

Paper SAS4582-2020

Optimize SAS® Viya®: Dividing Load through the Use of Multiple SAS® Cloud Analytic Services Servers

Brandon Kirk, Jerry Read, Jason Shoffner, SAS Institute Inc.

ABSTRACT

Analytic workloads mixed with data preparation and visualization can put SAS® Viya® to the test. Learn how to optimize your investment and orchestrate different workloads by leveraging multiple SAS® Cloud Analytic Services (CAS) servers. In this paper, we explain setting up and administrating additional CAS servers to keep processes optimized. Once the CAS servers are set up, we cover data administration and securing the data for the intended workload. Orchestrating the workload on SAS Viya keeps your general SAS® Visual Analytics report consumers and machine learning data scientists from stepping on each other's toes.

INTRODUCTION

The SAS Viya platform provides different deployment options to optimize processing power and efficiency. These options can include single machine deployment, full deployment with distributed CAS server, or clustered deployment for high availability. To build on this flexibility, you can deploy multiple CAS servers to meet the needs of your SAS Viya environment.

Sometimes having too many options can be daunting and leave you asking, "what if I **don't** initially choose that route?" The paper explores SAS Viya 3.5 deployment options at a high level with references to the topics in detail. A CAS server can be added after the initial deployment, and it can be created in symmetric multi-processing (SMP) or massively parallel processing (MPP) mode. This allows for scaling either by adding workers to a CAS server in MPP mode or by adding additional CAS servers.

The advantages of running multiple CAS servers could be as strategic as separating workload by usage type or as simple as providing a sandbox location to prevent the possibilities of development cycles interfering with production work. This paper focuses on helping you identify when additional CAS servers will meet the objective, and how to implement and administer additional CAS servers within your analytics cluster.

Note: The capacity for each additional CAS server must be within your license agreement.

SCENARIOS FOR OPTIMIZING CLOUD ANALYTIC SERVICES

As defined in the SAS administration guide, CAS is a server that provides the cloud-based, run-time environment for data management and analytics with SAS. Understanding all the options for scaling and optimizing the run-time environment empower design and implementation decisions that align with your environment goals. The paper introduces the

topics of scaling and optimizing, which are focused on addressing a challenge, and then dives deeper into managing multiple CAS servers. The goal is to provide the information needed to pursue the appropriate path to meet the needs of the environment.

INCREASING DATA CAPACITY

There are some simple ways to scale capacity. If the system is a small SMP, then additional memory for the host could be the appropriate answer. Another solution for an SMP environment might be to move to an MPP deployment, which can be accomplished through Ansible and the site playbook. If the deployment is already running in MPP mode, then the Ansible playbook `deploy-casworker` can be used without restarting the CAS server. You can find additional detail on this topic in the topic **"Convert SMP CAS Server to MPP CAS Server"** in *SAS® Viya® 3.5 Administration: SAS® Cloud Analytic Services*.

HIGH AVAILABILITY CONSIDERATIONS

The SAS Viya components can be clustered to increase availability and balance load. There are components that can be clustered post deployment, but to truly architect SAS Viya for high availability, the plan and design need to be created prior to deployment. For a good description of planning and designing for high availability, see Edoardo Riva's SAS Global Forum paper, listed in References (Riva 2019).

CAS RESOURCE TUNING

On Linux deployments, CAS interfaces with the Linux kernel use control groups (cgroups) that apply limits to resource usage (**CPU, memory, disk I/O, network, and so on**) based on policies. Resource management can be implemented broadly at the CAS server level through server configuration files or narrowly at the user or caslib level through policy templates that are applied through the SAS Viya command line interface. You can find additional information about CAS resource management in *SAS® Viya® 3.5 Administration: SAS® Cloud Analytic Services*.

MULTI-TENANCY SAS VIYA

A multi-tenant deployment might be the appropriate choice to meet your deployment objectives, but you should choose it only after careful consideration. Although services are shared in a multi-tenant deployment, data and content cannot be shared across tenants. This deployment option is very effective for a few scenarios such as these:

- complete security and segregation of content and data
- chargeback based on business unit usage requirements
- multiple LDAP connections to directory services

To learn more about multi-tenancy, see Eric Davis' SAS Global Forum paper, listed in References (Davis 2019).

ADDITIONAL CAS SERVERS

You can deploy multiple CAS servers to support your SAS Viya environment. After the initial deployment, you can add CAS servers to new machines, and each additional CAS server can support either SMP or MPP mode. For example, a single-machine CAS server that is deployed to support SMP mode can operate in the same environment as a distributed CAS server, which supports MPP mode. In a scenario where you have a department with small data, they could be located on an SMP CAS server. This prevents small data from being spread thin across a larger MPP footprint.

Here are a few advantages of deploying multiple CAS servers:

Multiple CAS servers support the separation of the CAS server data and processing demands of multiple work groups. This separation of data security and resource management increases control.

- For large enterprise environments that support multiple business units and that want to retain one portal for all reporting, multiple CAS servers allow the ability to right size each CAS server to meet the business unit requirements.
- In order to enable data validation on test data, you can stage an additional CAS server to manage and process data before promotion to a production CAS server. This scenario could be for model training or just report validation.

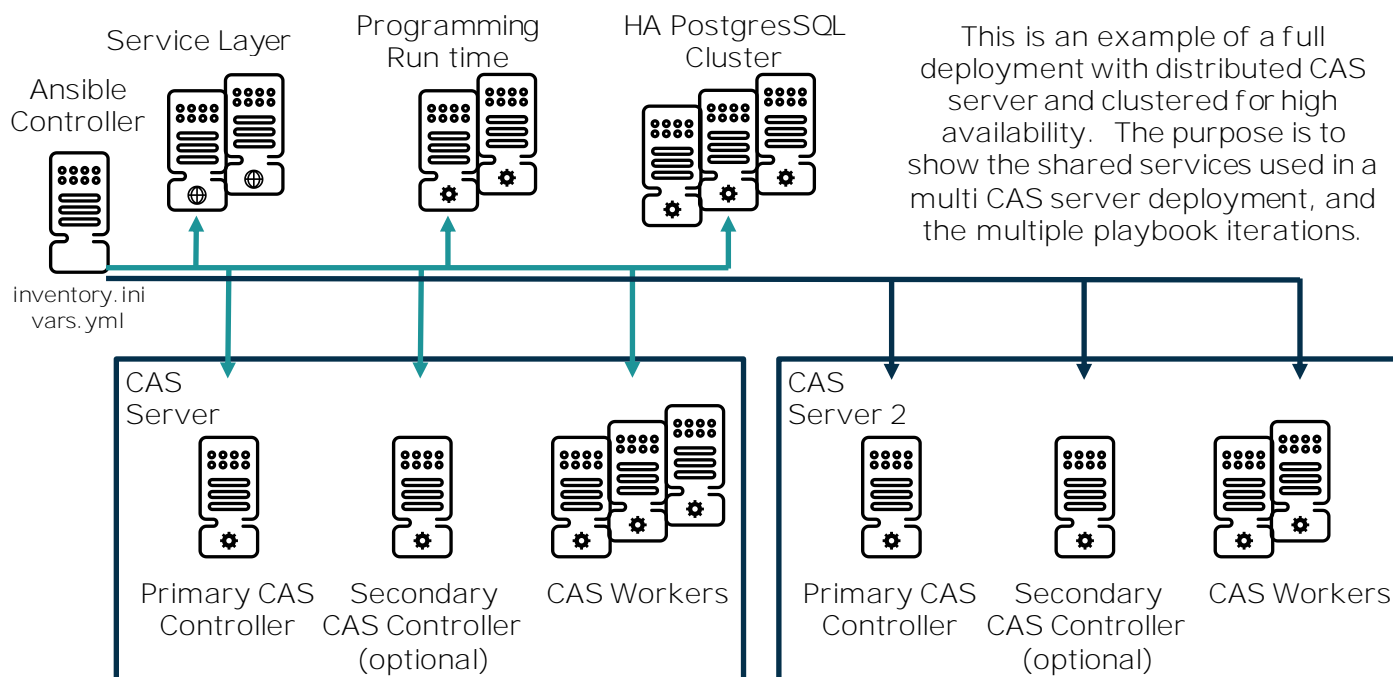


Figure 1: Example Full HA deployment with two MPP CAS servers. The colors depict that for each additional CAS server an additional run through the playbook will be required for deploy, upgrade and update activities.

HOW TO IMPLEMENT ADDITIONAL CAS SERVERS

ADDING A NEW CAS SERVER

The process for adding a new CAS server is very well documented, so there is no reason to go into the details in this paper. However, there is one item that should be appended to the documentation.

In the topic "Add a CAS Server" in *SAS® Viya® 3.5 Administration: SAS® Cloud Analytic Services* you are directed to only add new hosts to the `sas_casserver_primary`, `sas_casserver_secondary`, and `sas_casserver_worker` host groups. While this lists all the additions that are required, we also recommend adding the new hosts to the `CommandLine` host group. This ensures that the command line interface is installed on all new hosts.

STOPPING AND STARTING MULTIPLE CAS SERVERS

The topic "(Optional) Using the SAS Viya Administration Resource Kit" in *SAS® Viya® 3.5 Administration: General Servers and Services: Operate (Linux)* discusses using SAS® Viya®

Multi-Machine Services Utilities (MMSU) of the SAS® Viya® Administration Resource Kit (Viya-ARK) to stop and start services spread across one or more machines. If you have multiple CAS Servers, you will, by definition, have more than one machine, so we also recommend using this valuable tool. Because each CAS server requires its own inventory file, we have found the best way to stop and start services with more than one CAS Server is to create a “master” inventory file that merges the contents of **each CAS server’s** inventory file. See Figure 2 for an example of merging inventory files. Once created, you can reference the merged inventory file when running playbooks from MMSU:

```
ansible-playbook -i merged_inventory.ini /path/to/playbook.yml
```

Warning: Never use the merged inventory file when running the site.yml or update_only.yml inventory file, as only one CAS server can be upgraded at a time.

Default Inventory File	2 nd CAS Server Inventory File	Combined Inventory File
[sas_casserver_primary] cas1_ctrl_primary	[sas_casserver_primary] cas2_ctrl_primary	[sas_casserver_primary] cas1_ctrl_primary cas2_ctrl_primary
[sas_casserver_secondary] cas1_ctrl_secondary	[sas_casserver_secondary] cas2_ctrl_secondary	[sas_casserver_secondary] cas1_ctrl_secondary cas2_ctrl_secondary
[sas_casserver_worker] cas1_worker1 cas1_worker2	[sas_casserver_worker] cas2_worker1 cas2_worker2	[sas_casserver_worker] cas1_worker1 cas1_worker2 cas2_worker1 cas2_worker2

Figure 2: Sample of merged inventory file.

CONVERTING CONTENT TO CAS SERVER 2+

After the new CAS server is in place and you have caslibs and data available for use, you can start the next steps of migrating to this new CAS server. There are three main steps in the process of moving content to use the new CAS server:

1. identify
2. export
3. import (while mapping to new CAS server)

IDENTIFY

There are two main approaches to identify objects. The first one is for environments with not much content. This is done by using SAS® Lineage to explore the lineage of the objects.

This simply shows you the tables that are used in each report. You can start this exploration at the caslib level or at a table level. In Figure 3, we are exploring the lineage of the RAND_RETAILDEMO table.

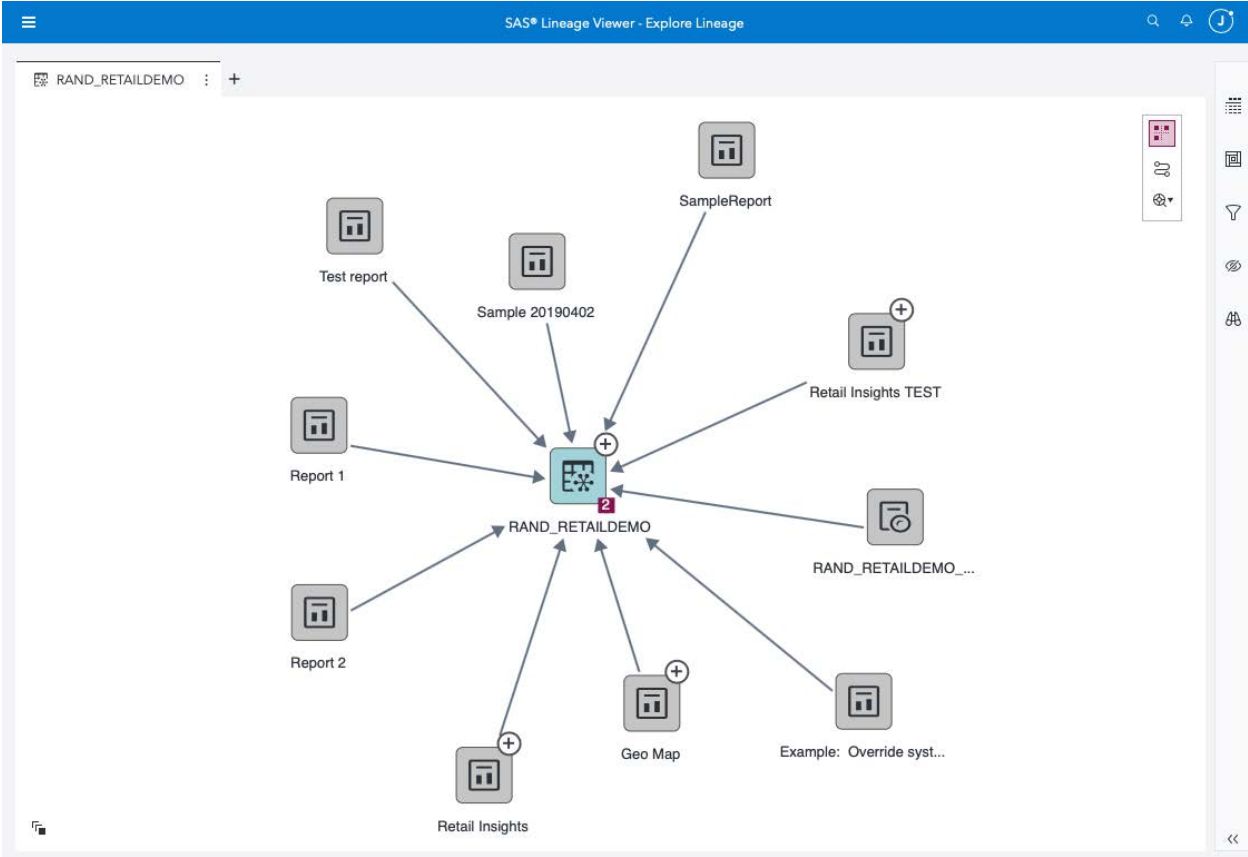


Figure 3: Lineage of the RAND_RETAILDEMO Table

The second approach is by using the relationships API. More information for this method can be found here -> <https://developer.sas.com/apis/rest/CoreServices/#relationships>.

No matter which approach you use, you should be able to determine what reports you need to convert to the use the new CAS server. In our example, RAND_RETAILDEMO is used in seven reports. However, we are only going to focus on the Retail Insights report.

EXPORT

In the Content area of SAS Environment Manager, you can export a json package of the reports that you want to remap to the new CAS server. This self-contained JSON package is a file that contains the report definition and other important attributes about the report.

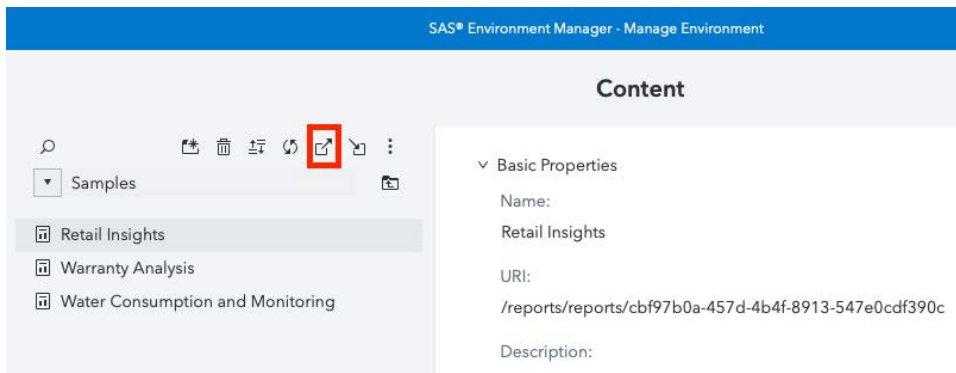


Figure 4: Export JSON Package

The JSON package will be stored locally to the client.

IMPORT

After you export the package, the next step is to import the same package. In the same Content area in SAS Environment Manager, you can import the JSON package.

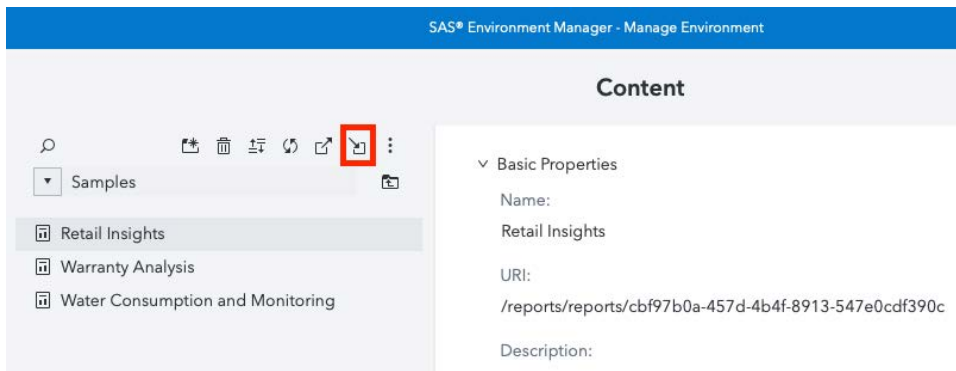


Figure 5: Import JSON Package

During the import process, you have the option to modify the report mappings of the data sources. This is where you will specify the Target Server to be the new CAS server.

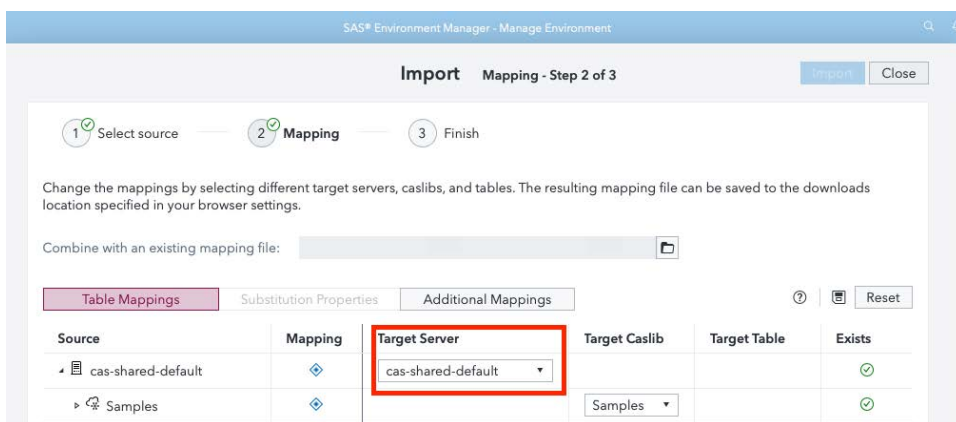


Figure 6: Specifying the New CAS Server as the Target Server

After the package is imported, the reports point to the new CAS server. Verify that the reports open and work as designed.

THINGS TO KNOW

We've discussed some of the benefits of multiple CAS Servers. However, before deciding to implement this on your own environment, here are some things of which you should be aware.

1. Once a new CAS Server is added to an environment, there is no simple process to remove it. The additional CAS Server must remain until an environment is redeployed.
2. If you add an additional CAS Server to an environment, all updates and upgrades will require running the appropriate playbook once for every CAS server (see Figure 1). This could mean extended outage times for updates and upgrades.

CONCLUSION

Adding additional CAS servers can be a very effective way to separate workload while still leveraging SAS Viya services for seamless sharing. Adding CAS servers provides the benefit of achieving total data separation while retaining a single web front end and content security layer. Your report viewers and data scientist will rest easy knowing they can fully leverage all CAS resources without affecting each other.

REFERENCES

SAS Institute Inc. 2019. "Convert SMP CAS Server to MPP CAS Server." In *SAS® Viya® 3.5 Administration: SAS® Cloud Analytic Services*. Cary, NC: SAS Institute Inc. Available <https://documentation.sas.com/?cdcId=calcdc&cdcVersion=3.5&docsetId=calserverscas&docsetTarget=n03003viyaservers000000admin.htm&locale=en#n03039viyaservers000000admin> (Accessed February 17, 2020).

SAS Institute Inc. 2019. "Add a CAS Server." In *SAS® Viya® 3.5 Administration: SAS® Cloud Analytic Services*. Cary, NC: SAS Institute Inc. Available <https://go.documentation.sas.com/?cdcId=calcdc&cdcVersion=3.5&docsetId=calserverscas&docsetTarget=n03003viyaservers000000admin.htm&locale=en#n03049viyaservers000000admin> (Accessed February 17, 2020).

Goddard, Michael. "Thinking about CAS resource management." December 19, 2019. Available <https://communities.sas.com/t5/SAS-Communities-Library/Thinking-about-CAS-resource-management/ta-p/610867>

Pendergrass, Jerry. 2017. "The Architecture of the SAS® Cloud Analytic Services in SAS® Viya™." *Proceedings of the SAS Global Forum 2017 Conference*. Cary, NC: SAS Institute Inc. Available <https://support.sas.com/resources/papers/proceedings17/SAS0309-2017.pdf>

Collum, Rob. 2018. "Five Approaches for High-Performance Data Loading to the SAS® Cloud Analytic Services Server." *Proceedings of the SAS Global Forum 2018 Conference*. Cary, NC: SAS Institute Inc. Available <https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/1854-2018.pdf>

Read, Jerry. 2018. "SAS® Viya®: Architect for High Availability Now and Users Will Thank You Later." *Proceedings of the SAS Global Forum 2018 Conference*. Cary, NC: SAS Institute

Inc. Available <https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/1835-2018.pdf>

Brown, Tony. 2019. "Engineering CAS Performance Hardware Network, and Storage Considerations for CAS Servers." *Proceedings of the SAS Global Forum 2019 Conference*. Cary, NC: SAS Institute Inc. Available <https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2019/3351-2019.pdf>

Davis, Eric. 2019. "Multi-tenancy in SAS® Viya®: Considerations and Implementation." *Proceedings of the SAS Global Forum 2019 Conference*. Cary, NC: SAS Institute Inc. Available <https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2019/3322-2019.pdf>

Riva, Edoardo. 2019. "Proper Planning Prevents Possible Problems: SAS® Viya® High-Availability Considerations." *Proceedings of the SAS Global Forum 2019 Conference*. Cary, NC: SAS Institute Inc. Available <https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2019/3481-2019.pdf>

ACKNOWLEDGMENTS

This content of this paper is a collaboration, with information from all team members who are working to provide SAS Analytic Services for users at SAS.

RECOMMENDED READING

- [SAS® Viya® 3.5 Administration: SAS® Cloud Analytic Services](#)

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Brandon Kirk
Brandon.Kirk@sas.com
<https://www.linkedin.com/in/brandonkirksas/>



Jerry Read
Jerry.Read@sas.com

Jason Shoffner
Jason.Shoffner@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.