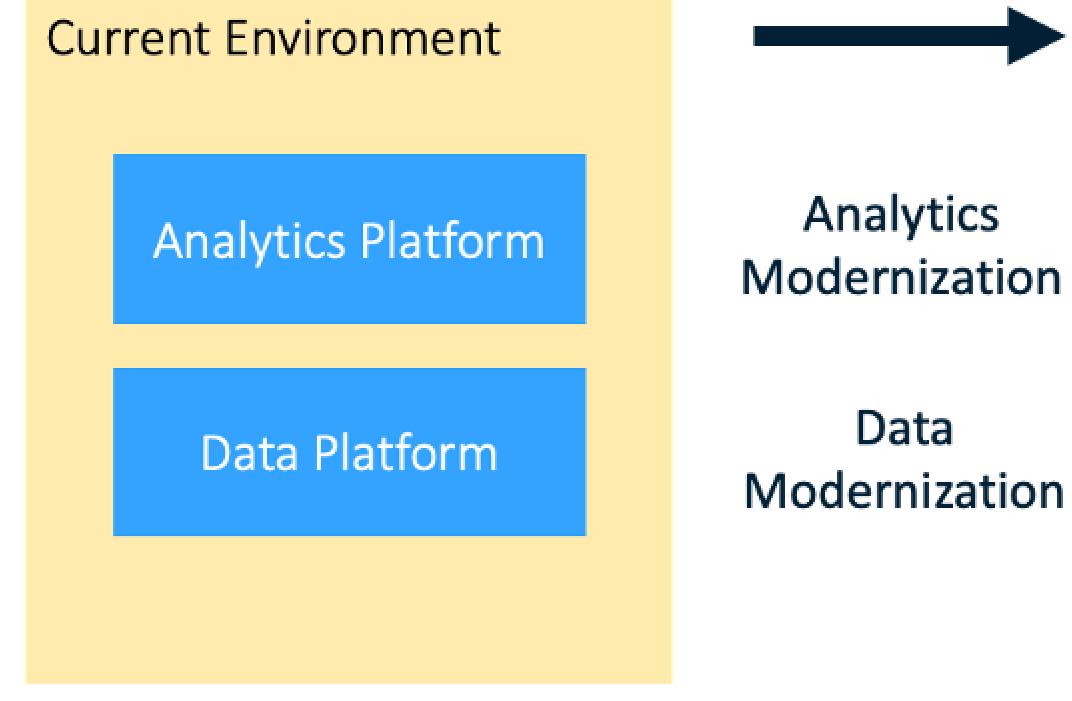
Governance and data quality while moving to the cloud with SAS Viya FANS Kristiansand

Lars Arne Skår, solution architect

Copyright © SAS Institute Inc. All rights reserved.



Moving to Cloud



Source: Jeff Stander @ SAS



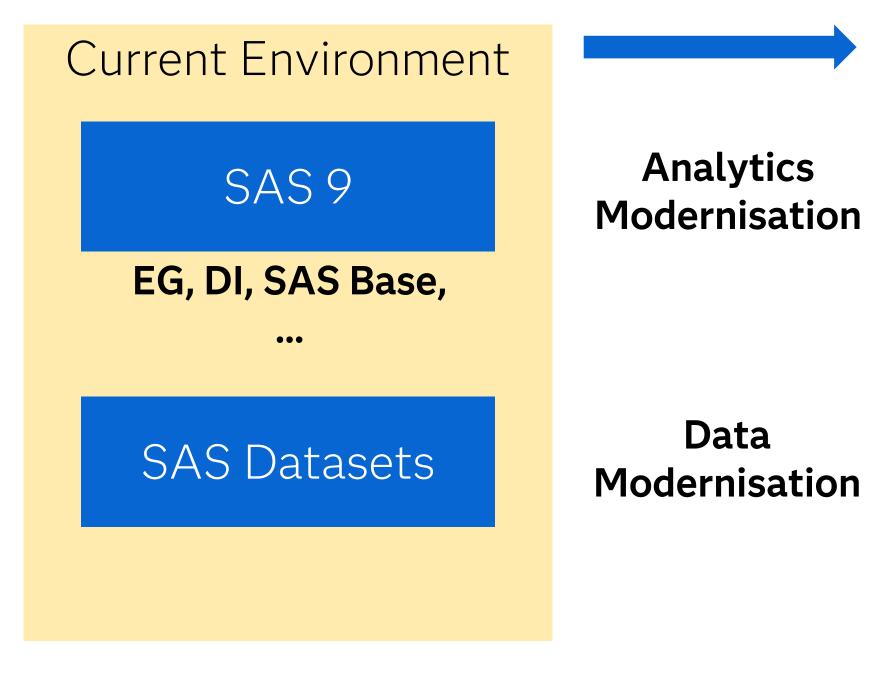
Cloud Environment

Modern Analytics Ecosystem

Modern Data Fabric



Moving SAS to the cloud



Source: Jeff Stander @ SAS

Copyright © SAS Institute Inc. All rights reserved.

Cloud Environment

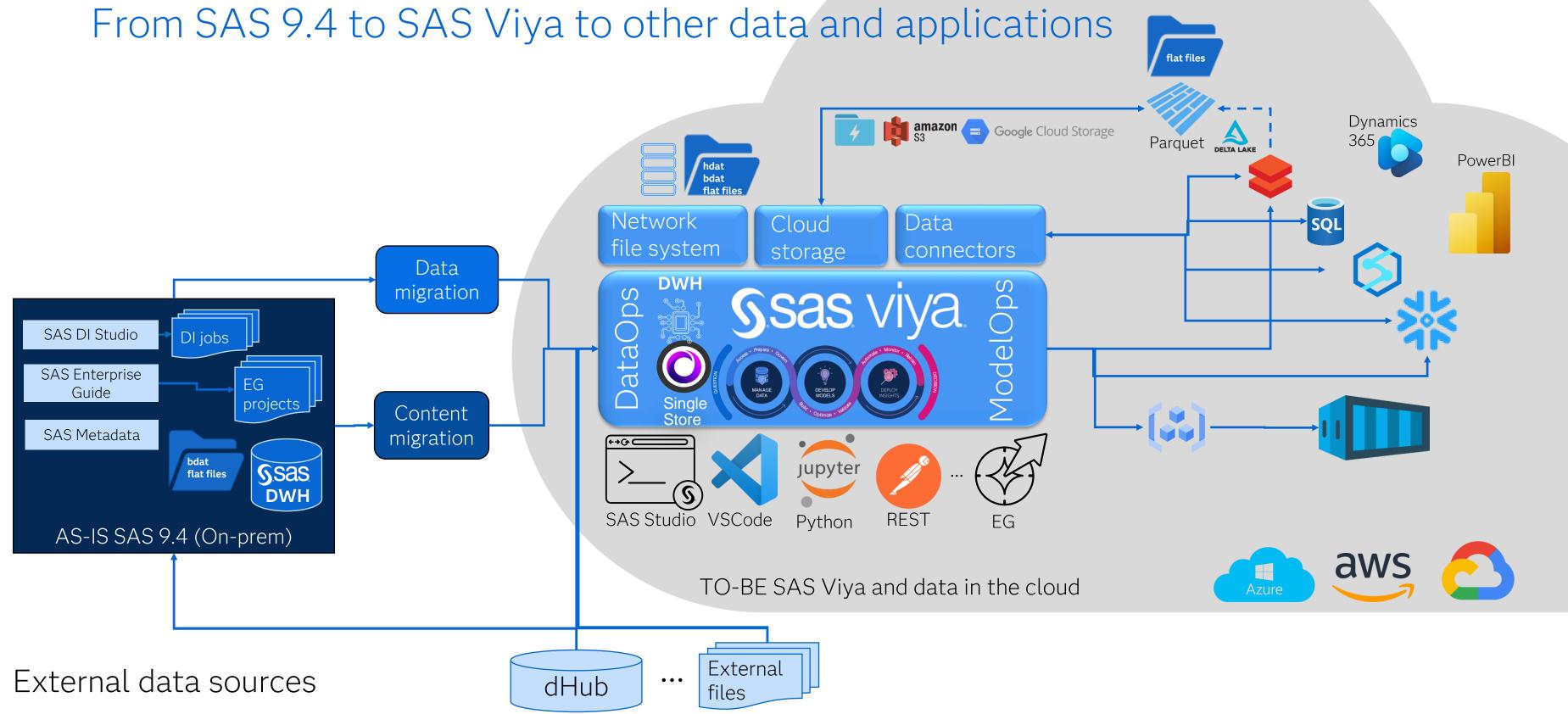
SAS Viya

Studio flows, SAS Base, Python, ...

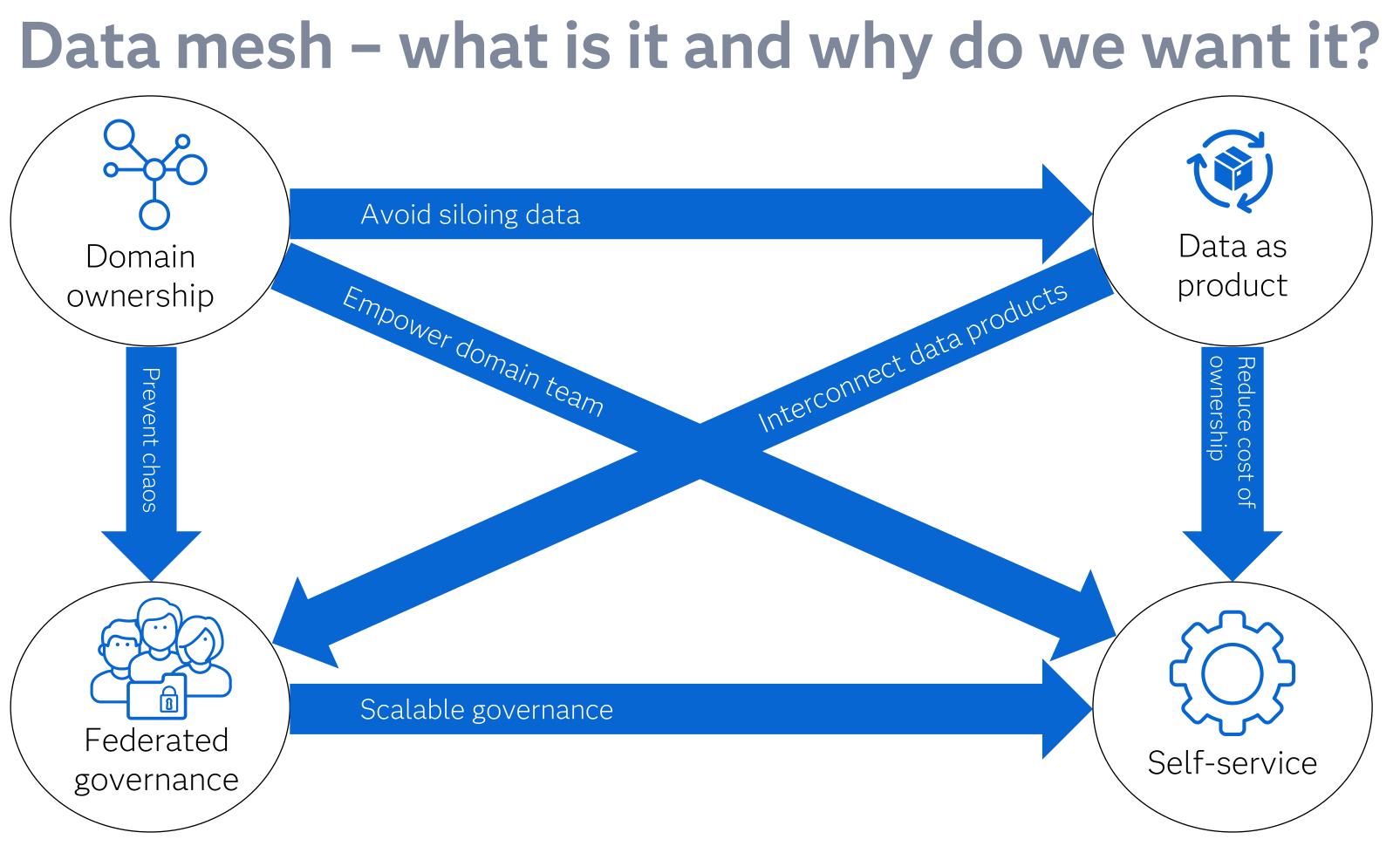
Optimal Storage



Data, analytics and Al architecture to the cloud From SAS 9.4 to SAS Viya to other data and applications





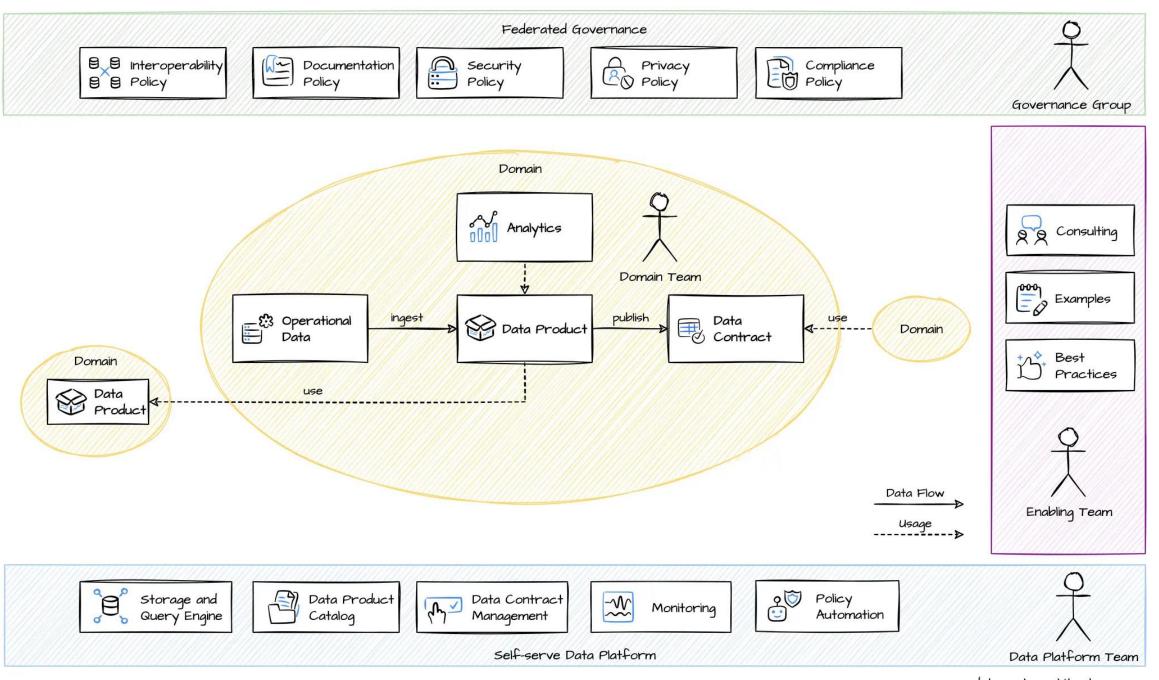


Copyright © SAS Institute Inc. All rights reserved.



Key capabilities from SAS for Data Mesh Federated governance, data access on-prem and cloud, self-service

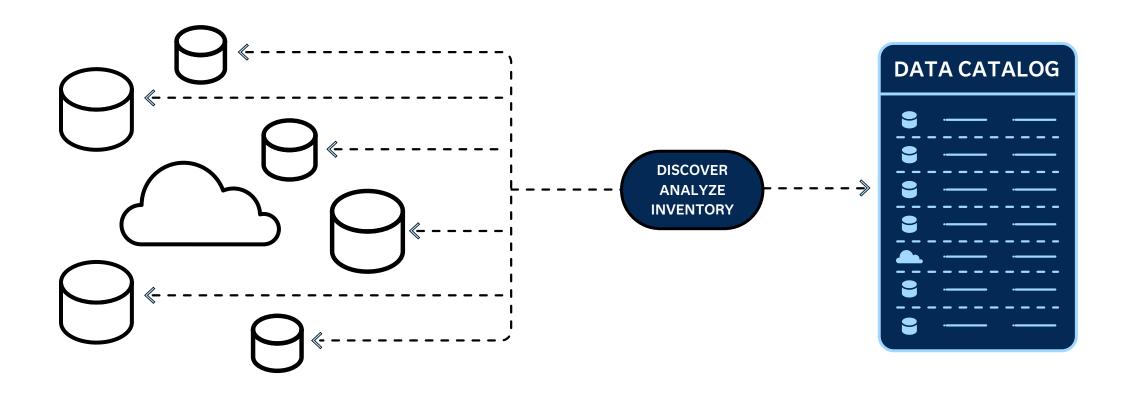
Data Mesh Architecture



datamesh-architecture.com

- Federated Governance through information catalogue and data lineage as well as embedded security enforcement in the platform
- **Self-service data** flow with SAS Studio flow and enablement through custom steps and popular languages for analytics and data prep (SAS, Python and R)
- **Operational and analytics processing** (in-DB, container, in-memory, on the edge, events)
- Providing **access to data** through external data sources, singlestore or REST-based access through MAS
- Storage and query engine through our access and connect modules and compute/CAS
- Data contract management through our information catalogue and business glossary
- Data exploration and visualization on metadata and content
- Data product catalog with our information catalogue using our discovery agent
- Monitoring and policy automation through DQ monitoring with business rules using QKB and discovery agents

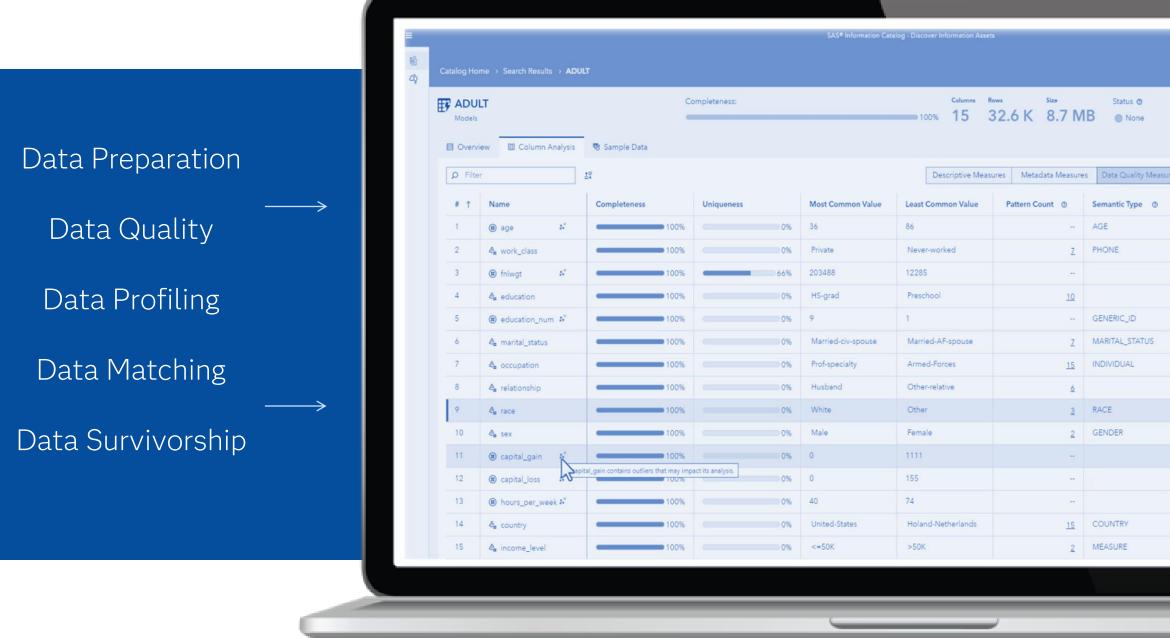






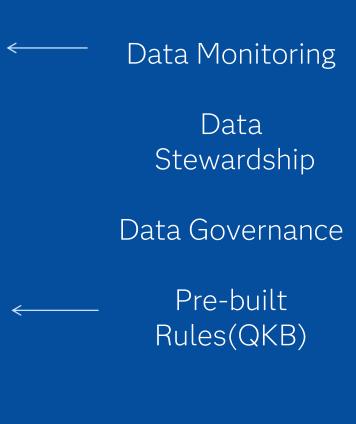
SAS Data Engineering Platform Overview

One single centralized platform for Data Citizens



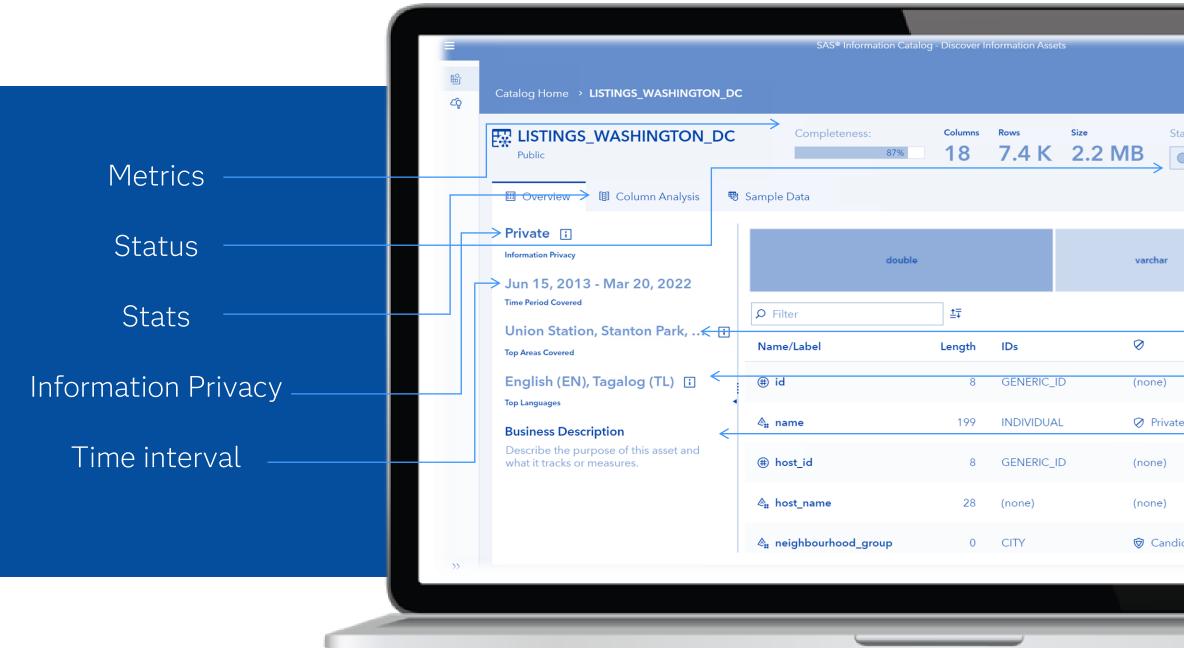


			♀ V Ø New search
			☆ Actions •
		Date and	elyzed: Jul 18, 2022 2:50 PM
11	res 🖪 :	♣ race	
	Informatio 🖡	Frequency distribution:	27,816
	Private	Black -	
	Private	Asian-Pac-Islander – Amer-Indian-Es –	
		Other -	
		0 10	0,000 20,000
		Туре:	varchar
	Sensitive	Actual type: O	String
	Private	Format: Length:	
		Minimum length:	6
		Maximum length:	19
	Sensitive	Primary key candidate: Ø	No
	Private	Logical type: O	Nominal
		Semantic type: 0	RACE
		Information privacy: O	Sensitive
	Candidate		





SAS Information Catalog



	_		-	
_	_	1 5	-)	
		♀ New search		
Status: ⑦ ● None ・	۲ ۲	Actions •		
	Date analyzed: J	un 20, 2022 8:58 PM		
	Asset type:	CAS table		Γ
	Date modified:	Jun 20,		
	Modified by:			
	Date created: Created by:	Jun 20, SAS Ad		
ate	Library: Encoding:	Public		
	Label:			
		_		
didate				
				-





Faceted Search

	rney		562 assets cataloged	
			What assets are you looking for?	
			() Help	
COLLECTIONS			Column.informationPrivacy: "private"	
			Olumn.informationPrivacy:"private"	
🕞 Recent			🕑 tables with customer	
🗈 Favorites		Name	 tables with customers tables with reviews 	
			© tables with reviews © product review	
		Customers_us	DatasetSize:>50000	
		AMAZON_REVIEWS500	DatasetSize > 50000	
		☐ food_ingredients	ColumnCount:>10	
			ColumnCount:>100	
		🕞 adult	al Name:	
		CONTACT_INFO_WITH_DUP		
		₿ joycetest	DateCreated:	
			M Status:	
			a Library:	

Import Data

1 [

\$

Ę

WELCOME

Ignite your analytics journey with SAS Information Catalog. Discover relevant and high quality data from many sources.

SEARCH

Perform free-text searches using standard search, or search on specific attributes using syntax search.

EXPLORE ASSETS

View table and column properties, quality metrics, and sample data for selected assets.

BUILD COLLECTIONS

Mark important assets as favorites and quickly access recent items.

TAKE ACTION

Open assets in related applications. Launch tasks to build a pipeline or create a report.

For more information, see <u>SAS Information</u> <u>Catalog User's Guide</u> .

6 items

lans

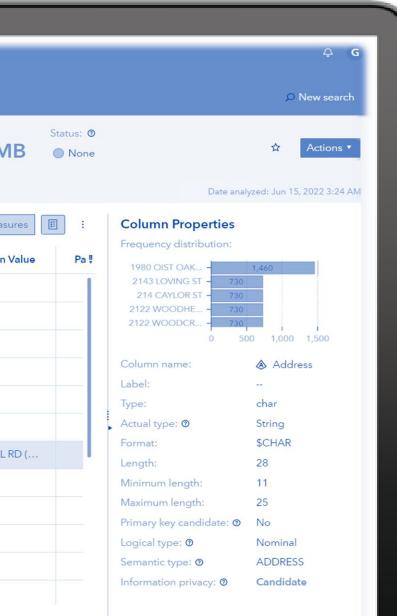
L Adm Acc



Define, discover, and profile data

Automated & prebuilt Data Quality Statistics:

			Comple	teness:	Columns Rows	^{Size} 7 K 9.6 ME
	Over Over		sis 👦 Sample Data	Descriptive Measu	rres Metadata Measures	Data Quality Measure
, blank & null	# ↑	Name	Completeness	Uniqueness	Most Common Value	Least Common Va
s etc.	1	(#) Year	100%	0%	2015	2014
	2	Month	100%	0%	12	2
	3	Day	100%	0%	28	31
	4	Date	100%	2%	12/31/2015	01/01/2014
•	5	Serial	100%	0%	2451	140
n	6	Property	100%	0%	1965	112
	7		100%	0%	1980 OIST OAK BLVD	1650 FEDERAL RD
	8	le City	100%	0%	HOUSTON	HOUSTON
	9	⊕ Zip	100%	0%	77094	77020
	10) Lat	100%	0%	-95.408872	-95.71878
	11	() Long	100%	0%	29.785747	29.651167
	12	Property_type	100%	0%	0	0

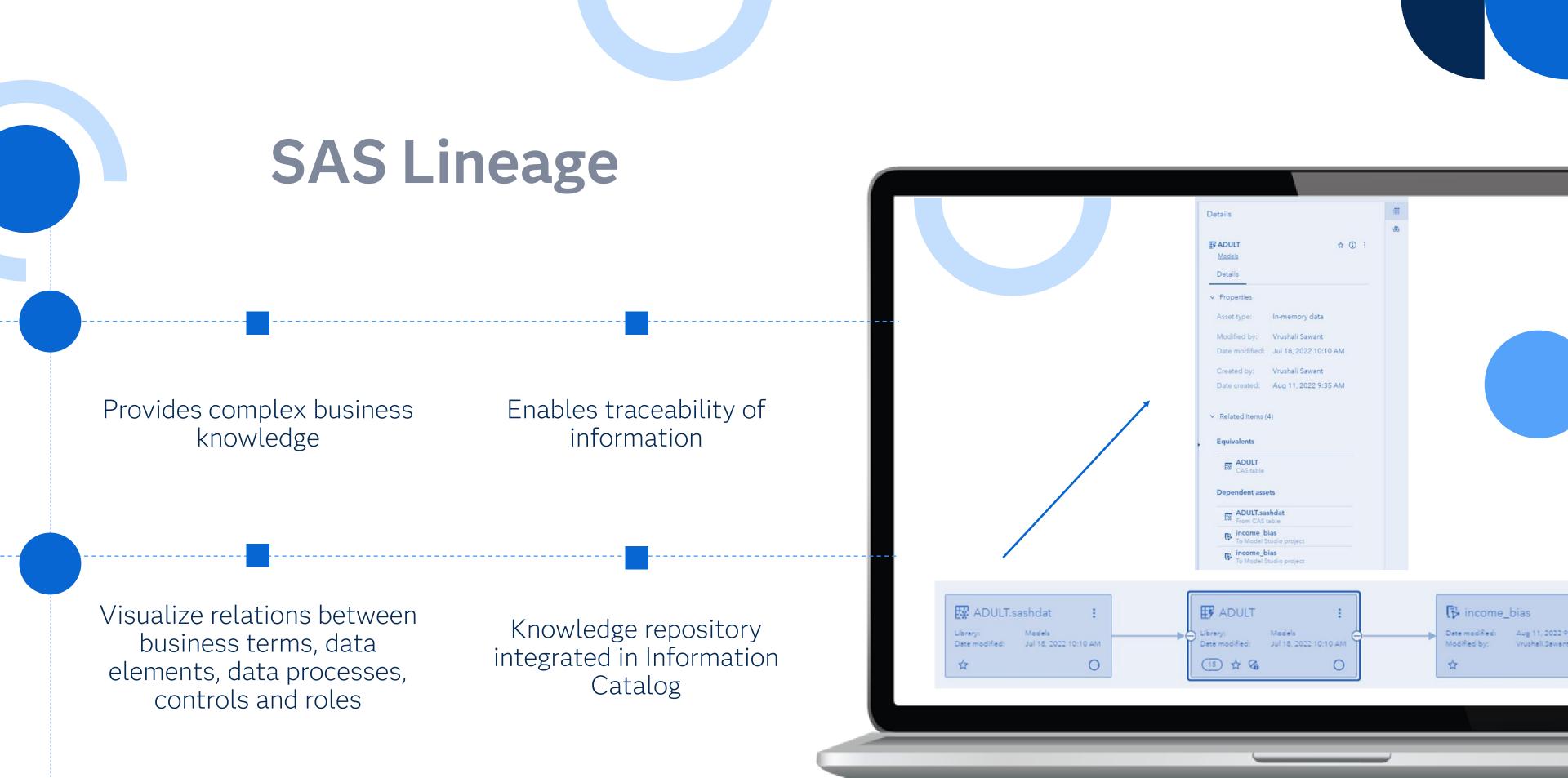


Frequency Distribution

Referential Integrity

Redundant Analysis





Copyright © SAS Institute Inc. All rights reserved.

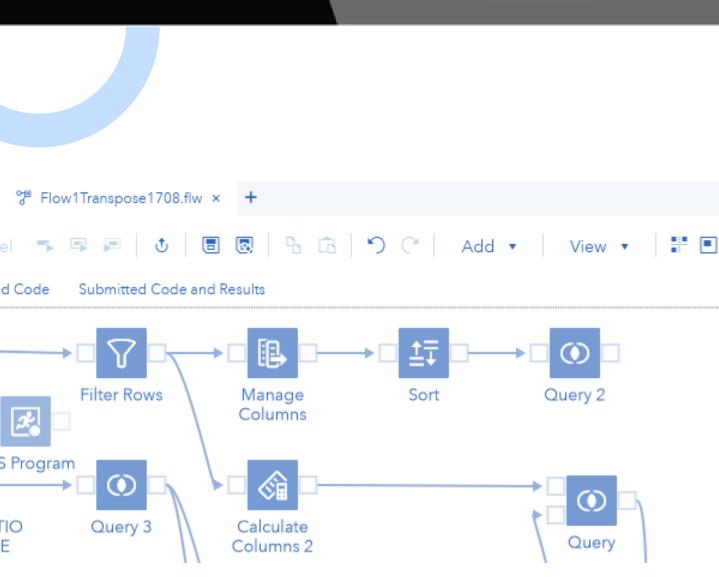


SAS Studio Overview

- Access virtually all data sources
- Extract, cleanse, transform, conform, aggregate, load and manage data.
- Support data warehousing, migration, synchronization, federation and provisioning initiatives.
- Support both batch-oriented and real-time master data management solutions.
- Create reusable data integration services in support of service-oriented architectures and data governance.

🕼 Start	Page
▶ Run	Cance
	Generated
TRZ_U	









Steps 1 🔭 💼 🛛 🕼	
© Type to filter list	SAS Studio
> type to internat	
SAS Steps : Shared	
- 🖆 Data (Input and Output)	
E Export	
⊞ File	Interactive data
🔄 Import	
📰 Table	Enterprise conne
Develop	
Python Program	
🛃 SAS Program	Data profiling
Transform Data	
🕼 Branch Rows	Extraction, trans
Calculate Columns	
♀ Filter Rows ♀ Insert Rows	
→- Insert Rows III- Manage Columns	Metadata manag
Query	
0 Rank Data	Migration and sy
 Remove Duplicates 	I'ligration and sy
≟∓ Sort	
🗗 Transpose Data	Works with both
🖆 Integrate	
12 Execute Decisions	
忆, Implement SCD	Leverage/reuse
Load Table	
Y Merge Table	Work collaborat
Enrichment Operado Data	
 Geocode Data Verify & Geocode Addresses - Logate 	
 Verify & Geocode Addresses - Loqate Verify Email Addresses - Logate 	
Convertiny Entail Addresses - Edgate	

Overview

- integration development environment
- ectivity

- sformation and load (ETL) and (ELT) ready
- Igement
- ynchronization
- h CAS and SPRE service context
- work by others (GitHub repository)
- tively in groups or teams



Enable Data Quality with SAS Studio

The Data Management solution comes with world-class data cleansing and enrichment capabilities

Improve the quality of source data

		≛ •	↑ 🗉
		р Туре	to filter list
		SAS Snip	pets : My Snippets
▶ Run	🔳 Cancel 🦐 📮 🖉 / 🗐 🐻 🕒 🖻 🖻 🏷 🤇 Add 🔹 🛛 View 🔻	▶ 💼 Cat	alogs
Flow	Generated Code Submitted Code and Results	▶ 🖪 Dat	a
		4 📫 Dat	a Quality
		E) .	Apply Schemes
		E)	Calculate Matchcodes
Parse	e Data Standardize Perform Data Identification	E;	Case Data
		-	
		Ξ;	Cluster (Entity Resoluti
		-	Cluster (Entity Resolut Extract Data
🔏 Per	form Identification Analysis	₿;	
🔏 Peri	form Identification Analysis	E) E)	Extract Data
		E) E) E)	Extract Data Guess Locale
Code	Node Notes /************************************		Extract Data Guess Locale Identify Gender Parse Data
Code	Node Notes /************************************		Extract Data Guess Locale Identify Gender Parse Data Perform Identification
Code 1 2	Node Notes /************************************		Guess Locale Identify Gender Parse Data Perform Identification Run Data Profiling - Pro
Code 1 2 3	Node Notes /************************************		Extract Data Guess Locale Identify Gender Parse Data Perform Identification
Code 1 2 3 4	Node Notes /************************************		Extract Data Guess Locale Identify Gender Parse Data Perform Identification Run Data Profiling - Pro

Spippote

ion) - Proc DQMATCH

Analysis

C DATAMETRICS

Standardization Matching Clustering Surviving record Country specific

