

# **SAS® Viya®: Architect for High Availability Now and Users Will Thank You Later**

Jerry Read, SAS Institute Inc.

## **ABSTRACT**

You have SAS® Viya® installed and running for your business. Everyone loves it so much, they are using it more than you initially anticipated. Fortunately, you had the forethought to architect for high availability. Now you don't have to lose sleep as you keep your growing user community happy. This paper guides you through the process of creating high availability for the SAS Viya services, microservices, and web applications that support it.

## **INTRODUCTION**

What is high availability? If you were to ask 10 different Information Technology professionals this question, chances are, you would get 10 different answers. All those answers would probably revolve around things such as continuous operation, uptime, and downtime. My definition, for this paper, is no different – high availability is the introduction of redundant systems or system components, to reduce downtime and enable long term system availability.

High availability of SAS Viya is achieved through the load balancing of system components. While most components can be load balanced, there is a catch. If you plan to use high availability, you must plan for it in your initial deployment. For most components, you cannot change your deployment to add high availability without uninstalling your SAS Viya software and re-installing.

This paper separates the components to be load balanced into four different categories:

1. SAS® Cloud Analytic Services (CAS)
2. SAS Viya Programming Run-Time Servers
3. Microservices
4. Infrastructure servers

This paper will discuss these four categories, identify components that can be highly available, and explain how the high availability is implemented.

## **SAS CLOUD ANALYTIC SERVICES**

SAS Cloud Analytic Services is a server that provides the run-time environment for data management and analytics. The server can be run on a single machine or as a distributed server on multiple machines. The distributed server, also known as a massively parallel processing architecture, is necessary to provide high availability, and consists of a controller, an optional backup controller, and one or more workers.

## **HIGHLY AVAILABLE COMPONENTS**

### **CAS Controller**

In a distributed server architecture, one machine is designated as the controller. That machine coordinates the processing that is performed by the worker nodes. High availability of the controller is achieved using a backup controller. The backup controller is optional, and only applies to a distributed server architecture.

The primary controller continuously synchronizes with the backup controller, and sends heartbeat messages. When an expected heartbeat message is not received, failover to the backup controller occurs.

## CAS Workers

In a distributed server architecture, CAS worker nodes perform data analysis on the node's in-memory data. A distributed server can have one or more worker nodes and, as more resources are needed, additional nodes can be added to distribute workload.

## IMPLEMENTATION

A distributed server architecture can be implemented upon the initial deployment. Machines and their functions are defined in the inventory file, used by Ansible®. There are 3 different host groups in the inventory file representing CAS hosts.

1. [sas-casserver-primary]
2. [sas-casserver-secondary]
3. [sas-casserver-worker]

The primary controller should be listed in the [sas-casserver-primary] host group. An entry in this host group is required for installing CAS.

The backup controller should be listed in the [sas-casserver-secondary] host group. This host group can be empty, but it means there will be no backup controller.

All worker nodes should be listed in the [sas-casserver-worker] host group. If this host group is empty, CAS will be installed with a single server architecture, on the host listed in the [sas-casserver-primary] host group.

Here is an example of these three host groups in an inventory file:

```
[sas-casserver-primary]
cas-controller-1

[sas-casserver-secondary]
cas-controller-2

[sas-casserver-worker]
cas-worker-1
cas-worker-2
cas-worker-3
```

In the example above, cas-controller-1 is the primary controller, cas-controller-2 is the secondary controller, and there are three CAS workers: cas-worker-1, cas-worker-2, and cas-worker-3.

## ADDING ADDITIONAL HOSTS

As data volume grows, and processing times increase, additional CAS workers may be added. This is done by updating the [sas-casserver-worker] host group and running Ansible.

SAS Cloud Analytic Services is the one category of components where high availability can be implemented after the initial deployment, without a complete uninstall and reinstallation. If the [sas-casserver-worker] host group was initially empty, creating a single server architecture, it can be converted to a distributed architecture by adding hosts to the host group and running Ansible. Additionally, if you have a distributed architecture, but no backup controller, one can be added by listing the host in the [sas-casserver-secondary] host group and running Ansible.

## SAS VIYA PROGRAMMING RUN-TIME SERVERS

SAS Viya servers that make up the programming run-time environment are referred to as SAS Viya programming run-time servers:

- SAS Compute Server
- SAS Launcher Server
- SAS Workspace Server
- SAS Object Spawner
- Embedded Web Application Server
- SAS/CONNECT® Server
- SAS/CONNECT Spawner

These SAS Viya Servers are installed on the same host(s), and all can be made highly available. However, multiple instances of these servers are unaware of one another, and failover is not supported. In the event of failure, a session will be established on another host, and the user must log on to reestablish state.

### IMPLEMENTATION

Hosts where these servers are installed must be listed in the [ComputeServer] and [programming] host groups, of the inventory file. To implement high availability, two or more hosts must be listed during the initial deployment. Both host groups must contain the same hosts. Here is an example of these host groups in the inventory file:

```
[ComputeServer]
sas-programming-1
sas-programming-2
```

```
[programming]
sas-programming-1
sas-programming-2
```

In the example above, SAS Viya programming run-time servers will be installed on both the sas-programming-1 and sas-programming-2 hosts.

### MICROSERVICES

Microservices are lightweight, self-contained pieces of software, that perform a single function. They are key components of the SAS Viya architecture. The inventory file contains multiple host groups, representing different categories of microservices:

- [AdvancedAnalytics]
- [CASServices]
- [ComputeServices]
- [CoreServices]
- [DataMining]
- [DataServices]
- [HomeServices]
- [ModelServices]
- [ReportServices]

- [ReportViewerServices]
- [ScoringServices]
- [ThemeServices]
- [configuratn]

High availability is implemented the same way for all microservices – by listing two or more hosts in the corresponding host group:

```
[AdminServices]
microservices-1
microservices-2
```

```
[AdvancedAnalytics]
microservices-1
microservices-2
```

```
[configuratn]
microservices-1
microservices-2
```

The example above only shows three host groups, but the same configuration can be made for any host groups representing microservices.

## INFRASTRUCTURE SERVERS

Infrastructure servers are stateful services that make up the underlying infrastructure required for SAS Viya to operate. These services include:

- Apache HTTP Server
- SAS Configuration Server (Consul)
- SAS Message Broker (RabbitMQ)
- SAS Infrastructure Data Server (PostgreSQL)

### IMPLEMENTATION

While high availability of infrastructure servers is achieved by adding multiple hosts to the appropriate host groups, some of these servers require additional configuration.

#### Apache HTTP Server

Hosts where the Apache HTTP Server will run should be added to the [httpproxy] host group:

```
[httpproxy]
deployTarget-1
deployTarget-2
```

In the example above, the HTTP proxy will be configured on both deployTarget-1 and deployTarget-2. However, to provide high availability of the Apache HTTP Server, a load balancer is also required.

#### SAS Configuration Server (Consul)

Hosts where the SAS Configuration Server will run should be added to the [consul] host group:

```
[consul]
deployTarget-1
deployTarget-2
```

```
deployTarget-3
```

In the example above, SAS Configuration Server will be installed on three hosts. This server is unique because three servers is the minimum required to implement high availability, as the server maintains a quorum voting structure among all hosts.

### SAS Message Broker (RabbitMQ)

Hosts where the SAS Message Broker will run should be added to the [rabbitmq] host group:

```
[rabbitmq]
deployTarget-1
deployTarget-2
```

In the example above, SAS Message Broker will be installed on deployTarget-1 and deployTarget-2.

### SAS Infrastructure Data Server (PostgreSQL)

The SAS Infrastructure Data Server is unique. Out of the box, SAS Viya uses High Availability (HA) PostgreSQL as the SAS Infrastructure Data Server. However, by default, HA PostgreSQL is not deployed in a highly available topology, but rather as a single node on a single machine. The information below shows how to implement a horizontal cluster with two nodes, each on a separate host. While HA PostgreSQL supports numerous other topologies, discussing all of them is outside the scope of this paper. To learn more, please review *Creating High Availability PostgreSQL Clusters*, in the *SAS Viya 3.3 for Linux: Deployment Guide*.

The [sasdatasvrc] host group contains hosts where the SAS Infrastructure Data Server will be installed:

```
[sasdatasvrc]
deployTarget-1
deployTarget-2

[pgpoolc]
deployTarget-1
```

In the example above, HA PostgreSQL nodes will be installed on two hosts, deployTarget-1 and deployTarget-2. The [pgpoolc] host group defines where Pgpool is installed. Pgpool is software that manages the load balancing for HA PostgreSQL.

In addition to updating the inventory file, implementing high availability of the SAS Infrastructure Data Server also requires updating the vars.yml file for your cluster:

```
INVOCATION_VARIABLES:
  deployTarget-1:
    pgpoolc:
      - PCP_PORT: '5431'
        PGPOOL_PORT: '5430'
        SANMOUNT: '{{ SAS_CONFIG_ROOT }}/data/sasdatasvrc'
        SERVICE_NAME: postgres
    sasdatasvrc:
      - NODE_NUMBER: '0'
        NODE_TYPE: P
        PG_PORT: '5432'
        SANMOUNT: '{{ SAS_CONFIG_ROOT }}/data/sasdatasvrc'
        SERVICE_NAME: postgres
  deployTarget-2:
    sasdatasvrc:
      - NODE_NUMBER: '1'
```

```
NODE_TYPE: S
PG_PORT: '5432'
SANMOUNT: '{{ SAS_CONFIG_ROOT }}/data/sasdatasvrc'
SERVICE_NAME: postgres
```

Additional information on the vars.yml file can be found in *Creating High Availability PostgreSQL Clusters*, in the *SAS Viya 3.3 for Linux: Deployment Guide*.

## ADDITIONAL INFORMATION

### DISCLAIMERS

- This paper is based upon an installation containing SAS® Visual Analytics, SAS® Visual Statistics, and SAS® Visual Data Mining and Machine Learning, but the concepts can be applied to other SAS Viya products.
- The information contained in this paper is for SAS Viya 3.3. The information may not apply to future releases.

### EXCEPTIONS

- There are two host groups in the inventory file which do not allow for high availability:
  1. [Operations] – This host group contains services that accumulate metric, log, and notification events from RabbitMQ, then process those into CAS tables which are consumed by the SAS Environment Manager application.
  2. [pgpoolc] – This host group contains Pgpool, used to manage load-balancing and other functions of High Availability PostgreSQL.
- CAS is the only component of SAS Viya that allows for the implementation of high availability after the initial deployment, and without reinstallation.

## CONCLUSION

High availability is very important to enterprise applications, and SAS Viya provides the ability to make almost all components highly available. However, if you determine you might need high availability now or in the future, whether it's to reduce downtime or to improve performance, you should implement it at the initial deployment.

## RECOMMENDED READING

- [SAS® Viya® 3.3 Administration](#)
- [SAS® Viya® 3.3 for Linux: Deployment Guide](#)

## CONTACT INFORMATION

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

Jerry Read  
SAS Institute  
(919) 531-9201  
jerry.read@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.