Top 5 Gotchas for Getting Cloud Ready in SAS® Viya®

Charu Shankar, SAS Institute

.

ABSTRACT

As a sas9 programmer, you may be interested to learn about the cloud and why SAS Viya is such a big leap forward. In addition to hearing about the big picture and getting cloud savvy, you are also interested in a deeper dive in how things are going to be in SAS Viya vs SAS 9.

This paper will discuss the top 5 gotchas for getting cloud ready in SAS : from database connections to data types, exploring metadata, loading tables in memory, new engines.

All levels welcome. Some awareness of SAS 9 would be beneficial.

INTRODUCTION

SAS Viya is a cloud-enabled, in-memory analytics engine that provides quick, accurate and reliable analytical insights. The latest enhancement of the SAS platform, SAS Viya is an open, cloud-enabled, analytic runtime environment with a number of supporting services. One of those supporting services is SAS Cloud Analytic Services, or CAS. CAS provides a powerful in-memory engine that delivers blazing speed to accurately process your big data. It uses scalable, high-performance, multi-threaded algorithms to rapidly perform analytical processing on in-memory data of any size.
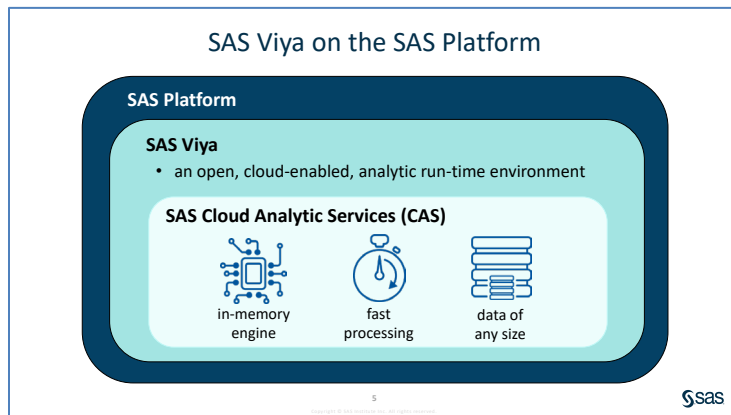
TERMINOLOGY

File is used to refer to the source data that is in a Caslib's data source. For a Caslib that uses a path-based data source, this is natural. For a caslib that uses a database as a data source, the tables in the database are referred to as files.

Table is used to refer to in-memory data. After a file (using the preceding definition) is loaded into the server, it is referred to as a table.

CAS - in Viya, SAS Cloud Analytic Services (CAS) is the star of the show, providing lightning, fast analytics of in-memory data for SAS Visual Analytics and other software offerings.

SPRE - Foundation SAS, is also referred to as the SAS Programming Runtime Environment (SPRE). SPRE provides a user interface and data processing environment for executing classic SAS program code. It offers the Foundation SAS software we're all familiar with, including Base SAS, SAS/ACCESS engines, , and more as well as the SAS Studio web application.

WHAT'S SAS VIYA – IN MEMORY

SAS Viya is the latest enhancement of SAS.
- It is cloud enabled, allowing scalable, web-based access for your data processing needs.
- Its in-memory engine and parallel processing capabilities can have a big impact on execution speed.
- Code that you are accustomed to writing in SAS®9 will execute as is in SAS Viya, but with a few modifications, you can take advantage of in-memory processing for accelerated performance.
- You can also use Python, R, or other open-source languages because Viya provides integration with open-source tools.
- Together, these features in Viya provide a modern analytics platform that can take you from data access and preparation to relevant, actionable results.
-
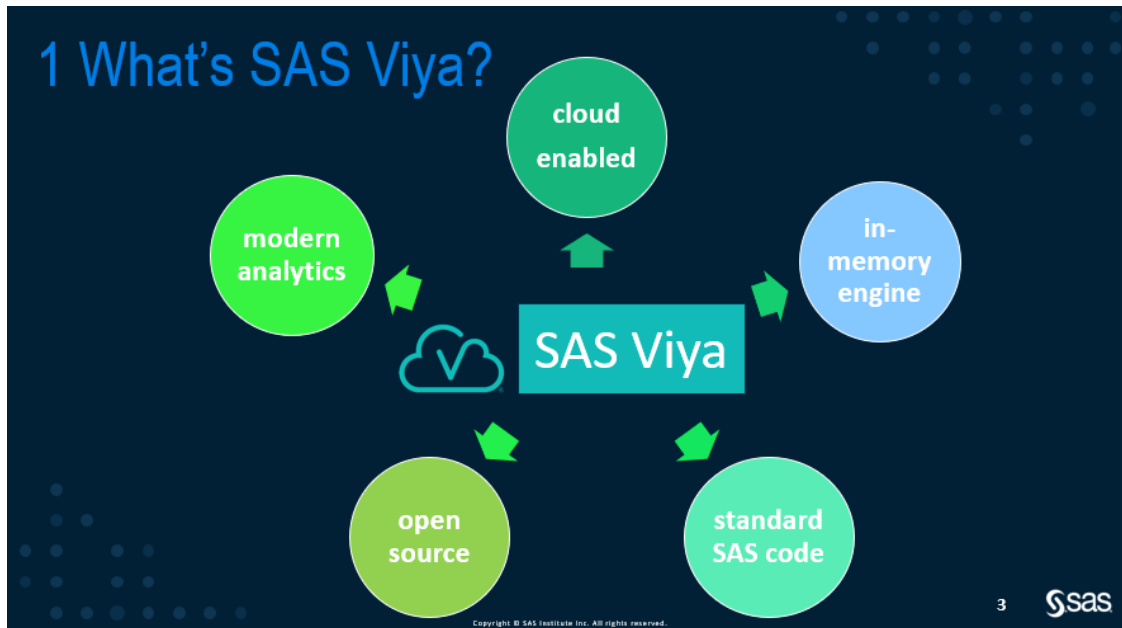
ENGINES

SAS Viya includes multiple servers to execute SAS code. The two primary servers are the SAS Compute Server and SAS Cloud Analytic Services, or CAS. You can think of the SAS Compute Server as equivalent to the SAS®9 workspace server. Your familiar SAS®9 programs can be submitted as is to SAS Viya, and, by default, they'll run on the Compute Server. There's no need to learn new syntax to use the Compute Server.

SAS Cloud Analytic Services is the high-performance server that performs parallel processing on in-memory data. It's likely that you'll use CAS for your big data and complex analytics. Often, only very minor code modifications are required in order for programs to run in CAS. This course focuses on how to write or modify SAS programs to run in CAS.
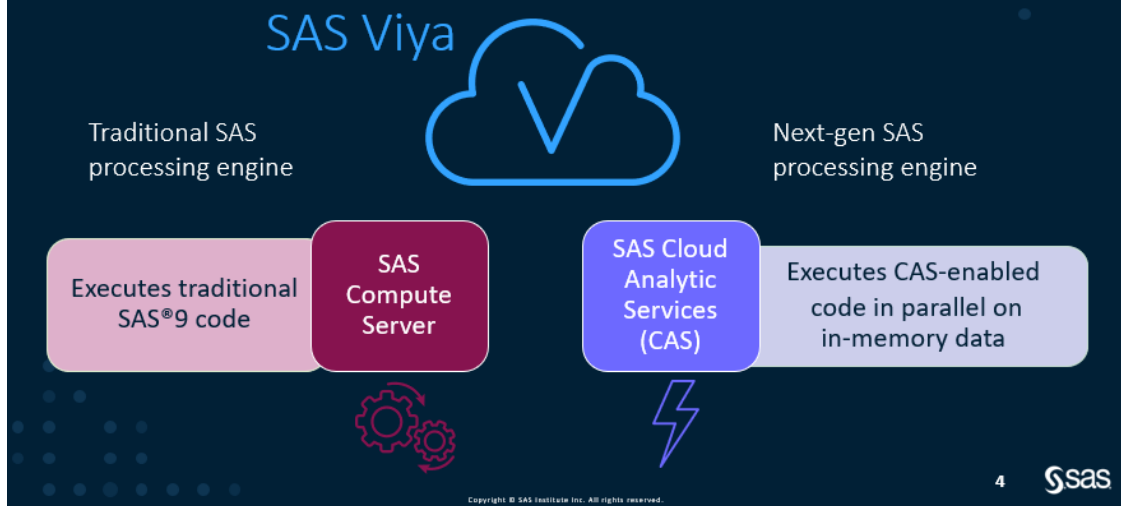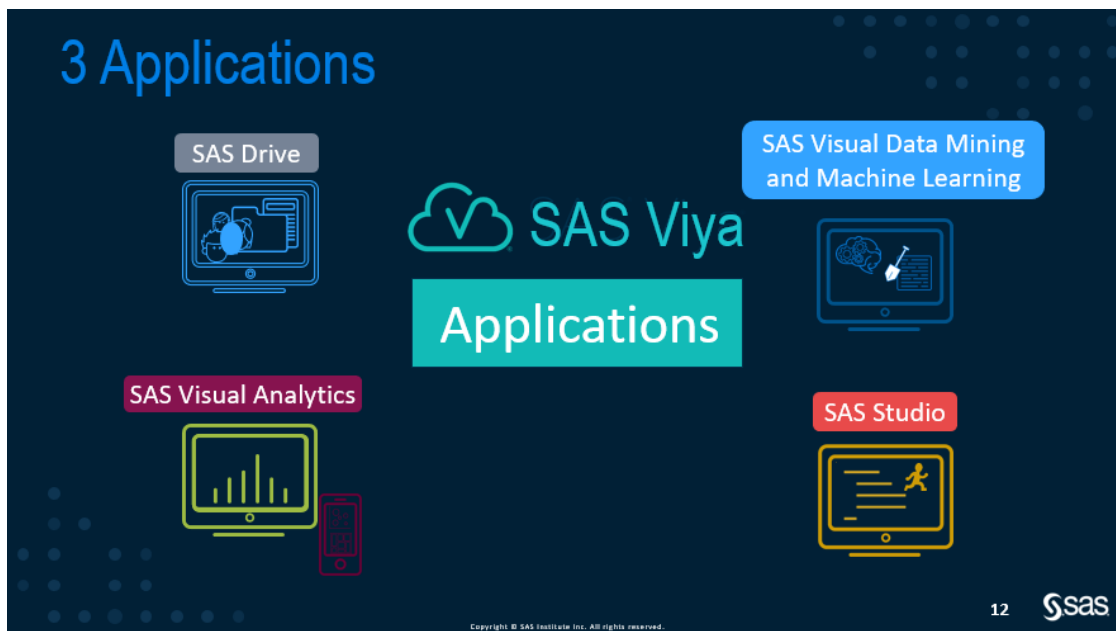
APPLICATIONS

SAS Viya includes several services and applications. Here are a few of these applications:

- SAS Drive, for organizing and accessing all SAS content
- SAS Visual Analytics for web-based reporting and dashboards
- SAS Visual Data Mining and Machine Learning
- SAS Studio for writing and submitting SAS code

DATA

In SAS Viya, you can access a variety of data sources. In addition to SAS tables, you can access data that is on-premises or in the cloud. Data can be in relational databases or unstructured data. Viya can also use other familiar file formats such as XML, JSON, CSV, or XLSX.
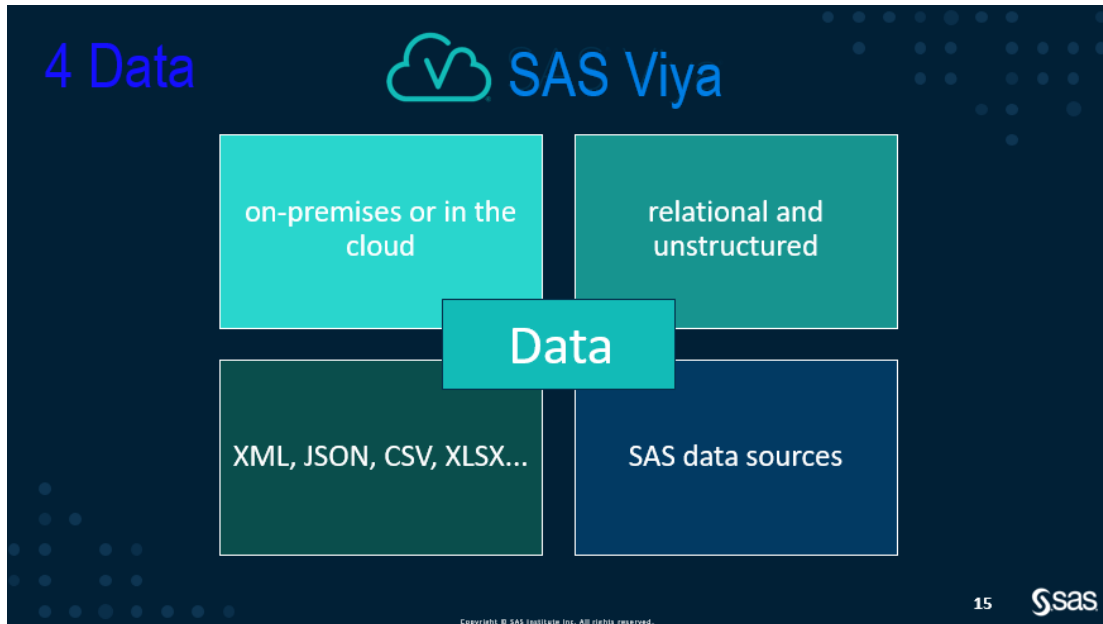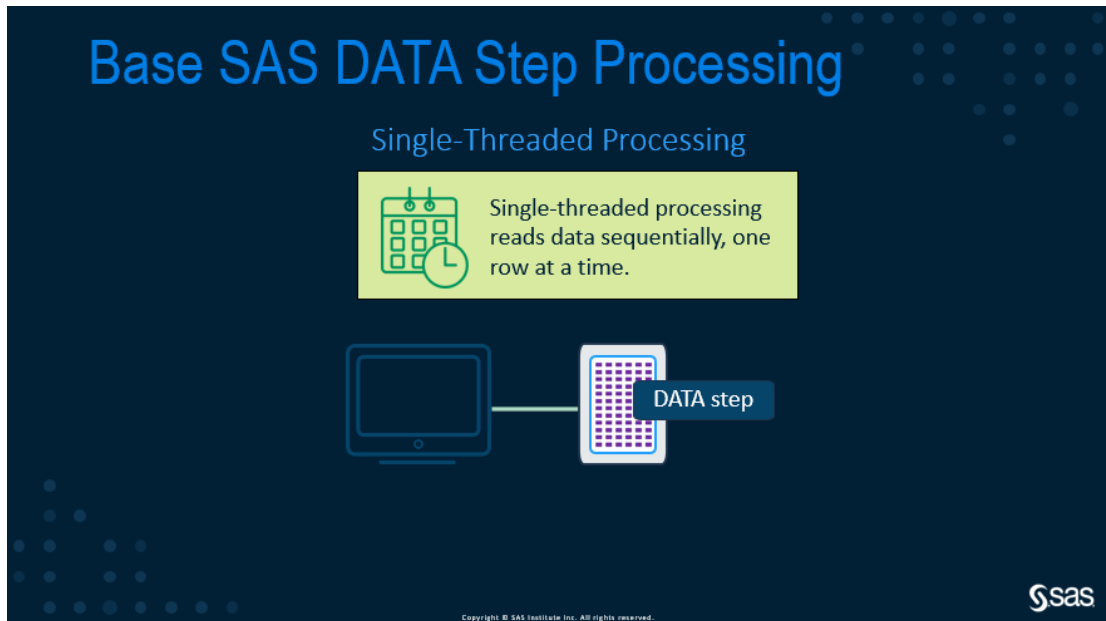

CODE

SAS Viya is not a replacement for SAS 9. You can still leverage your SAS programming knowledge and make tweaks to existing SAS code to allow it to run in SAS Viya.

When a DATA step is executed in Base SAS, it runs in a single thread on the SAS workspace server. Processing data in a single thread reads data sequentially, one row at a time.

SAS Viya enables data to be divided and processed simultaneously on multiple threads. When a data step executes in CAS, each thread executes the program statements on its data, and returns the results to the controller.
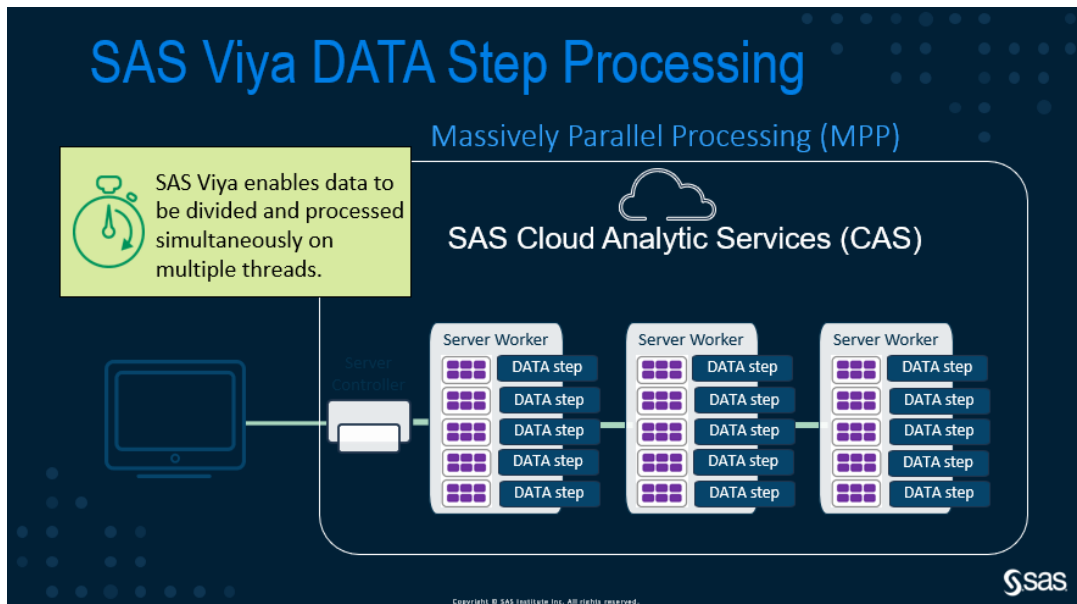
Sometimes to get the DATA step to process in CAS, it's as simple as modifying the library reference on the DATA statement and the SET statement to use a caslib. When both the output and input tables are CAS tables, the DATA step will process in CAS.

Modifying the DATA Step to Run in CAS

```
data casuser.Departments;
   set casuser.mycustomers end=eof;
      select(Continent);
         when ('Africa', 'Asia')
             Department="General Shoes";
         when ("Oceania")
             Department="Men's Shoes";
         when ("North America", "Europe")
             Department="Women's Shoes";
         otherwise Department='Unknown';
      end;
   keep City Continent Department;
   if eof then put _threadid_ =    _N_=;
run;
```

If the output table and input table use a caslib library reference, the DATA step runs in CAS.

pv03d02

COMPUTE SERVER OR CAS

So, the question becomes, Should I use the compute server, or should I use CAS & there's really no black and white answer to that. Because CAS performs in-memory parallel processing, does that mean that you should modify all of your code to run in CAS? Probably not. Each have their unique strengths. It depends on the size of your files and the nature of your code as to which steps leverage those strengths. It might be most efficient for your code to execute with a combination of the two servers.

Compute Server or CAS?

Is CAS always faster?

It depends ...

Compute Server

CAS

So how do you decide which steps you should write or modify to run in CAS?

As you work in your Viya environment and consider writing or modifying programs to run in CAS, there are a few best practices to consider.

- CAS generally outperforms the Compute Server if you have data sources larger than 50 GB,

- steps that run for 30 minutes or longer,

- PROCs that are computationally demanding, or DATA steps with extremely long or complex logic.

CONCLUSION

This paper is an introduction to the 5 new gotchas to be aware of and hopes to serve as a starting point for SAS 9 users who are contemplating the move to SAS Viya and want to increase their awareness.

ACKNOWLEDGEMENTS

The author is grateful to the many SAS users that have entered her life and the Pharmasug committee for creating new opportunities to learn about advanced programming techniques.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:
Charu Shankar
SAS Education
Charu.shankar@sas.com

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.

REFERENCES

CAS Concepts

https://go.documentation.sas.com/?docsetId=calserverscas&docsetTarget=n05000viyaservers000000admin.htm&docsetVersion=3.4&locale=en

An intro to Viya Programming

https://go.documentation.sas.com/api/docsets/pgmdiff/3.4/content/pgmdiff.pdf

Differences in the SAS® 9 and SAS® Viya™ 3.1 Platforms

https://go.documentation.sas.com/api/docsets/whatsdiff/3.1/content/whatsdiff.pdf?locale=en#nameddest=n0evbd1ha0clqvn1sbz5yag06xi6

SPRE (SAS Programing Runtime Environment)

https://communities.sas.com/t5/SAS-Communities-Library/Deploying-the-SPRE-in-SAS-Viya-3-4/ta-p/602891

SAS® Cloud Analytic Services 3.1: Language Reference

https://go.documentation.sas.com/api/docsets/casref/3.1/content/casref.pdf?locale=en#nameddest=p05ccny5glgvwan19mkisxi8z1jk