:
- Column.status:
- Column.Count:
- Completeness:
- CreatedBy:
- DataSetSize:
- dataSetDataFields: {}
- dataStoreDataSets: {}
- dataStoreDataSets.dataStore: {}
- DateAnalyzed:
- DateCreated:
- DateModified:
- DefinitionID:
- Description:
- FileExtension:
- FreeText:
- InformationPrivacy:
- Keywords:
- Label:

Library:

How to enter text value

Library: "public"

public

sashelp

Drilling into table assets – Column Analysis

Descriptive measures

Metadata measures

Data Quality measures

SAP Information Catalog - Discover Information Assets

manipulators Public

Columns: 17 Rows: 199 Size: 337.4 KB

Completeness: 86%

Status: None

Column Properties

Frequency distribution:

manipulators Public

Columns: 17 Rows: 199 Size: 337.4 KB

Completeness: 86%

Status: None

Metadata Measures

#	Name	Label	Type	Actual Type	Logical Type	Format	Length	Minimum Length	Maximum Length
1	ID	ID	double	--	ID	F	8	--	--
2	MANUFACTURER	MANUFACTURER	char	String	Nominal	\$	50	5	50
3	STREET_ADDR	STREET_ADDR	char	String	Nominal	\$	255	7	46
4	CITY	CITY	char	String	Nominal	\$	50	5	12
5	STATE_PROV	STATE_PROV	char	String	Nominal				
6	POSTAL_CD	POSTAL_CD	char	String	Nominal				
7	COUNTRY	COUNTRY	char	String	Nominal				
8	PHONE	PHONE	char	String	Nominal				
9	CONTACT	CONTACT	char	String	Nominal				
10	CONTACT_ADDRESS	CONTACT_ADDRESS	char	String	Nominal				
11	CONTACT_CITY	CONTACT_CITY	char	String	Nominal				
12	CONTACT_STATE_P...	CONTACT_STATE_PROV	char	String	Nominal				
13	CONTACT_POSTAL...	CONTACT_POSTAL_CD	char	Integer	Nominal				
14	CONTACT_CNTRY	CONTACT_CNTRY	char	String	Nominal				

manipulators Public

Columns: 17 Rows: 199 Size: 337.4 KB

Completeness: 86%

Status: None

Data Quality Measures

#	Name	Completeness	Uniqueness	Most Common Value	Least Common Value	Pattern Count	Semantic Type	Information Privacy
1	ID	100%	100%	408	1	--	GENERIC_ID	
2	MANUFACTURER	100%	90%	Kraft Foods North A...	Albertson's Inc. (19+...	168	ORGANIZATION	
3	STREET_ADDR	47%	38%	P.O. Box 049003	101 Utah Ave. So. (1...	55	DELIVERY ADDRESS	Candidate
4	CITY	96%	50%	Melville	Eugene (19+ more)	40	CITY	Candidate
5	STATE_PROV	81%	33%	NY (1 more)	B.C. (19+ more)	12	STATE/PROVINCE	Candidate
6	POSTAL_CD	93%	52%	11747	01301 (19+ more)	2	POSTAL CODE	Candidate
7	COUNTRY	93%	5%	US	Canada	2	COUNTRY	Candidate
8	PHONE	50%	38%	800-434-4246	(585) 374-9108 (19+...	14	PHONE	Private
9	CONTACT	100%	67%	Mark Antonucci	Alan Cunningham (1...	80	INDIVIDUAL	Private
10	CONTACT_ADDRESS	99%	64%	14 Lowery Ave (1 m...		114	DELIVERY ADDRESS	Candidate

Investigating privacy of information

Automatic classification of semantic type based on field name and content

SAS® Information Catalog - Discover Information Assets

Catalog Home > CUSTOMER_SENSITIVE

CUSTOMER_SENSITIVE
SAMPGEL

Columns: 24 Rows: 2 Size: 969 bytes

Completeness: 83%

Status: Flagged

Actions

Date analyzed: Jun 23, 2021 8:06 PM

#	Name	Uniqueness	M...	Lea...	Pattern Count	Semantic Type	Information Privacy
22	Religion	100%	Ba...		1	RELIGION	Sensitive
19	MARITAL_STATUS	100%	M...		1	MARITAL_STATUS	Sensitive
17	DOB	100%	14...	11...	1	BIRTH DATE	Private
14	PHONE	100%	4...		--	PHONE	Private
18	AGE	100%	55...	43	--	AGE	Private
15	GENDER	100%	F		1	GENDER	Private
16	EMAIL	100%	ja...		1	E-MAIL	Private
13	NAME	100%	Jo...	Jan...	1	INDIVIDUAL	Private
11	ZIP	100%	31...	3032	--	POSTAL CODE	Candidate
6	ADDR1	100%	6/...	30 ...	2	ADDRESS	Candidate

Column Properties

Frequency distribution:

Column name: GENDER
Label: --
Type: varchar
Actual type: Boolean
Format: --
Length: 1
Minimum length: 1
Maximum length: 1
Primary key: Yes
candidate:

Investigating privacy of information

How private ?

SAS® Information Catalog - Discover Information Assets

Catalog Home > CUSTOMER_SENSITIVE

CUSTOMER_SENSITIVE
SAMPGEL

Columns: 24 Rows: 2 Size: 969 bytes

Completeness: 83%

Logged

Actions


Overview Column Analysis Sample Data

Date analyzed: Jun 23, 2021 8:06 PM

#	Name	Uniqueness	M...	Lea...	Pattern Count	Semantic Type	Information Privacy
22	Religion	100%	Ba...		1	RELIGION	Sensitive
19	MARITAL_STATUS	100%	M...		1	MARITAL_STATUS	Sensitive
17	DOB	100%	14...	11...	1	BIRTH DATE	Private
14	PHONE	100%	4...		--	PHONE	Private
18	AGE	100%	55...	43	--	AGE	Private
15	GENDER	100%	F		1	GENDER	Private
16	EMAIL	100%	ja...		1	E-MAIL	Private
13	NAME	100%	Jo...	Jan...	1	INDIVIDUAL	Private
11	ZIP	100%	31...	3032	--	POSTAL CODE	Candidate
6	ADDR1	100%	6/...	30 ...	2	ADDRESS	Candidate

Column Properties

Frequency distribution:



Column name: GENDER
Label: --
Type: varchar
Actual type: Boolean
Format: --
Length: 1
Minimum length: 1
Maximum length: 1
Primary key: Yes
candidate:

Start in Information Catalog

Work with the assets anywhere

SAS Information Catalog - Discover Information Assets

Catalog Home > Search Results

Search indexes: (no filter)

Top 100 Results

Name	Status	Asset Type	Date Modified
Model 1	--	Model	Jan 2, 2022
cv_model	--	Model	Feb 11, 2022 2:36 PM

Actions menu for Model 1:

- Open model
- Register model
- Explore lineage

SAS Information Catalog - Discover Information Assets

Name: cust*

Search indexes: (no filter)

Top 30 Results

Name	Status	Asset Type	Date Modified
CUSTOMERS	●	CAS table	Dec 19, 2021
CUSTOMERS.sashdat	--	CAS table	Apr 12, 2022
CUSTOMERS	●	Table	Jan 1, 11:00 AM
CustomerFocusGroup	●	Table	Jan 1, 11:00 AM
CUSTOMER	●	CAS table	Feb 7, 2022
CUSTOMER	●	Table	Feb 7, 2022 2:24 PM
CUSTOMER_US_MASKED.s...	--	CAS table	May 19, 2021 6:08 AM
CUSTOMERS_DETAILS	●	In-memory ...	Feb 14, 2022 11:02 AM
CUSTOMERS_MASKED	●	In-memory ...	Feb 14, 2022 11:02 AM
CUSTOMERS	●	In-memory ...	Feb 14, 2022 9:07 AM

Actions menu for CUSTOMERS:

- Change table status >
- Open
- Prepare data
- Manage data
- Explore and visualize
- Build model
- Explore lineage

Different objects have different options.

You can start your data journey in Information Catalog and move to SAS Visual Analytics to build reports or check lineage or build a model.



mariusz.zychowicz@sas.com

tel. 693 410 366

sas.com