

SAS Platform Overview: Data Management (eli miten taipuu datahommat SAS Viyalla)

3.4.2024

Jarno Lindqvist

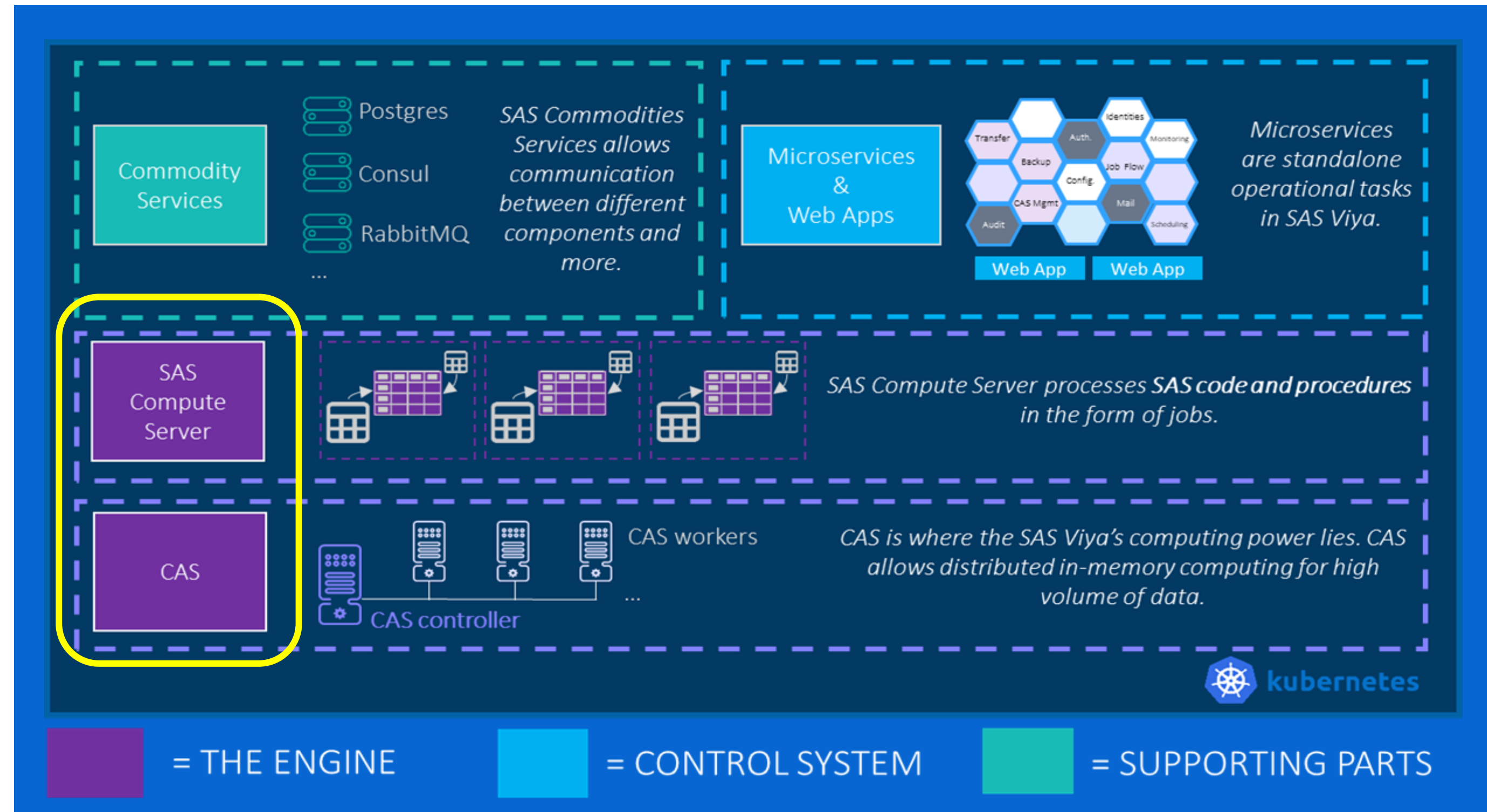
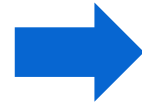
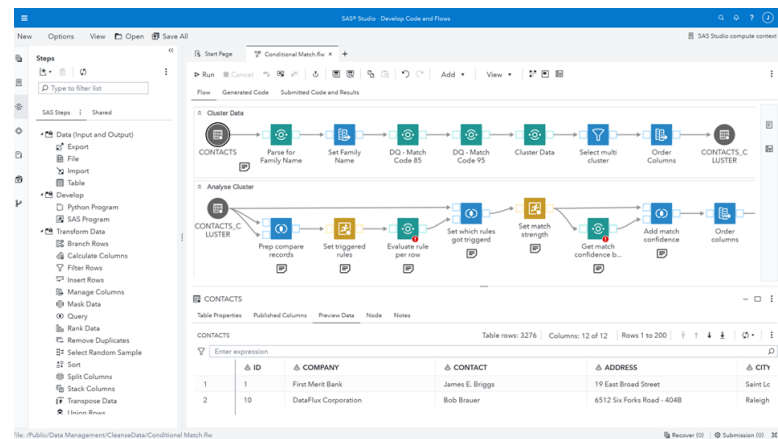


Alkuun pari faktaa SAS Viyasta

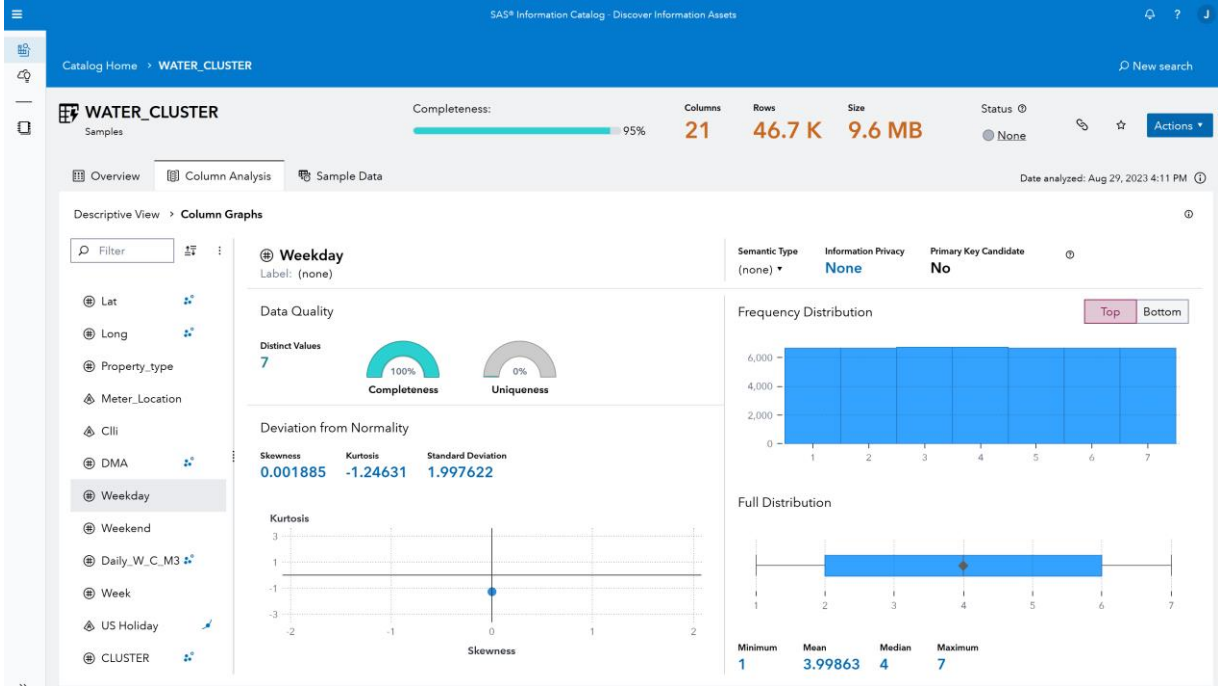
- SAS Viya on muistinvaraiseen prosessointiin kykenevä alusta joka on rakennettu skaalautuvaksi ja pilvinatiiviksi.
- SAS Studio on ensisijainen väline hallita SAS9:sta tuotua sisältöä, sillä sinne voi tuoda mm. SAS-koodin, Enterprise Guide -projektit sekä Data Integration Studio -jobit
- SAS Viya suorittaa koodia sekä perinteisellä Compute-serverillä, että muistinvaraisella CAS-serverillä
- Compute-serverillä ajettaessa SAS Viya on SAS-koodin suhteen lähes täysin yhteensopiva SAS9 koodin kanssa
- Muistinvaraisen CAS-prosessoinnin täysimääräinen hyödyntäminen vaatii joitakin SAS koodiin tehtäviä muutoksia. Muistinvarainen suoritus tehostaa prosessointi-intensiivisten työvaiheiden suoritusta merkittävästi

CAS ja Compute – SAS Viyan turbomootorit

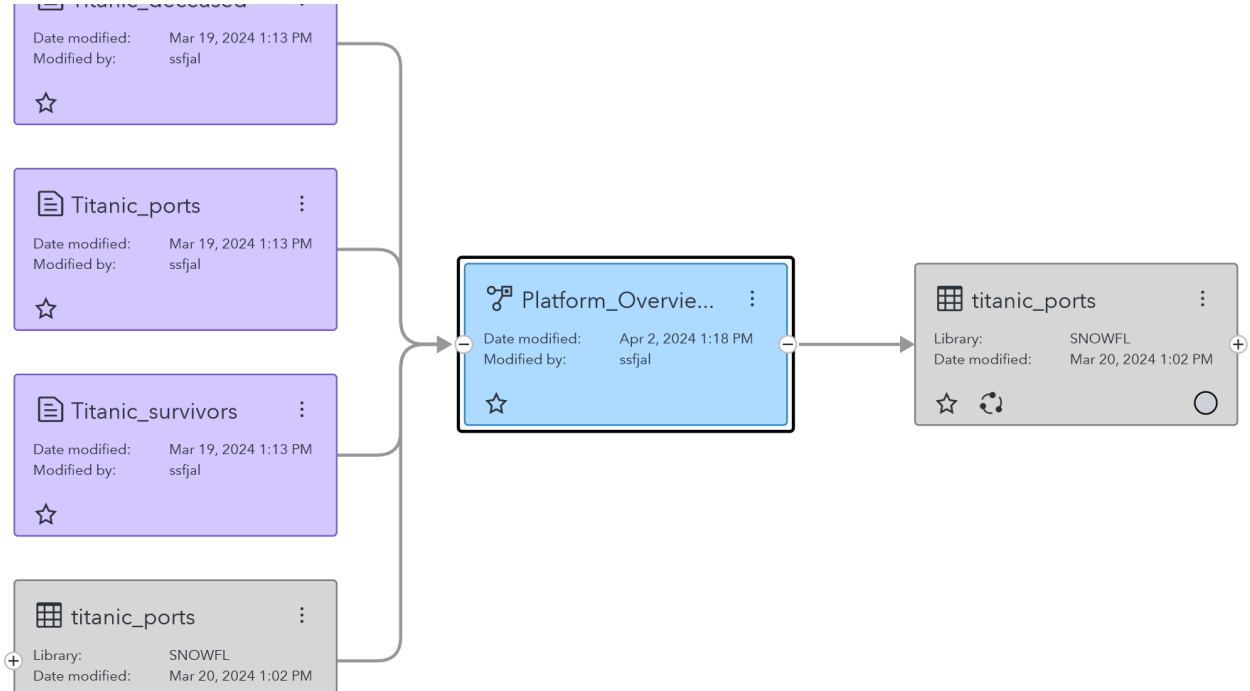
- SAS Studio voi hyödyntää sekä Compute-, että CAS-serveriä, myös saman workflown sisällä



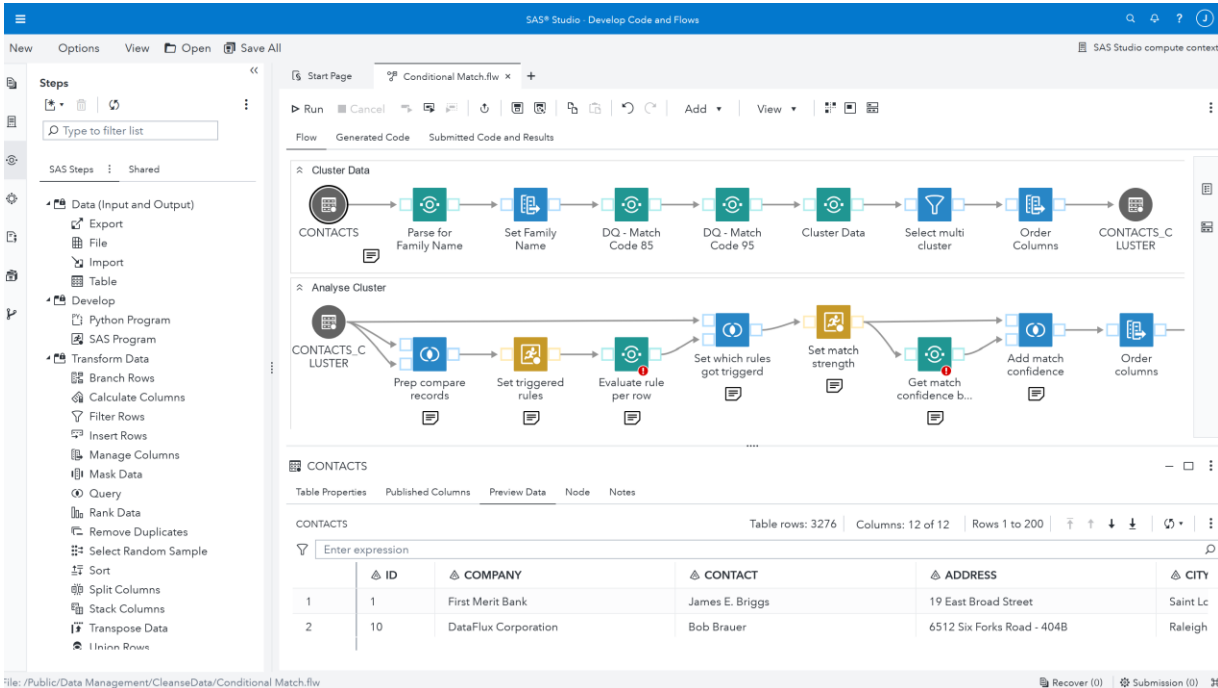
Datan hallinnan käyttöliittymät Viya 2024.02



SAS Information Catalog



SAS Lineage



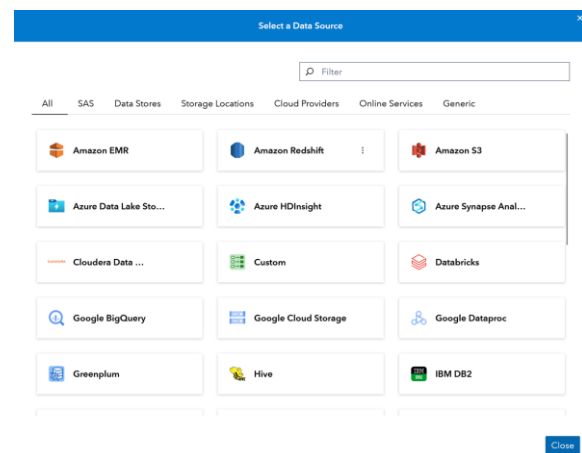
SAS Studio

The screenshot displays the SAS Data Explorer interface for a table named '_customers.sashdat'. The table has 24 columns and 57.8 MB of data. A sample of 100 rows is shown, with columns including City, Continent, Postal_Code, State_Province, Street_Name, xyContinent, and Custom. The data includes entries for various cities like Madrid, Las Palmas, Barcelona, and Granada.

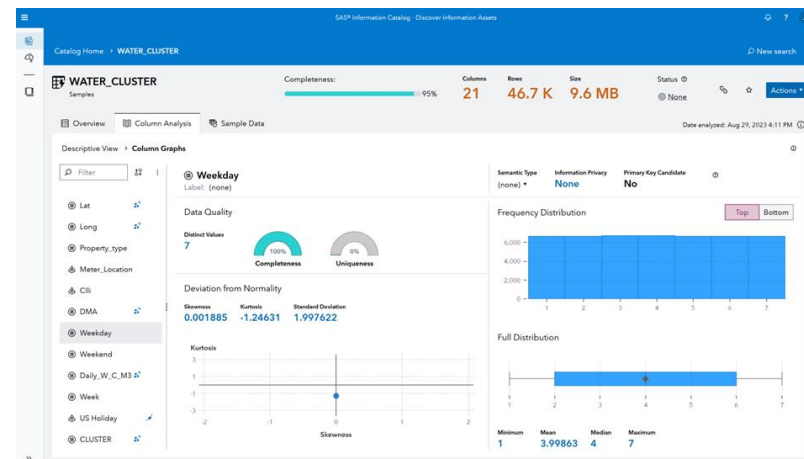
SAS Data Explorer



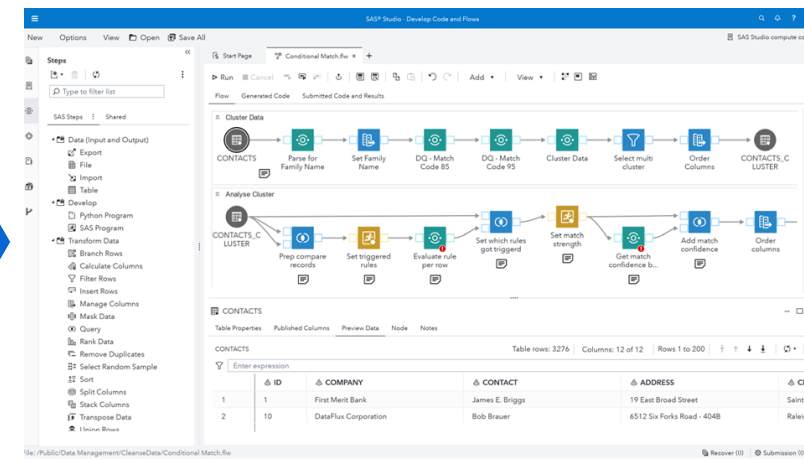
Jokaisella SAS Viya –käyttöliittymällä on paikkansa dataketjussa (ei aina tässä järjestyksessä)



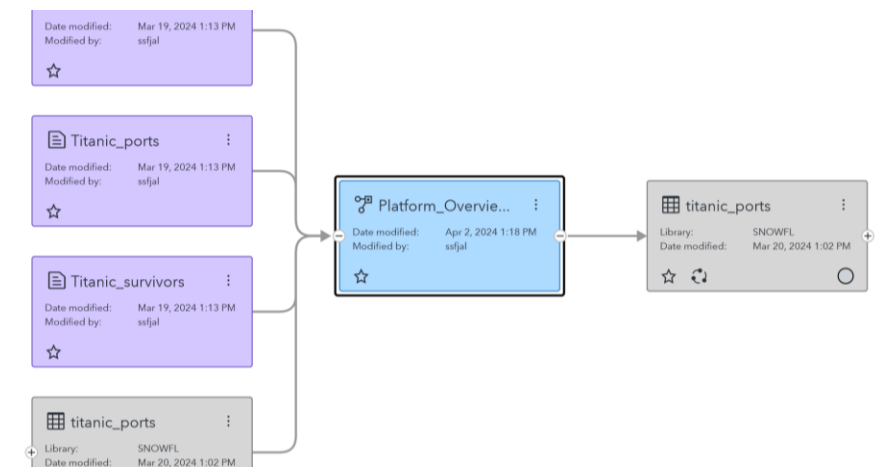
SAS Data Explorerin avulla määritellään CAS-kirjastot, SAS Studioon avulla vastaavasti Compute-kirjastot



SAS Information Catalog kahlaa valitut kirjastot läpi jonka jälkeen data on helposti löydettävissä ja sen kelpoisuus voidaan arvioida



SAS Studiossa tehdään datan muokkaus ja yhdistely, joko koodilla tai graafisen Flow-työkalun avulla



SAS Lineage visualisoi datan elinkaaren ja kertoo mihin dataobjekteihin se on linkitetty

Develop Code and Flows
Discover Information Assets
Manage Data
Explore Lineage

Develop Code and Flows
Discover Information Assets
Manage Data
Explore Lineage

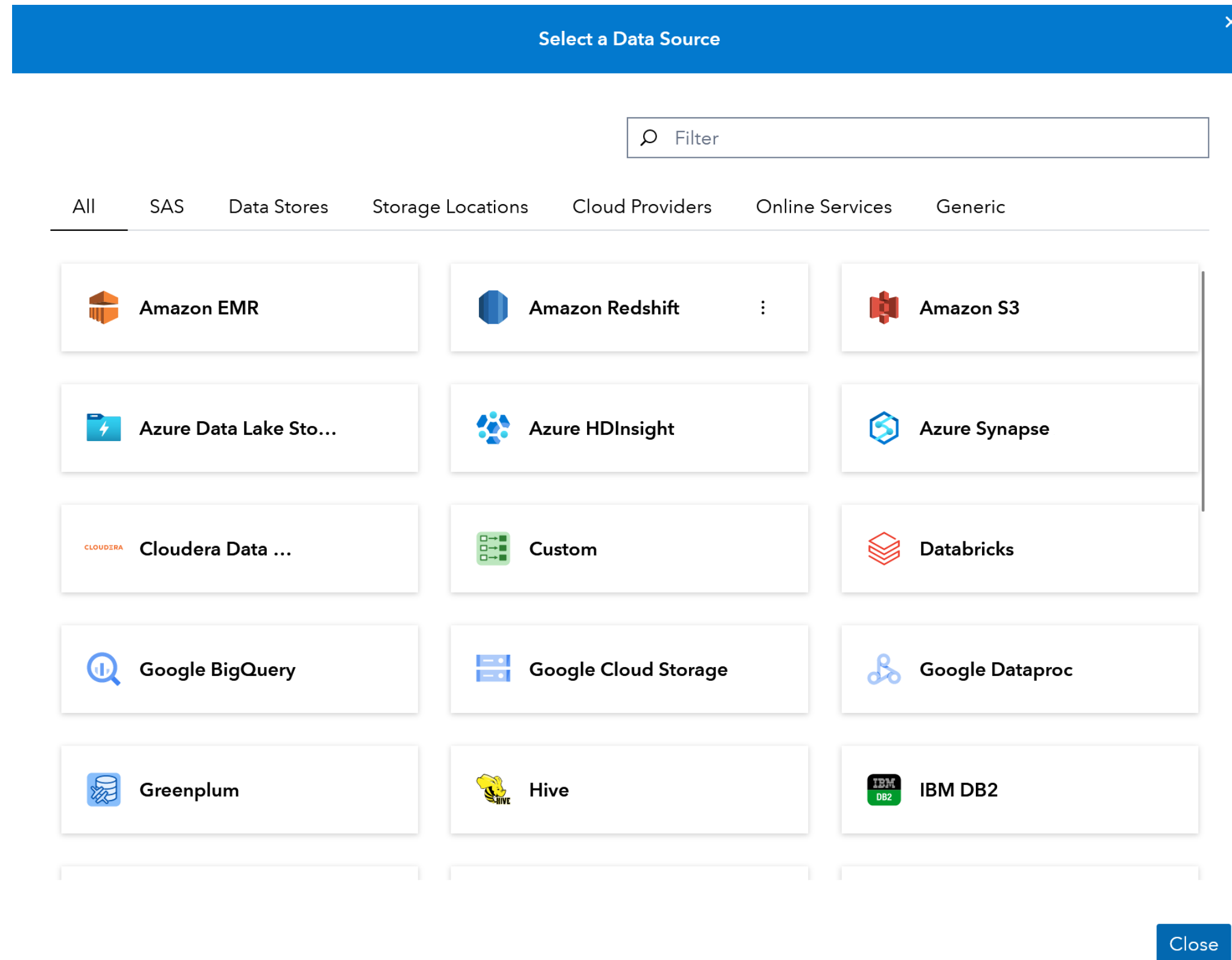
Develop Code and Flows
Discover Information Assets
Manage Data
Explore Lineage

Develop Code and Flows
Discover Information Assets
Manage Data
Explore Lineage

Kuten aina ennenkin, SASilla pääsee kiinni kaikkiin datalähteisiin



Nyt uusien datalähteiden määrittely on selkeää



SAS Viya sai loppuvuodesta yhtenäisen React-pohjaisen Data Explorer käyttöliittymän josta hallitaan niin tietokantalähteet, kuin pilvessä tai levyllä olevat datat

Dataa helppo tuoda SAS-kirjaston tai importin kautta

New Library Connection

Connection name: *

JarnoLib

Library name (libref): *

jarnolib

Library type:

SAS Base Engine

Assign and connect to data sources at startup

Properties Connection Options

Physical name support: *

/export/home/users/ssfjal/jarnolib

Library attributes:

Input encoding:

Output encoding:

Output representation:

Test connection

Kirjaston lisääminen SAS Studiossa
(huom. levyoikeudet pitää olla!!!)

Contacts.csv

Step 1 - File Import Options

File Specifications Advanced

Input file delimiter:

Comma

Scanned rows:

20

Locale:

Enter locale

Source encoding: *

UTF-8

First row contains column names

Convert character columns to variable size

Step 2 - Output File Options

Name: *

Contacts

Location (library): *

cas-shared-default/CASUSER(ssfjal)

Label:

Enter label

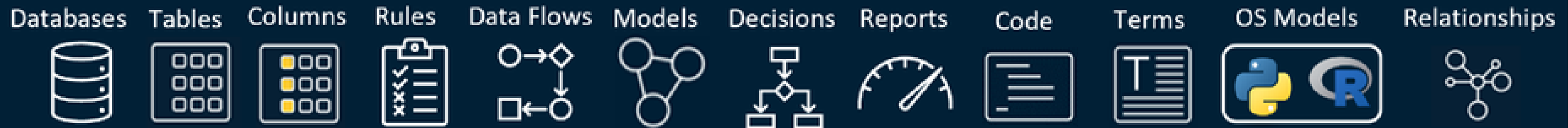
Format:

sashdat

CSV-import SAS Data Explorerissa

*Yksinkertaisin tapa
tuoda esim. CSV tai
Excel data SAS
Viyaan on pudottaa
se hiirellä Viya-
ikkunan päälle*

SAS Information Catalog

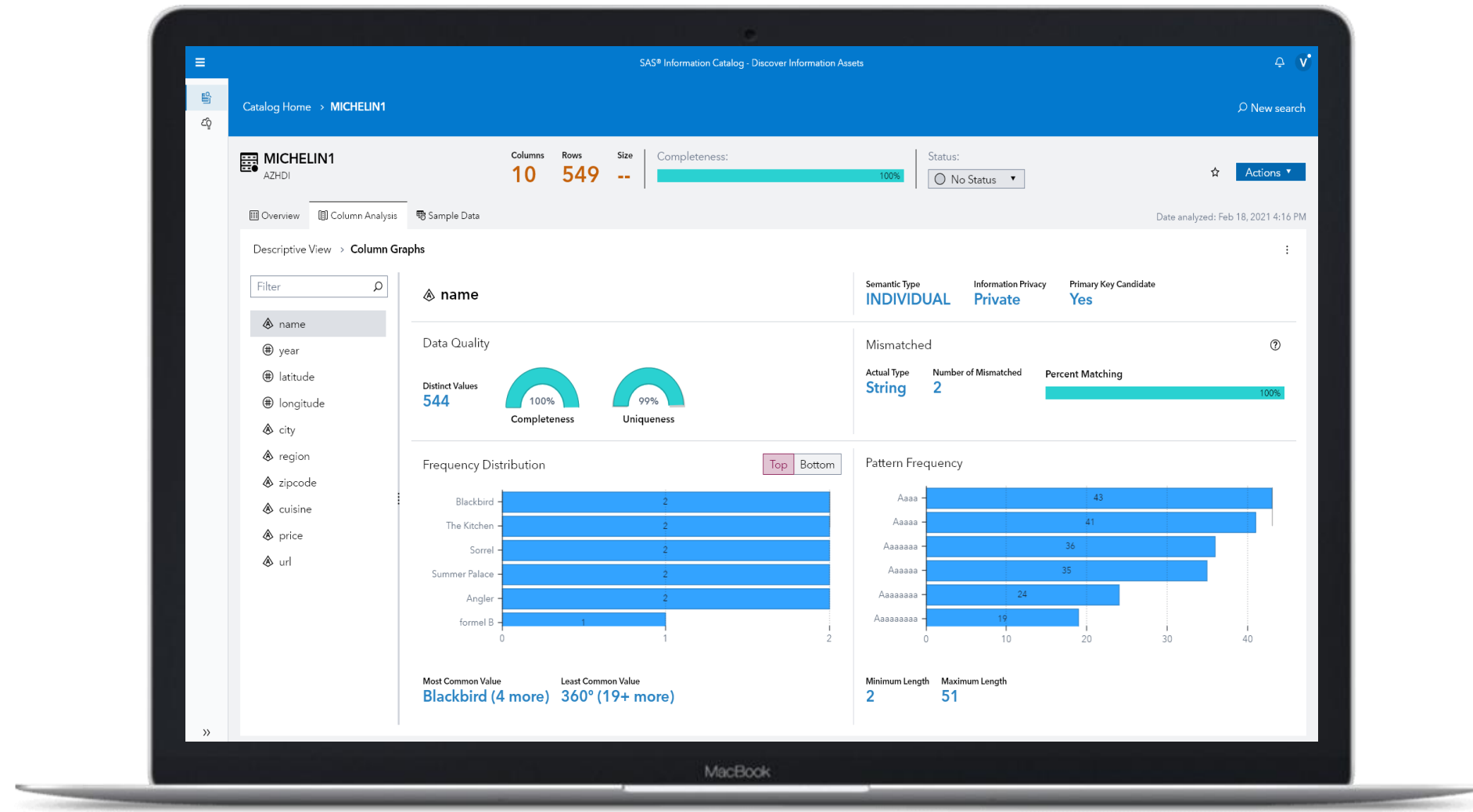


- Data Governance SAS Viyaassa on kiinteä osa SAS Viya -alustaa
- Käyttöliittymät ovat SAS Information Catalog ja SAS Lineage
- Näihin pääsee kätevästi muista SAS Viya – käyttöliittymistä
- Information Catalog auttaa löytämään datan ja arvioimaan sen soveltuvuuden käyttötarkoitukseen
- Information Catalog tarjoaa keskitetyn dashboardin kaikkiin data-resursseihin, siis taulut, workflowt, raportit, mallit ja paljon muuta!
- SAS Lineage kuvaa datapolun alusta loppuun

The screenshot shows the 'Actions' menu for a data asset. The 'Discover information assets' option is highlighted with a blue box. Below the menu, a search bar contains the text 'cars'. A table of search results is displayed below the search bar.

Name	Library	Date Modified	Modified By
BATCH_SEARCH	Public	Mar 14, 2024 12:09 PM	sukhsn
MARITIME	Public	Mar 14, 2024 12:09 PM	sukhsn
PROB_DEP	Public	Mar 14, 2024 12:09 PM	sukhsn
PROB_LOC	Public	Mar 14, 2024 12:09 PM	sukhsn
PROB_TOTAL	Public	Mar 14, 2024 12:09 PM	sukhsn
SALVAGE_MAP	Public	Mar 14, 2024 12:09 PM	sukhsn
SEARCH_ANIM	Public	Mar 14, 2024 12:09 PM	sukhsn

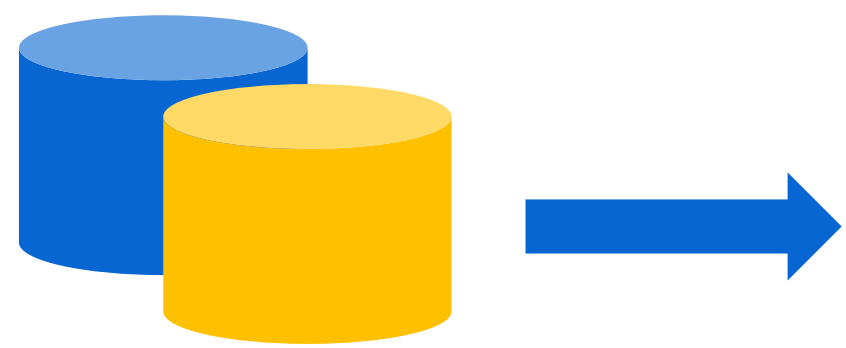
Information Catalog löytää datan




Hakuportaali, dashboard ja laatuanalyysit

Haku-palvelut

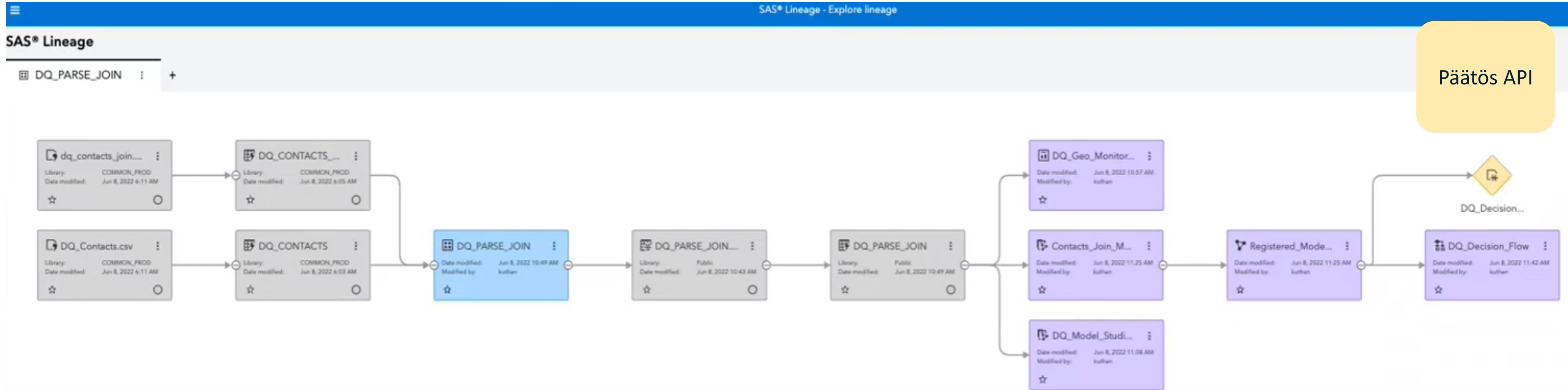
Datan esianalysointi



Automaattinen hakuagentti 

Metadatan tallennus (ei siis datan) 

Lineage kuvaa datapolun alusta loppuun



Lähteet

CAS taulut

Datan yhdistely

Validoitu data

Välitulokset

Mallit ja raportit

Rekisteröity malli

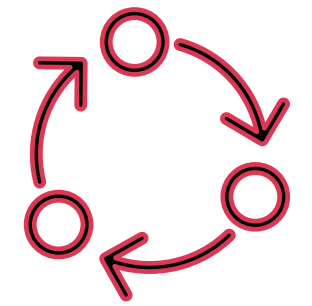
Päätös flow

SAS Studio on datanhallinnan ydin Viyassa

SAS Enterprise Guide



SAS Studio



SAS Data Integration Studio



Base SAS ja SAS-koodi



Koodaa

SAS-koodieditori,
jossa ennakoiva
tekstinsyöttö ja
pikalinkit help-
sivuihin

Rakenna

SAS-koodi, valmiit
taskit, Flow stepit,
kyselyt

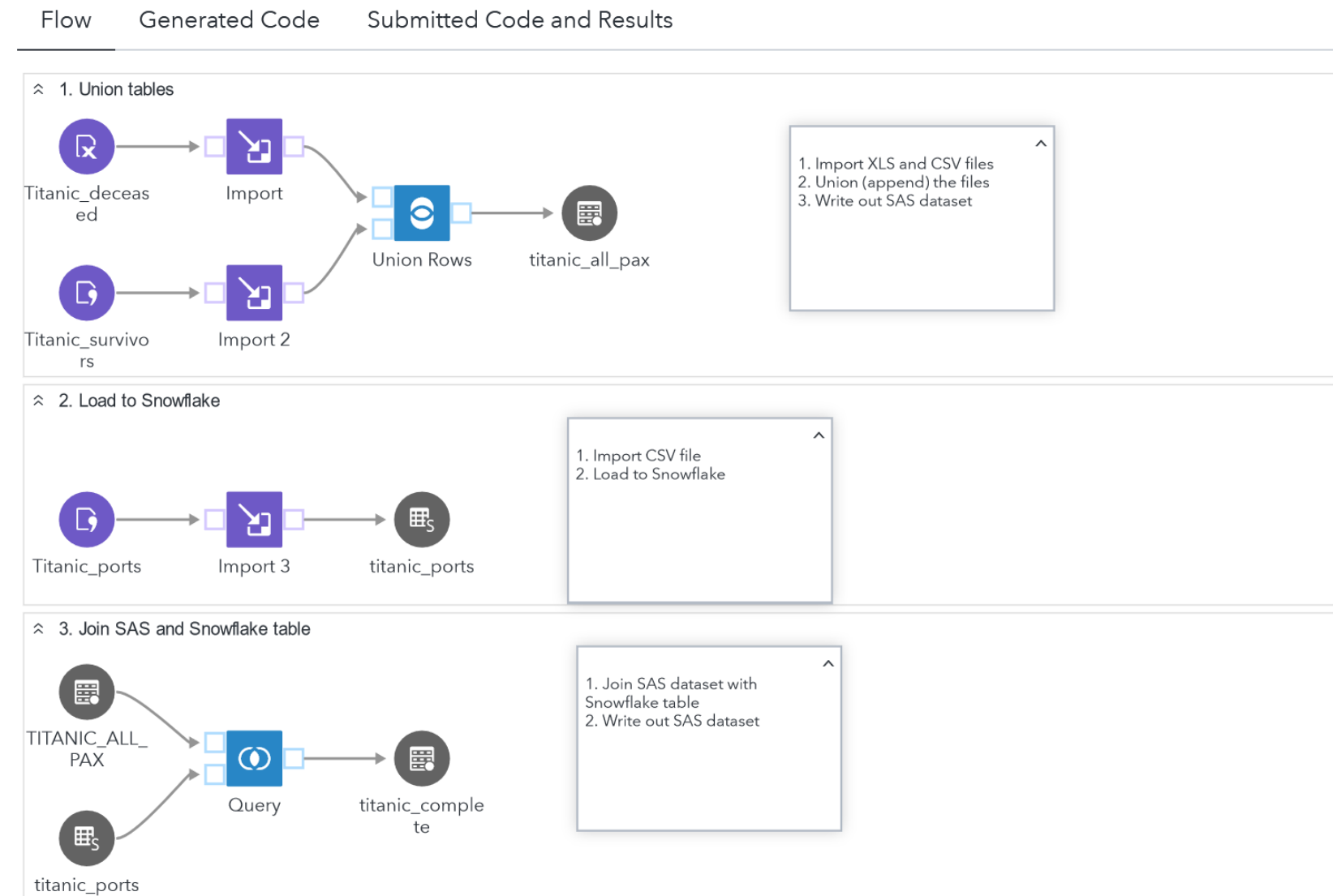
Tee yhdessä

Git-integraatio,
jaetut custom stepit
ja taskit

https://www.sas.com/en_us/software/studio.html

SAS Studioissa sekä koodin kehitys että Flow

```
Code Log
10
11 options validvarname=any;
12 ods noproctitle;
13 ods graphics / imagemap=on;
14
15 ⊖ proc sort data=SASHELP.CARS out=WORK.TempSorted5215;
16     by Origin;
17 run;
18
19 /* Scatter plot matrix macro */
20 ⊖ %macro scatterPlotMatrix(xVars=, title=, groupVar=);
21     proc sgscatter data=WORK.TempSorted5215;
22         matrix &xVars / %if(&groupVar ne %str()) %then
23             %do;
24                 group=&groupVar legend=(sortorder=ascending) %end;
25     ;
26     title &title;
27     by Origin;
28 run;
29
30 title;
31 %mend scatterPlotMatrix;
32
33 %scatterPlotMatrix(xVars=EngineSize Cylinders Horsepower,
34     title="Scatter plot matrix grouped by Make", groupVar=Make);
35 %scatterPlotMatrix(xVars=EngineSize Cylinders Horsepower,
36     title="Scatter plot matrix grouped by Model", groupVar=Model);
37
38 ⊖ proc datasets library=WORK noprint;
39     delete TempSorted5215;
40 run;
```



Koodia voi joko kirjoittaa editorilla itse tai generoida valmiilla SAS-taskeilla

SAS Studio Flow generoi SAS-koodia, eli työtapa on visuaalinen ja samalla itsedokumentoitu

Code ja low code yhdistyvät helposti

“Code to Flow” –napilla saa SAS-koodin helposti stepiksi “flow-puolelle”

The screenshot illustrates the 'Code to Flow' process in SAS Studio. On the left, the 'Code to Flow' button is highlighted with a red circle. A blue arrow points from this button to the 'Recode Values' node in a flow diagram. The flow diagram shows a sequence of steps: 'Recode Values' (a yellow node with a person icon), 'CARS_NEW' (a grey node with a database icon), and 'Query' (a blue node with a magnifying glass icon). A second blue arrow points from the 'Recode Values' node to the SAS code editor. The code editor shows the generated SAS code for the 'Recode Values' node, which includes a data step to create a new variable 'Make_new' based on the 'Make' variable from the 'SASHELP.CARS' dataset.

Code to Flow

Code Log Output Data (1)

Errors (0) Warnings (0) No

```
77
78 /*
79 *
80 * Task code generated by S/
81 *
82 * Generated on '4/19/22, 1:
83 * Generated on server 'sas-
84 * Generated on SAS platform
85 * Generated on SAS version
86 * Generated on browser 'Mo:
87 * Generated on web client
87 ! 'https://ssemmonthly.demo.sa:
88 */
89
90 data work.CARS_NEW;
91 length Make_new $ 13;
92 set SASHELP.CARS;
93
```

Flow Generated Code Submission

Recode Values

Code	Node
1	/*
2	*
3	* Task code generated by SAS® Studio 6.0
4	*
5	* Generated on '4/19/22, 11:29 AM'
6	* Generated on server 'sas-launcher-de0f8624-0
7	* Generated on SAS platform 'Linux LIN X64 5.4
8	* Generated on SAS version 'V.04.00M0P011322'
9	* Generated on browser 'Mozilla/5.0 (Windows I
10	* Generated on web client 'https://ssemmonthly
11	*/
12	
13	data work.CARS_NEW;
14	length Make_new \$ 13;
15	set SASHELP.CARS;
16	
17	select (Make);
18	when ('Volkswagen') Make_new='VW';
19	otherwise Make_new=Make;
20	end;
21	run;

Vakio-stepit SAS Studiossa 2024.02

Visualize Data

- Bar Chart
- Bar-Line Chart
- Box Plot
- Bubble Map
- Bubble Plot
- Choropleth Map
- Heat Map
- Histogram
- Line Chart
- Pie Chart
- Scatter Map
- Scatter Plot
- Series Plot
- Text Map

Examine Data

- Characterize Data
- Describe Missing Data
- List Data
- List Table Attributes

Develop

- Python Program
- SAS Program

Integrate

- Execute Decisions
- Implement SCD
- Load Table
- Merge Table

Statistical Process Control

- Analysis of Means
- Control Charts
- Pareto Analysis

Machine Learning

- Moving Window Principal Component Analysis

Manage Models

- Register Python Model
- Register SAS Model

Data (Input and Output)

- Export
- File
- Import
- Table

Enrichment

- Geocode Data
- Verify & Geocode Addresses - Loqate
- Verify Email Addresses - Loqate
- Verify Phone Numbers - Loqate

Statistics

- Coin Toss Simulation
- Combinations
- Correlation Analysis
- Dice Roll Simulation
- Distribution Analysis
- One-Way Frequencies
- Permutations
- Poker Hand Probability
- Same Birthday Probability
- Summary Statistics
- t Tests
- Table Analysis

Transform Data

- Branch Rows
- Calculate Columns
- Filter Rows
- Insert Rows
- Manage Columns
- Mask Data
- Query
- Rank Data
- Remove Duplicates
- Select Random Sample
- Sort
- Split Columns
- Stack Columns
- Transpose Data
- Union Rows

Data Quality

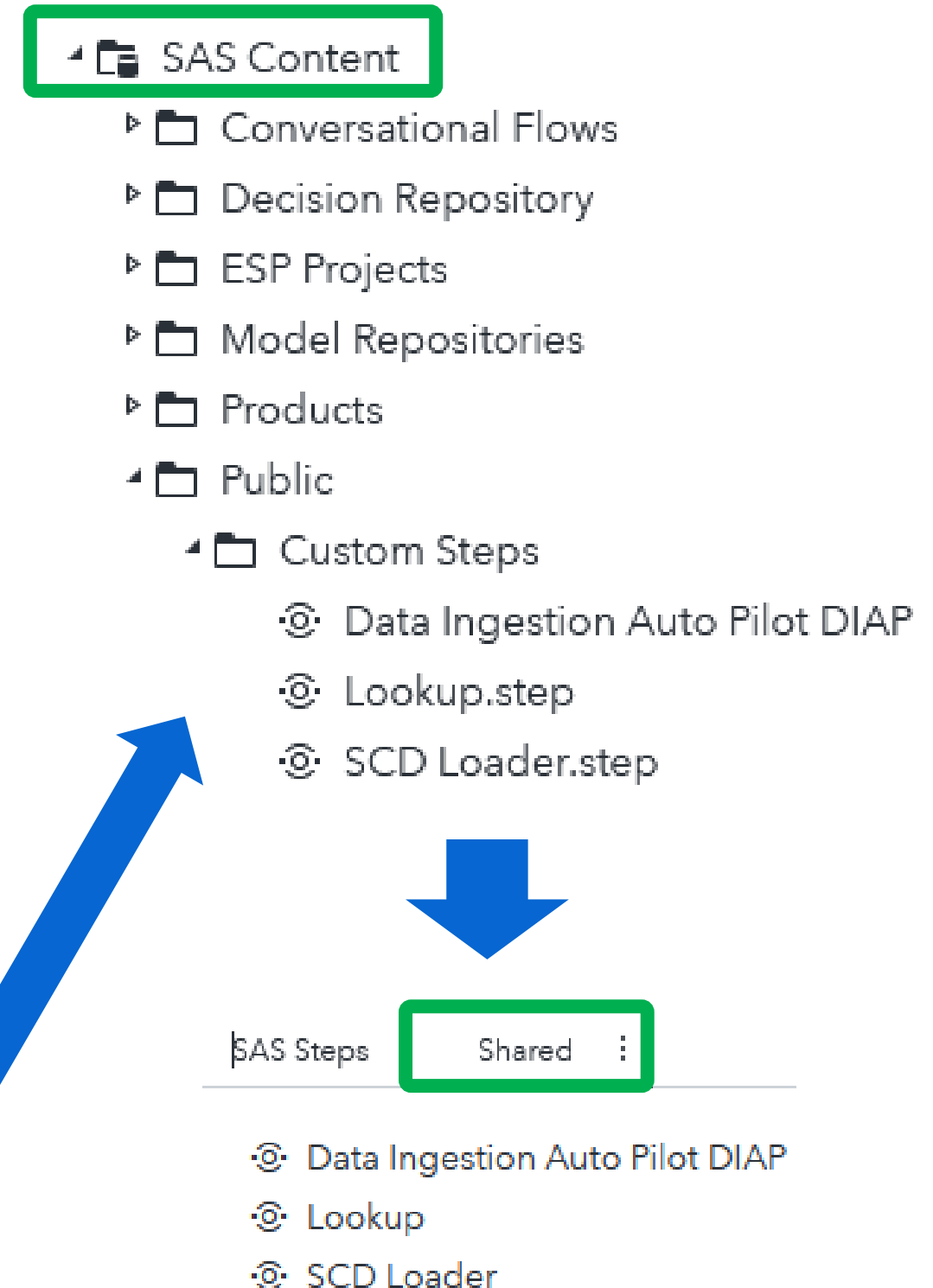
- Clean Data
- Match Codes
- Parse Data

SAS Studio Custom stepit

Custom step on SAS Studio Flown uudelleen käytettävä, muokattava komponentti:

- Kehittäjä rakentaa logiikan ja upottaa sen Custom Stepiin
- Custom step “piilottaa” koodin sellaiselta käyttäjältä joka haluaa käyttää graafista käyttöliittymää
- Nopeuttaa kehitystä tarjoamalla valmiiksi mietittyjä valikkokomponentteja
- Custom Stepin jakaminen on helppoa, se tarvitsee vain tallentaa yhteisesti näkyviin kansioihin niin se on kaikkien käytettävissä
- Stepejä voi jakaa oman organisaation sisällä tai muiden tekemiä *ladata SASin GitHub-repositorysta*

<https://github.com/sassoftware/sas-studio-custom-steps>



Miksi SAS Studio ja Flow on hyvä juttu

- SAS9-koodin suoritus SAS Studiossa on suoraviivaista, sillä valtaosa koodista toimii sellaisenaan SAS Viyan Compute-serverillä
 - Toki kirjasto- ja tiedostoviittaukset tulee muokata vastaamaan uutta ympäristöä
- Tukee all code / low code / no code -työtapoja
- Toimii yhteen SAS-taulujen, CAS-taulujen, tietokantojen ja pilvistoragen (S3, ADLS2 jne.) kanssa
- SAS Studio selainkäyttöliittymä ei vaadi asennuksia tai päivityksiä
- Ominaisuuksia voi täydentää Custom Stepeillä (GitHubista lataamalla tai itse rakentamalla)
- Tukee myös Python-koodia
- SAS9:n Data Integration Studio -jobit ja Enterprise Guide -projektit voidaan migroida SAS Studioon
 - SAS Content Assessment -työkalun avulla tiedät etukäteen mikä migroituu ja mikä vaatii muutoksia

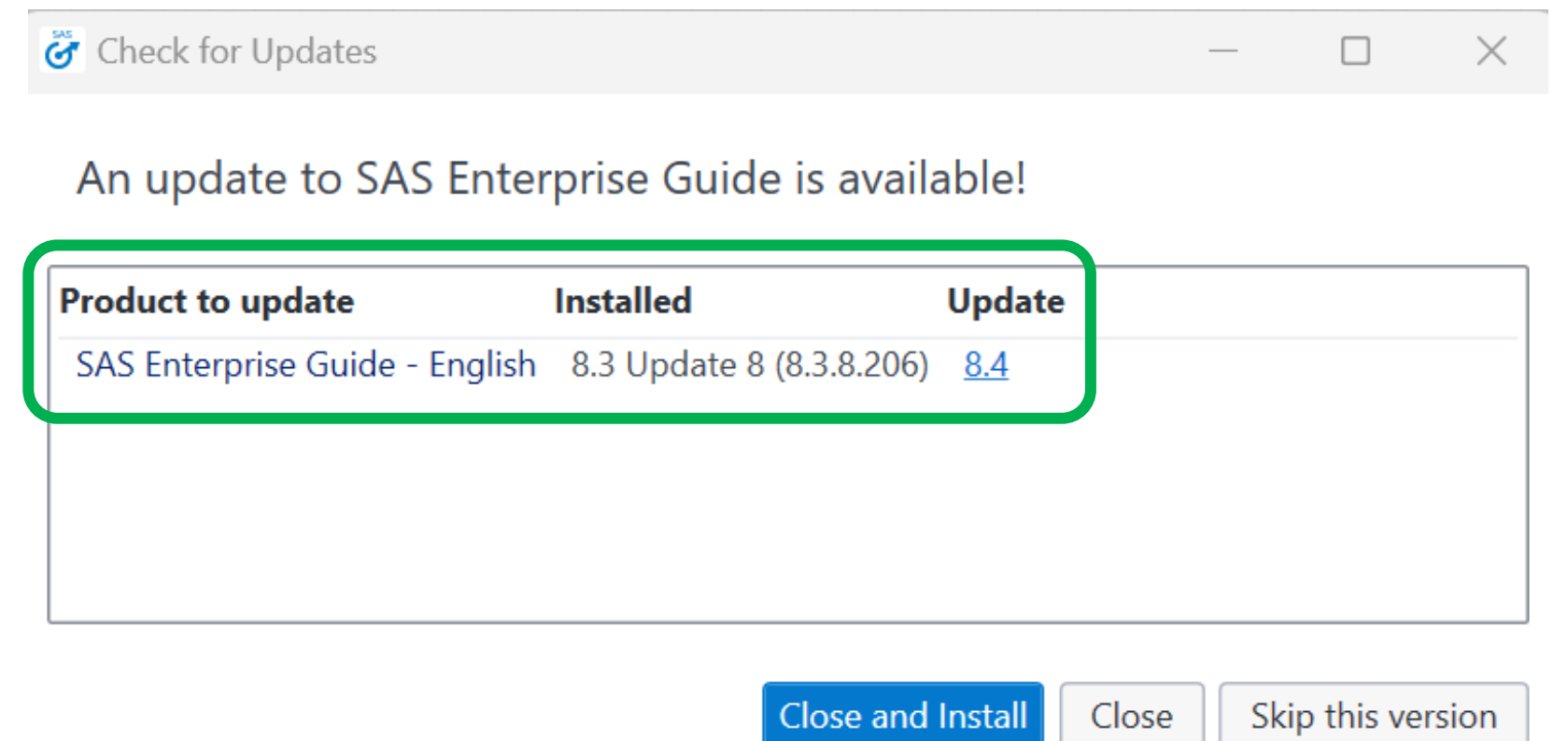
SAS EG:n Viya 4 yhteensopiva versio 8.4 julkaistiin 19.3!

Mitä tuetaan

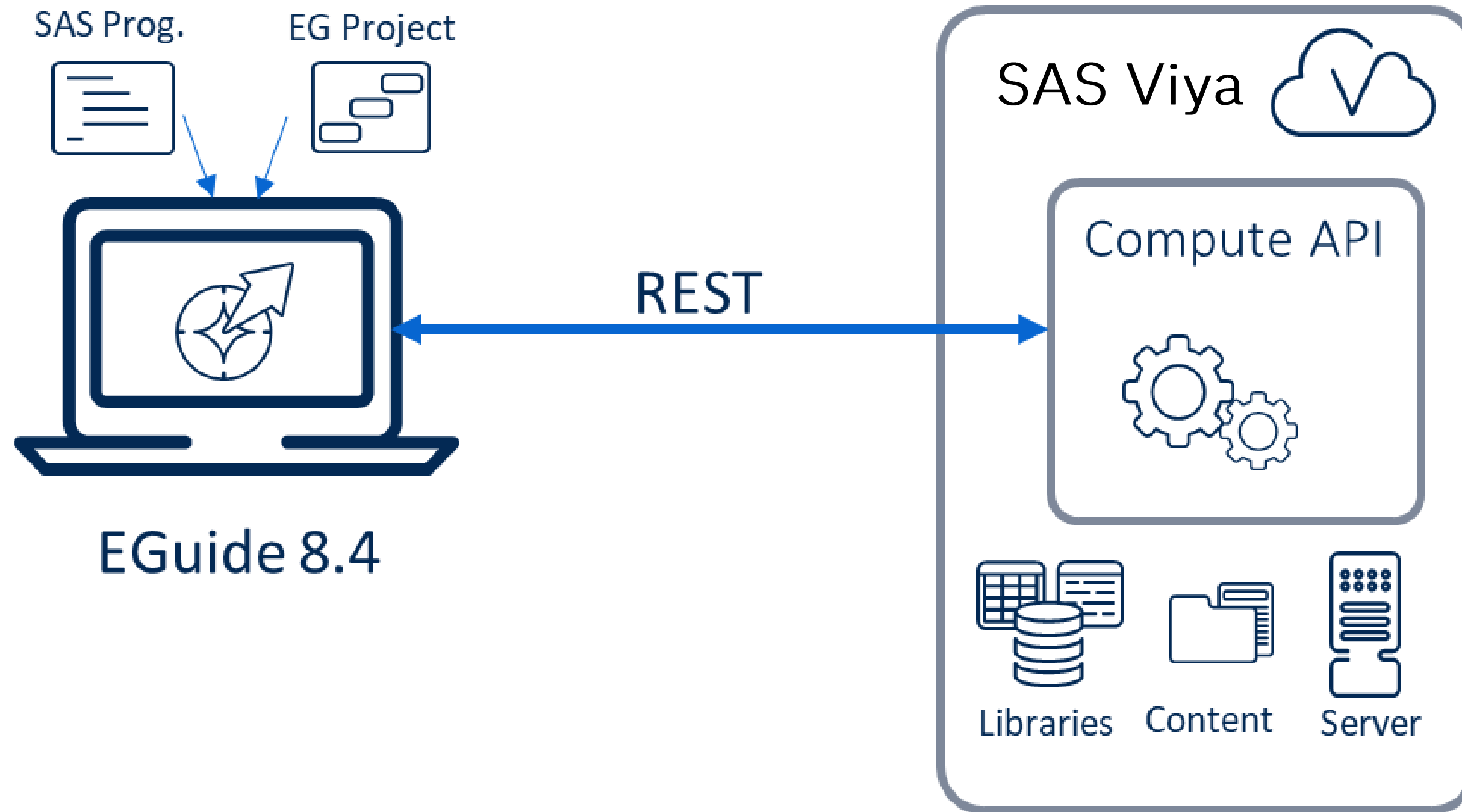
- Tukee nyt Viya 4:n Compute Serveriä
- SAS 9.4:n tuki säilyy sellaisenaan
- Viyan osalta näkyy SAS Content, SAS Libraries ja tiedostokansiot
- EG:n perustoiminnot:
 - Koodin suoritus
 - Taulujen avaus
 - Tulokset ja lokifileet
 - Promptit (poislukien SAS9 Metadattaa hyödyntävät)
 - DATA Step Debugger
- EG 8.3 päivitys klikkaamalla "Check for Updates"
 - EG 8.4 julkaisun jälkeen 19.3 päivitysikkunan pitäisi pompahtaa automaattisesti näkyviin kun vanhemman EG:n käynnistää

Mikä on ennallaan

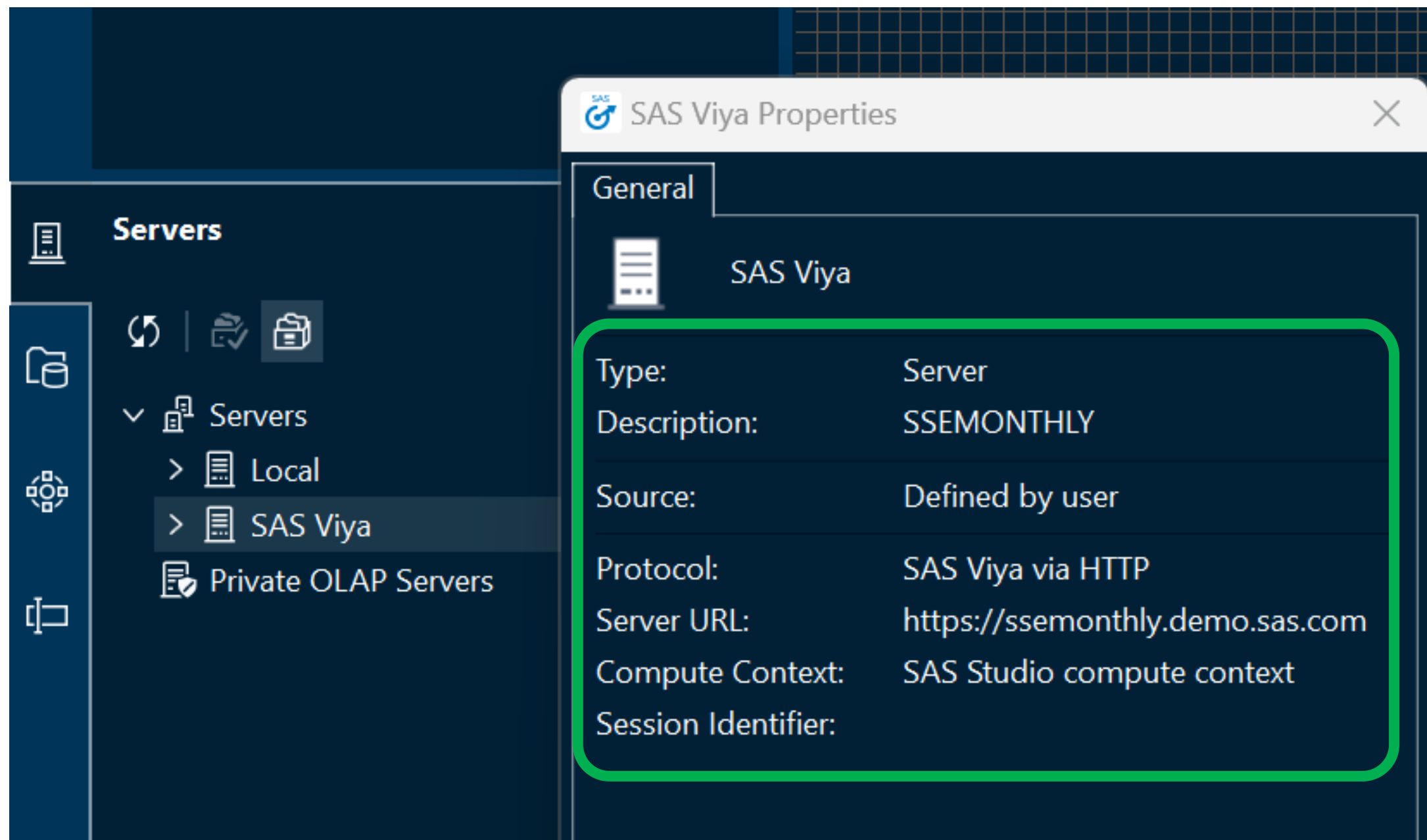
- Käyttöliittymä pysyy samana
- Ei uusia EG-taskeja
- Ei SAS9 metadata-toiminnallisuuksia SAS Viyan kanssa
- Stored prosessit
 - Ei voi ajaa sellaisenaan, mutta voidaan migroida Viyan jobeiksi



Enterprise Guide hyödyntää SAS Viyan Compute Serveriä REST-rajapinnan kautta



EG:n käyttö Viyan kanssa on hyvin samanlaista kuin ennenkin



- SAS Viya 4 palvelin näkyy serverilistalla kuten SAS9-palvelimet
- Käyttäjä voi kätevästi vaihtaa SAS 9 palvelimen ja Viyan välillä esim. kehitystä ja testailua varten
- Käytännössä ainoa ero on kommunikointi EG:n ja palvelimen välillä:
 - Viyan kanssa REST
 - SAS9:n kanssa IOM

Lisätietoa täällä:

https://www.sas.com/en_us/software/studio.html

www.sas.com

