­­­

SAS Architecture Plan

Viya 4 Platform

Prepared for

Standard Bank

Version 2

Status: Draft for review

Date: 27 September 2022

|  |  |
| --- | --- |
| **Notices** |  |
| **Copyright** | Copyright © 2022, SAS Institute Inc., Cary, NC, USA. All rights reserved. Produced by SAS Institute Australia / New Zealand.  Style Version: 1.0, July 2008. Content Version: 1.0, February 2021. |
|  | No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc. |
| **U.S. Government Restricted  Rights Notice** | Use, duplication, or disclosure of this software and related documentation by the U.S. government is subject to the Agreement with SAS Institute and the restrictions set forth in FAR 52.227-19, Commercial Computer Software-Restricted Rights (June 1987). |
| **Trademarks and Patents** | SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.  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 registered trademarks or trademarks of their respective companies. |
| **Disclaimer** | This document is provided “AS IS”, for informational purposes only. SAS' contractual commitments are contained solely in the agreement for professional services executed between SAS and the Client. |
| **Confidentiality** | This document is the confidential and proprietary property of SAS Institute Inc. This document may contain approaches, techniques and other information proprietary to SAS. This document shall not be disclosed in whole or in part to third parties without the prior written consent of SAS |

Contents

[Document Distribution v](#_Toc84841731)

[Revision History v](#_Toc84841732)

[Documentation Acceptance v](#_Toc84841733)

[Meetings and Workshops vi](#_Toc84841734)

[1 Introduction 1](#_Toc84841735)

[1.1 Purpose 1](#_Toc84841736)

[1.2 Method and Approach 2](#_Toc84841737)

[1.3 Document Scope 2](#_Toc84841738)

[1.3.1 In-scope 3](#_Toc84841739)

[1.3.2 Out-of-Scope 3](#_Toc84841740)

[2 Requirements Summary 4](#_Toc84841741)

[2.1 Environments 4](#_Toc84841742)

[2.2 Availability and Recovery Requirements 4](#_Toc84841743)

[2.3 Security requirements 4](#_Toc84841744)

[2.3.1 User Authentication 4](#_Toc84841745)

[2.3.2 Users, Groups and Identities 5](#_Toc84841746)

[2.3.3 Encryption 5](#_Toc84841747)

[2.4 Batch Processing Requirements 5](#_Toc84841748)

[2.5 User Community and Workloads 6](#_Toc84841749)

[3 Software Information 7](#_Toc84841750)

[3.1 SAS Viya Products 7](#_Toc84841751)

[3.2 Third-Party Software Products 8](#_Toc84841752)

[3.2.1 Viya 4 Deployment Tool Prerequisites 8](#_Toc84841753)

[4 Solution Overview 10](#_Toc84841754)

[4.1 System Context 10](#_Toc84841755)

[4.2 Architectural Decisions 11](#_Toc84841756)

[4.3 Technical Assumptions 13](#_Toc84841757)

[4.4 Architecture Overview 14](#_Toc84841758)

[4.4.1 Kubernetes Cluster Overview 15](#_Toc84841759)

[4.5 Workload Placement Plan 16](#_Toc84841760)

[4.5.1 SAS Viya PreProd 16](#_Toc84841761)

[4.5.2 Workload Placement Summary 17](#_Toc84841762)

[5 Environment and Infrastructure Details 18](#_Toc84841763)

[5.1 AKS Cluster Details 18](#_Toc84841764)

[5.1.1 Nodes and Node Pools 18](#_Toc84841765)

[5.1.2 Storage 21](#_Toc84841766)

[5.2 Namespaces 23](#_Toc84841767)

[5.3 Networking 23](#_Toc84841768)

[5.4 Database Drivers and Client Software 24](#_Toc84841769)

[5.5 Approach to Kubernetes Administration 25](#_Toc84841770)

[5.6 Private Registry Information 25](#_Toc84841771)

[5.6.1 Registry Update Approach 25](#_Toc84841772)

[5.7 Deployment Assets 25](#_Toc84841773)

[6 SAS Viya Configuration Details sasviya-preprod 26](#_Toc84841774)

[6.1 SAS Viya Deployment Approach 26](#_Toc84841775)

[6.2 Namespace Information 26](#_Toc84841776)

[6.3 Directory Structure 26](#_Toc84841777)

[6.4 SAS Viya Configuration 27](#_Toc84841778)

[6.4.1 Ingress Host and SAS Services information 27](#_Toc84841779)

[6.4.2 Security Aspects 27](#_Toc84841780)

[6.4.3 SAS Infrastructure Data Server 28](#_Toc84841781)

[6.4.4 CAS Server mode and configuration 28](#_Toc84841782)

[6.4.5 SASWORK Configuration 29](#_Toc84841783)

[6.4.6 Configure SAS/CONNECT Settings 29](#_Toc84841784)

[6.4.7 SITEDEFAULT.YAML Configuration 29](#_Toc84841785)

[6.4.8 High-Availability Configuration 30](#_Toc84841786)

[6.4.9 Mail Service Configuration 30](#_Toc84841787)

[6.4.10 Workload Placement Implementation 30](#_Toc84841788)

[7 Systems Management 31](#_Toc84841789)

[7.1 Logging and Monitoring 31](#_Toc84841790)

[7.2 Backup and Recovery Recommendations 31](#_Toc84841791)

[8 Installation and Configuration 32](#_Toc84841792)

[8.1 SAS Viya PreProd Summary 33](#_Toc84841793)

[Appendix A. Pre-Installation Checklist 34](#_Toc84841794)

[Appendix B. Definitions of Terms and Glossary 37](#_Toc84841795)

List of Figures

[Figure 1. System Context Diagram 10](#_Toc84841796)

[Figure 2. Architecture overview 14](#_Toc84841797)

[Figure 3. AKS cluster overview 15](#_Toc84841798)

[Figure 4. Workload placement summary 17](#_Toc84841799)

[Figure 5. Networking overview 24](#_Toc84841800)

List of Tables

[Table 1. Stakeholders 2](#_Toc84841801)

[Table 2. User community summary 6](#_Toc84841802)

[Table 3 SAS Software 7](#_Toc84841803)

[Table 4 Required 3rd-Party Software 8](#_Toc84841804)

[Table 5 Viya 4 Deployment tool prerequisite software 8](#_Toc84841805)

[Table 6. Architectural decisions 11](#_Toc84841806)

[Table 7. Assumptions 13](#_Toc84841807)

[Table 8. SASViya PreProd workload placement plan 16](#_Toc84841808)

[Table 9. AKS cluster details 18](#_Toc84841809)

[Table 10. Node pool capacity estimates 18](#_Toc84841810)

[Table 11. Node Pool and node specification 20](#_Toc84841811)

[Table 12. Storage specification 21](#_Toc84841812)

[Table 13. Kubernetes namespaces 23](#_Toc84841813)

[Table 14 Installation summary – SAS Viya PreProd 33](#_Toc84841814)

Document Information

Document Distribution

Each new version of this document will be distributed to the following people.

| Recipient | Role | Email |
| --- | --- | --- |
| Emmanuel Mdhluli | Platform Owner – SAS | Emmanuel.mdhluli@standardbank.co.za |
| Tumelo Moiloanyane | SAS Administrator | Tumelo.moiloanyane@standardbank.co.za |
| Le Roux, Jacques JM | SAS Administrator | Jacques.LeRoux2@standardbank.co.za |
|  |  |  |
| Donovan White | Snr. Cloud Solution Architect | Dowhite@microsoft.com |
| Gerhard Coetzee |  | Gerhard.Coetzee@microsoft.com |
| Andre Mills |  | Anmills@microsoft.com |
| Warren du Toit |  | Wadutoit@microsoft.com |
|  |  |  |
| Koketso Moeng | SAS Technical Account Manager | Koketso.Moeng@sas.com |
| Gavin Hindman | SAS Technical Consultant | Gavin.Hindman@sas.com |
| Kathleen Langman | SAS Account Executive | Kathleen.Langman@sas.com |
| Rajesh Soma | SAS Technical Architect | Rajesh.Soma@sas.com |
| Shane Empey | SAS Technical Architect | [Shane.Empey@sas.com](mailto:Shane.Empey@sas.com) |
| Jeremia Mtimunye | SAS Technical Architect | Jeremia.Mtimunye@sas.com |

Revision History

| Version | By | Date | Summary of change |
| --- | --- | --- | --- |
| 0.1 | Rajesh Soma | 02-09-2021 | Draft – for review |
| 0.2 | Rajesh Soma | 07-09-2021 | Updated Draft |
| 0.3 | Rajesh Soma | 16-09-2021 | Updated Draft |
| 0.4 | Rajesh Soma | 06-10-2021 | Updated Security & Namespace |
| 1.0 | Rajesh Soma | 15-10-2021 | Final Release |
| 2.0 | Shane Empey | 19-09-2022 | Add in Production |
| 2.0 | Jeremia Mtimunye | 27-09-2022 | Draft Release for Production |

Documentation Acceptance

| Recipient | Title / Role | Organisation |
| --- | --- | --- |
| Emmanuel Mdhluli | Platform Owner – SAS | SBSA |

Meetings and Workshops

The following table lists the various meetings and workshops held to support the preparation of this document.

| Venue / Location | Date | Subject |
| --- | --- | --- |
| MS Teams | Weekly | SBSA SAS Viya |
| MS Teams | 12-09-2022 | Production Kick-Off |
| MS Teams | 13-09-2022 | SBSA - Architecture Workshop 1 |
| MS Teams | 21-09-2022 | SBSA - Architecture Workshop 2 |
| Call between Jeremia & Tumelo | 26-09-2022 | SBSA – Architecture: HA, CAS & Compute |

# Introduction

Standard Bank south Africa (SBSA) have engaged SAS Institute (“SAS”) to produce a design for the installation of SAS software (SAS Visual Analytics and SAS Visual Machine Learning) in Microsoft Azure cloud services.

To add context, the following provides a high-level description of the scenario:

* SBSA has an on-premise SAS9.4 M6 grid platform. M7 GRID Platform
* SBSA has upgrade the existing version of the SAS Viya software purchased, from SAS Viya3.5 to SAS Viya4.
* This is the first SAS Viya deployment at SBSA. Second Deployment after Pre-Prod
* The SAS Viya software will be deployed on the MS Azure cloud infrastructure.
* The corporate network is connected to their Azure subscription using an ExpressRoute connection.
* The environment will use Azure Active Directory for SAS Identities.
* TLS encryption is required.
* The customer TLS certificate will be used for all secure connection
* Encryption of data at rest is not required.
* An ARM template will be used to build a dedicated AKS Cluster for Viya in the SBSA Azure subscription.
* A single AKS cluster will be used for the Viya environment.
* The default workload classes will be used.
* A Jump host will be provided for administration.
* See more in the requirements and architectural decisions.

## Purpose

The purpose of this document is to specify the solution architecture for the SBSA Viya 4 platform based on workshops and discussions with SAS, SBSA and Microsoft stakeholders. A solution architecture is a high-level description of how the SAS Viya platform will be deployed to meet SBSA’s requirements.

This document is a deliverable in SAS Professional Services' Intelligence Platform Implementation (IPI) Methodology.

The document contains information that can be used to understand:

* SAS and any SAS distributed third-party to be installed.
* The recommended installation design to support the requirements.
* The recommended architecture and hardware/infrastructure for the installation.
* The various environments to be installed.

This design document will form the basis for preparing the required infrastructure to support the SAS software installation and implementation activities.

Terms and definitions used in this document are explained in Appendix B

## Method and Approach

Prior to the delivery of this document, SAS consultants met with relevant representatives at SBSA and Microsoft and worked through a series of workshops and meetings to finalize the requirements and the specific deployment aspects.

The following table lists the participants and stakeholders in this design exercise.

Table 1. Stakeholders

| Name | Role | Organisation |
| --- | --- | --- |
| Emmanuel Mdhluli | Platform Owner - SAS | SBSA |
| Tumelo Moiloanyane | SAS Administrator | SBSA |
| Kim Chetty | SBSA Security | SBSA |
| Donovan White | Snr. Cloud Solution Architect | Microsoft |
| Gerhard Coetzee  Vladimirs Davidovs |  | Microsoft |
| Andre Mills |  | Microsoft |
| Warren du Toit |  | Microsoft |
| Gavin Hindman | Solution Architect | SAS |
| Rajesh Soma | Technical Architect | SAS |
| Shane Empey | Technical Architect | SAS |
| Jeremia Mtimunye | Technical Architect | SAS |

## Document Scope

This document covers the technical architecture design for the SAS Viya software deployment on Kubernetes, including any infrastructure requirements and dependencies on third party applications and services. It validates suitability for a baseline installation but does not examine whether the resulting SAS environment(s) will have any particular performance characteristics and has not explored any details of the planned workload.

The key aspects relating to the design are summarized below.

### In-scope

The following items are within scope of this document:

* Architecture and installation design for SAS software (listed in Section ) based on the identified requirements and decisions between SBSA, Microsoft and SAS.
* SAS software component details.
* Detailed specification of the SAS configuration.
* Storage details for the SAS Viya environment.
* Network configuration related to the SAS Viya environment.
* Security details for the SAS platform, the base platform configuration.
* Source data systems and databases to be integrated to the platform.
* Design and implementation of any Disaster Recovery (DR) environment or processes for Production Only.

### Out-of-Scope

The following items are out-of-scope for this document and the installation services:

* A detailed requirement gathering process and documentation.
* The business design, process data or model requirements.
* This document does not cover detailed security requirements and design considerations, in particular aspects related to content and process authorizations (definition of permissions).
* This document does not provide any formal sizing recommendation on hardware capacity required to support planned usage. Capacity planning and on-going monitoring is recommended to ensure the system is running optimally given changes to increased users, usage and data volumes over time.
* Hardware, Cloud Provider resources and other manufacturers are referred to for indicative purposes only, and not to be construed as a recommendation or endorsement by SAS.

# Requirements Summary

The following sections capture the requirements that affect the deployment and installation of the SAS software. This is not intended to be the definitive reference or a complete list of the Viya4 Platform requirements. There may be other requirements captured in SBSA documentation that are beyond the scope the SAS services.

## Environments

The following SAS Viya 4 environment will be built:

* SBSA Pre-Prod Environment  
  The SAS Visual Data Science solution plus Risk Modelling Add-on software will be used for the SAS Viya Pre-Prod (environment).
* SBSA Production Environment  
  The SAS Visual Data Science solution plus Risk Modelling Add-on software will be used for the SAS Viya Production (environment).

## Availability and Recovery Requirements

**Pre-Prod**

There is no HA requirement for the SAS Pre-Prod environment.

See decision AD013 for details of the decision relating to the HA requirements.

No disaster recovery (DR) requirements have been specified for the SAS Pre-Prod environment, and no dedicated DR or standby environments will be implemented.

SAS licensing permits a cold recovery environment being maintained, as long as no SAS processing occurs on that environment whilst the environment is active.

SBSA will develop and test backup and recovery procedures for the SAS platform, based on their IT standards and server backup policies.

**Production**

There is an HA requirement for the SAS Production environment.

See decision AD013 for details of the decision relating to the HA requirements.

Disaster recovery (DR) requirements have been specified for the SAS Production environment only. Cold DR to be designed.

SAS licensing permits a cold recovery environment being maintained, as long as no SAS processing occurs on that environment whilst the environment is active.

SBSA will develop and test backup and recovery procedures for the SAS platform, based on their IT standards and server backup policies.

## Security requirements

The following security requirements have been defined.

### User Authentication

The following authentication requirements have been identified:

* All users must authenticate using their SBSA credentials (username & password).
* Open ID connect was used for authentication in Pre-Prod.
* SAML is recommended for authentication in Production.
* Azure Active Directory (AAD) will be used as a service provider for SAML/OpenID Connect.

### Users, Groups, and Identities

The existing processes will be used for user and group management. The SBSA Active Directory is the definitive source for identity information. Active Directory is currently synchronized with Azure AD.

* Identities will be pulled from Azure Active Directory (AAD)
* LDAPS will be used to import SBSA users and groups into the SAS environment.

### Encryption

Encryption in motion is supported on all network communications and it is configured by default.

The following encryption requirements have been identified:

* TLS (SSL) encryption must be used to secure all end-user environment access.
* The SBSA Certificate Authority (CA) will be used to provide TLS certificates for platform access. The Ingress access will use the SBSA certificate.
* SBSA will provide certificates for all internal network sessions. (Session traffic internal to the Viya platform).

## Batch Processing Requirements

The batch processing requirements and processes will be implemented after the SAS environments have been built. The definition of any batch processes it beyond the scope of this document.

## User Community and Workloads

The SAS environment(s) will be supporting the following users.

Table 2. User community summary

|  |  |  |
| --- | --- | --- |
| Environment | User Group | User population |
| SBSA Viya Pre-Prod | SAS Users | 685 |
| SBSA Viya Production | SAS Users | 685 |
|  |  |  |

# Software Information

This section details the SAS software that will be installed and configured.

In addition to the SAS software, third-party software is required to support the running and operation of the SAS environment(s), both on the servers and client workstations. An example of these products are browsers and database clients.

It is SBSA’s responsibility to procure and install these products prior to the commencement of the SAS installation. A list of the required third-party software products is provided below.

## SAS Viya Products

lists the SAS software and versions to be installed. The SAS Viya software is provided in two different cadence versions:

* Stable – ships as monthly version releases and has a support window of 4 months. Being the current Stable version and three previous versions
* Long-Term Support (LTS) – ships as 6 monthly version releases and has a support window of 2 years. Being the current LTS version and three previous versions.
* Pre-Prod was deployed using Stable version.
* Production to be on latest LTS.

Table 3 SAS Software

| Products | Cadence / Version |
| --- | --- |
| SAS Visual Data Science - Solution (contain SAS Products below) | Stable or LTS |
| SAS Visual Analytics (on SAS Viya)  SAS/ACCESS Products (on SAS Viya)  SAS Data Preparation - includes all of SAS Data Quality  SAS Studio Analyst  SAS Information Governance  SAS Job Flow Scheduler (on SAS Viya)  SAS/CONNECT  SAS Econometrics  SAS Optimization  SAS IML  SAS/QC (on SAS Viya)  SAS Visual Forecasting  SAS Visual Statistics (on SAS Viya)  SAS Visual Data Mining and Machine Learning  SAS Model Manager (on SAS Viya)  SAS Visual Text Analytics optional  SAS Text Analytics | Stable / 2021.1.5  (Pre-Production)  Upgraded to 2022.1  LTS 2022.01 (Production) |
| Risk Modelling Add-on |  |

## Third-Party Software Products

This section provides the details of the required third-party software and versions to be installed. This information including version information is correct and up to date at time of writing this report.

Additional 3rd party software may be required such as Adobe Acrobat Reader to view PDF documents or printer driver software to print documents. These are assumed to be installed and provisioned as part of SBSA’s standard operating environment.

The term “SAS provided” below denotes SAS sourcing, downloading and installing the relevant software on the server or client workstation. “Customer Provided” denotes software to be sourced and installed on machines prior to SAS software installation by SBSA personnel.

details the required software to run the SAS Viya software.

Table 4 Required 3rd-Party Software

| Product | Version | SAS Provided or Customer Provided |
| --- | --- | --- |
| Kubernetes | 1.19.x – 1.21.x  1.22 and later not supported | Customer Provided |
| Kustomize | 3.7.0 | Customer Provided |
| Nginx Ingress Controller | 3.9.0 | Customer Provided |
| cert-manager | 1.1.0 or later | Customer Provided |
| kubectl | 1.18.8 | Customer Provided |

### Viya 4 Deployment Tool Prerequisites

The Viya 4 Deployment tool supports running both from ansible installed on a local machine or via a docker container. The Management Jump Host will be used for administration access and will require the following software to run the tool locally.

Table 5 Viya 4 Deployment tool prerequisite software for Pre-Prod

| Product | Version | SAS Provided or Customer Provided |
| --- | --- | --- |
| Terraform | 0.13.4 | Customer Provided |
| Ansible | 2.9 | Customer Provided |
| unzip | - | Customer Provided |
| tar | - | Customer Provided |
| Kubectl | 1.18 | Customer Provided |
| Kustomize | 3.7.0 | Customer Provided |
| python 3 |  | Customer Provided |
| pip3 |  | Customer Provided |
| openshift pip module |  | Customer Provided |
| dnspython python module |  | Customer Provided |
| Helm | 3 | Customer Provided |
| Git client |  | Customer Provided |

To confirm the latest prerequisites for the deployment tool, see [here](https://github.com/sassoftware/viya4-deployment#prerequisites).

Table 5 Viya 4 Deployment tool prerequisite software for Production (as of August 26, 2022) See [here](https://github.com/sassoftware/viya4-deployment/blob/main/docs/user/Dependencies.md)

| Product | Version | SAS Provided or Customer Provided |
| --- | --- | --- |
| Terraform | v1.0.0 | Customer Provided |
| Ansible | 2.10.7 | Customer Provided |
| unzip | - | Customer Provided |
| tar | - | Customer Provided |
| Kubectl | 1.22-1.24 | Customer Provided |
| Kustomize | 3.7.0 | Customer Provided |
| python 3 | 3.x | Customer Provided |
| pip3 | 3.x | Customer Provided |
| openshift pip module |  | Customer Provided |
| dnspython python module |  | Customer Provided |
| Helm | 3 | Customer Provided |
| Git client |  | Customer Provided |

# Solution Overview

The following sections detail the proposed SAS software environment given the IT standards, identified requirements and architectural decisions.

## System Context

The system context model, sometimes referred to as the business context model, is used to define the scope of the SAS implementation, by depicting the major interactions with external systems and user groups. The model initially represents the entire system as a single object and identifies the interfaces between the system and external entities. This is depicted in the for Pre-Prod, and Figure 2 for Production.

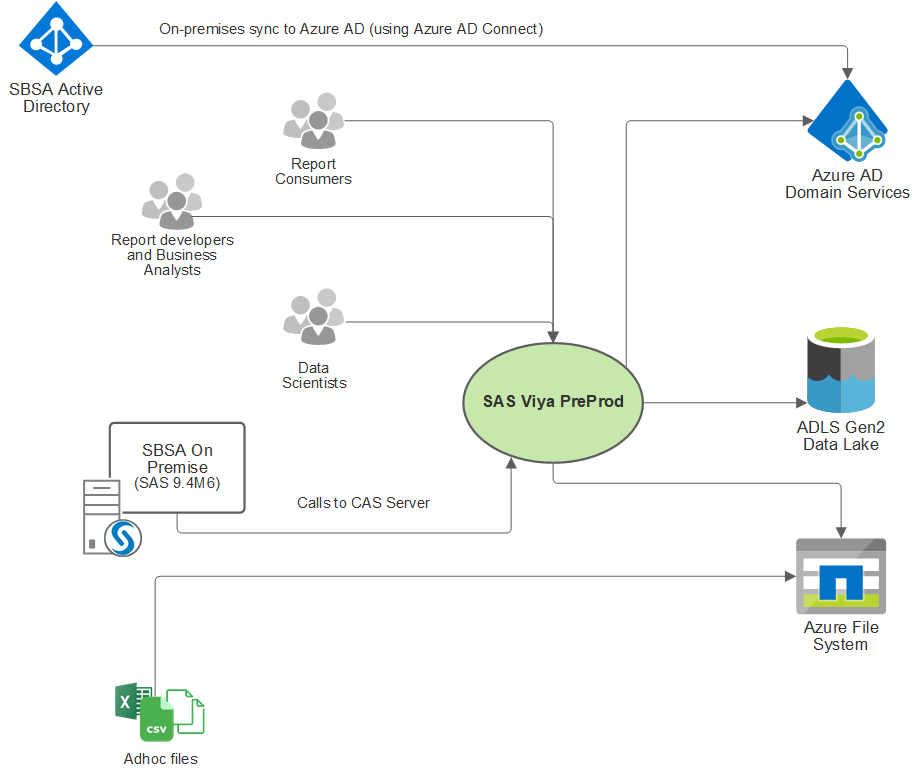


Figure 1. System Context Diagram for Pre-Prod Environment

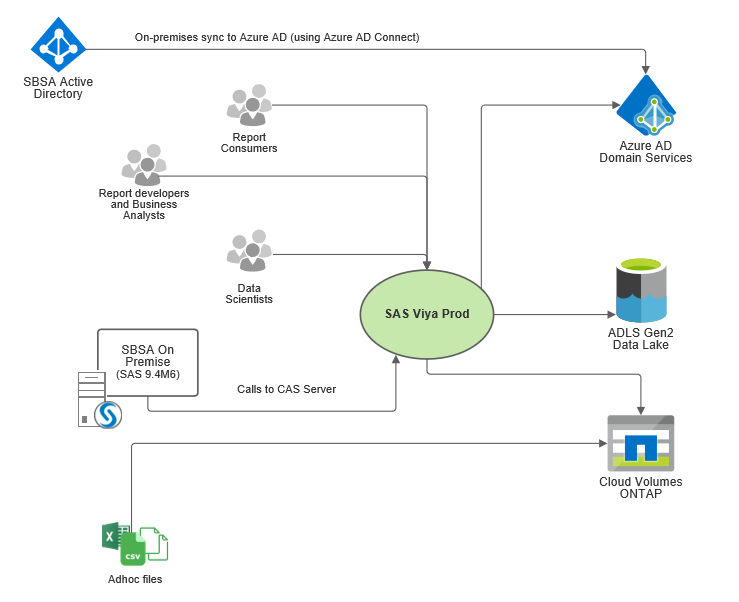


Figure 2: System Context Diagram for Production Environment

## Architectural Decisions

The following table details the key architectural and implementation decisions that have been made.

Table 6. Architectural decisions

| ID | Decision | Decision Details / Rationale |
| --- | --- | --- |
| AD001 | A decision is required on whether the SAS software would use a shared Kubernetes (K8s) cluster or whether a cluster will be dedicated to the SAS environments. | SBSA have made the decision to use a dedicated K8s cluster for the SAS environments.  Pre-Prod and Production will have separate K8s clusters |
| AD002 | A decision is required on how to implement the K8s cluster. Will the Terraform scripts (using the [SAS Viya 4 Infrastructure as Code (IaC) for Microsoft Azure](https://github.com/sassoftware/viya4-iac-azure) GitHub project be used to create the AKS cluster, or will the cluster be manually built? | SBSA have made a decision to manually build the K8s cluster and will make use of ARM template scripts. |
| AD003 | The following namespace will be used for the cluster | sasviya-preprod  sasviya-prod |
| AD004 | The following default names will be used for each of the AKS node pods | **Pre-Prod**  Default name: cas, compute, connect, stateless, stateful  **Production**  Default name: cas, compute, services (stateless and stateful), system (for ingress), default (ops4viya) |
| AD005 | A decision is required to define the node pools availability zone. | Default of single zone will be used.  Multi-Zone support in the works for future releases. |
| AD006 | A decision is required on whether the Internal or External PostgreSQL instance will be used. | The decision was made to use the Internal Instance of PostgreSQL. |
| AD007 | A decision is required on whether a local (mirror) registry will be used for the SAS Viya software. | The decision was made to create a mirror registry using the Azure Container Registry (ACR).  Implication(s)  An implication of using a local registry is that the SAS images need to be kept in sync with the Deployment Assets.  This is required for any future SAS Viya version updates (to install future versions of SAS Viya). The Deployment Assets are linked to a specific Viya version.  Automation will be required to pull the SAS images and push them to the ACR. The use of the SAS Mirror Manager will be scripted and run on a scheduled basis. |
| AD008 | A decision is required on whether the ‘SAS Deployment Operator’ will be used.  The SAS Viya Deployment Operator is a Kubernetes operator that automates some of the manual tasks that are required to deploy and update the software. | As the Deployment Operator does not automate all aspects of the installation and update of the SAS Viya environment (deployment) the decision was made not to use the Deployment Operator at this point in time.  As you still have to manually maintain kustomization.yaml file, the manual method will be used.  Production configured to use deployment operator. Pre-prod will be aligned at a later stage. |
| AD009 | A decision is required on the approach to logging and monitoring. Whether the SAS Viya Monitoring for Kubernetes GitHub project will be used or whether another approach (set of tools) will be used. | The decision was made to use the SAS Viya Monitoring for Kubernetes GitHub project.  <https://github.com/sassoftware/viya4-monitoring-kubernetes> |
| AD010 | Secure platform access. The approach to securing platform access needs to be defined. | To support SBSA IT standards all ingress access will be secured using SBSA’s TLS (X.509) certificate.  SBSA will also provide certificates for internal TLS sessions (session flows between SAS Viya pods). |
| AD011 | A decision is required on the approach for the SAS Identities microservice. Two options are support for the identities provider, using LDAP or using System for Cross-domain Identity Management (SCIM). | SBSA will make use of LDAP (Azure AD) for the Identities provider. SBSA will have to use Azure Active Directory Domain Services (AD DS). SBSA are already using the AD DS.  To support CASHostAccountRequired, SSSD needs to be enabled for the environment. This will happen inside of the pods and does not require any node-level configuration. |
| AD012 | A decision is required for the authentication method | SBSA have decided on OpenID Connect as the preferred authentication method for Pre-Prod.  SAML / SCIM multi-domain configured. Get more info. |
| AD013 | A decision is required on whether the SAS environments will use a HA (High-Availability) configuration. | **Pre-Prod**  The decision was made not to implement an HA configuration for the SAS Viya Pre-Prod environment.  By default, the stateful services have multiple replicas (2 or 3), but the stateless services (webapps and microservices) only have 1 replica.  **Production**  HA must be configured within a single Zone.  Zone redundancy is a future requirement and will be looked at when fully supported.  Due to costs, SBSA decided to keep components to a minimum due to cost implications:   * CAS will be deployed in SMP. No HA. * Only one Node will be used for CAS, and Compute will be reviewed later. |
| AD013x | A decision is required on whether the SAS environments will use a DR (Disaster Recovery) configuration. | DR to be cold  There will be no failover for production,  SBSA indicated that they can tolerate up to 4 hours downtime. |
| AD014 | A decision is required on whether the [SAS Viya 4 Deployment](https://github.com/sassoftware/viya4-deployment) GitHub project will be used. | The SAS Viya 4 Deployment project can be used to prepare the Kubernetes cluster and install the required software, such as NGINX and cert-manager, as well as the storage provisioners.  This tool supports running both from ansible installed on your local machine or via a docker container.  The decision was made to use the deployment tool. The ‘Management Jump Host’ will be used to run any client software for managing the SAS environments.  The bank is using ARM template scripts for this. |
| AD015 | Manual deployment | This cannot be used as here is no NFS server.  The initial deployment will be done using a manual approach for Pre-Prod only. |
| AD016 | Deployment using deployment operator | This will be investigated after the initial deployment; this may not add value to the SBSA environment as the cluster setup is automated to reach out to the internet for container images stored in a container repository.  Production will be deployed using the Deployment Operator. Pre-prod to be updated at a later stage. |
| AD017 | A decision regarding the SAS Viya software cadence is required. | After consulting with other SAS deployment experts, it was recommended to use the LTS release for this deployment.. Since this is a PreProd environment it will be a good learning curve for SBSA staff to also get experience in upgrades.  Production will be deployed using the latest LTS version. Pre-Prod will be upgraded later if required. |
| AD018 | A decision is required on the approach to shared storage for the default storage class. | The NetApp Files storage is currently unavailable in Azure South Africa.  The Azure files premium storage is unavailable in Azure South Africa.  SAS does not recommend Azure Files for Production workloads.  Preliminary decision is to use Cloud Volumes ONTAP as advised by Microsoft Cloud Architect. |
| AD019 | A decision is required on how to implement temporary storage for SASWORK, SASUTIL and CAS Disk Cache.  Will the default (emptyDir) storage be used or will dedicated (fast) storage be provisioned. | The decision was made to configure the platform and SAS environments to use instance ephemeral disk (local temporary SSD) for the SAS temporary storage through the use of a hostPath volume.  SSD Fast storage will be provisioned, and we will redirect the SASWORK, SASUTIL and CAS Disk Cache to these disks. |
|  |  | AKS cluster will be deployed as a private cluster. Using user defined routing pointing to firewall.  Load balancer with user defined routing. |
|  |  |  |

## Technical Assumptions

The design specifications and recommendations made in this document are based on the information and assumptions provided during the preceding consultation and discussions. Changes in requirements or assumptions will necessitate a review to determine impact and/or revision of recommendations (including previous estimates provided).

Table 7. Assumptions

|  |  |
| --- | --- |
| ID | Assumption Description |
| AS01 | The Azure infrastructure provides sufficient availability service levels to meet SBSA availability requirements for the SAS environment. |
| AS02 | Workload: It is assumed that the workload in the Pre-Prod environment will not be high as SBSA want to test this technology.  Infrastructure proven itself. Production workload can be defined. |
| AS03 | sasviya-preprod  sasviya-prod |

## Architecture Overview

**Pre-Production:**

Figure 3 provides an architecture overview of the SAS environment. The architecture overview diagram illustrates the major structuring and components of the solution. Two key decisions that drive the deployment architecture are AD001 and AD004. That is, a dedicated Kubernetes (AKS) cluster will be implemented for the sasviya-preprod platform and the standard labels, taints and tolerations will be used.

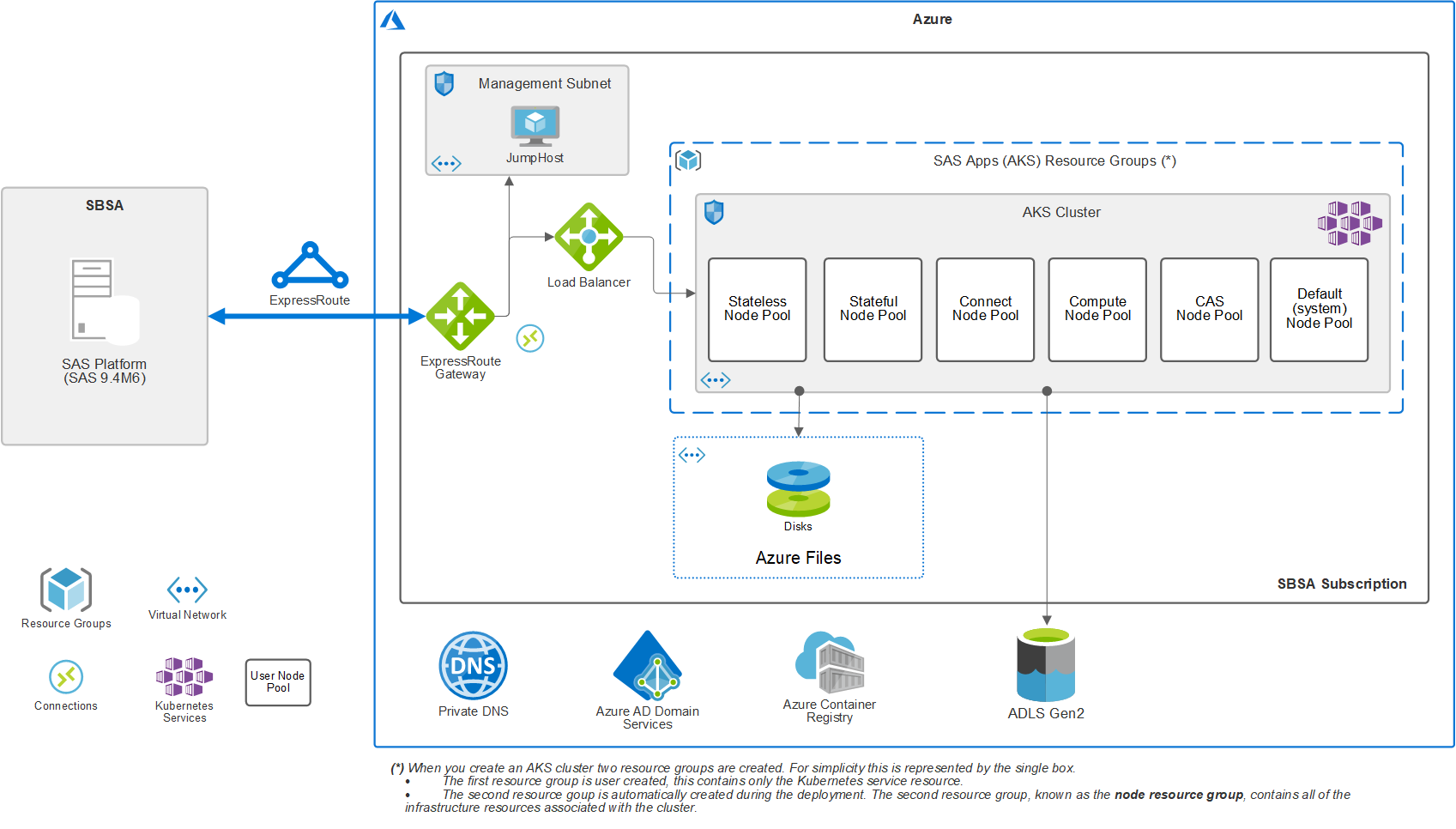


Figure 3: Architecture overview of the SAS environment

**Production:**

Two key decisions that drive the deployment architecture are AD001 and AD004. That is, a dedicated Kubernetes (AKS) cluster will be implemented for the sasviya-prod platform and the labels (services [stateless, stateful], cas, compute, system), standard taints and tolerations (stateful and stateless will share nodes, with cas and comp will be used.

**Graphical user interface, diagram

Description automatically generated**

Figure 4: Architecture Overview (Production)

Running the ARM templates scripts will create the required resource groups and other resources to support the AKS cluster for the SAS environments. Note, the default name will be used for the AKS ‘node resource group’.

The ARM template scripts we customized to create a Load Balancer with private routing tables to prevent the use of Public IP Address. The ExpressRoute CIDR will be used in the ARM template scripts to create the Network Security Group (NSG) rules to secure access to the SAS environment.

The firewall rules for the existing ExpressRoute connections will have to be updated to allow connectivity and user access to the SAS environments.

As per decision AD007, the Azure Container Registry will be used to create a local mirror repository for the SAS software. This can also be handled by the IaC project.

### Kubernetes Cluster Overview

**Pre-Production:**

Figure 5 provides a more detailed overview of the Kubernetes cluster, illustrating the five ‘user’ node pools that were implemented and the default (system) node pool in PreProd.

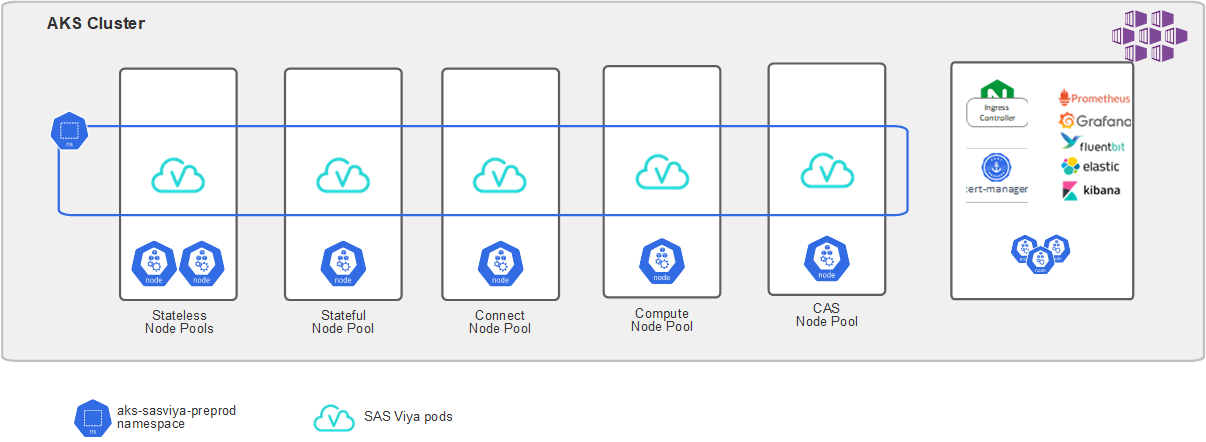


Figure 5. AKS cluster overview for Preprod

The default node pools for the SAS Viya software are:

* Stateless
* Stateful
* Connect
* Compute
* CAS.

Additionally, the default, system, node pool will be used for the non-SAS (third-party) applications. These applications are required to support the operation of the SAS Viya software (such as the ingress controller and cert-manager) and to support the logging and monitoring applications.

**Production:**

Following some changes on the SAS architecture side, we will be dropping the Connect Node Pool, see here for details for Production. The following diagram provides an overview of the Kubernetes Cluster in Production. Stateless and Stateful might be combined to one pool as per decision AD004

Graphical user interface, application

Description automatically generated

Figure 6: AKS Cluster Overview for Production

## Workload Placement Plan

The following workload placement plan was developed using the default node labels and taints.

### SAS Viya PreProd

The SBSA Viya PreProd environment will use the default labels and taints. The following tolerations will be used for the environment:

Table 8. SASViya PreProd workload placement plan

|  |  |
| --- | --- |
| Pod type | Affinity and Tolerations |
| Stateless | As the aks-sasviya-preprod environment will use HA for the stateless pods, the stateless pods will have a toleration for both the stateless and stateful taints to enable the pods to be scheduled across both the stateless and stateful node pools.  The stateless pods will have a preference to be scheduled on nodes that are labeled: workload.sas.com/class=stateless  The default tolerations will be used, the stateless pods will tolerate the following taints:   * workload.sas.com/class=stateless:NoSchedule * workload.sas.com/class=stateful:NoSchedule |
| Stateful | The stateful pods will have a toleration for the stateful and stateless taints to enable the pods to be scheduled across both the stateless and stateful node pools  The stateful pods will have a preference to be scheduled on nodes that are labeled: workload.sas.com/class=stateful  The default tolerations will be used, the stateful pods the tolerate the following taints:   * workload.sas.com/class=stateful:NoSchedule * workload.sas.com/class=stateless:NoSchedule |
| CAS | The cas pods will have a preference to be scheduled on nodes that are labeled: workload.sas.com/class=cas  The default tolerations will be used, the cas pods will tolerate the following taints:   * workload.sas.com/class=cas:NoSchedule   The default CAS resource requests (the CAS auto-resources option) will be enabled for the CAS pods. |
| Compute | The compute pods will have a preference to be scheduled on nodes that are labeled: workload.sas.com/class=compute  The default tolerations will be used, the compute pods will tolerate the following taints:   * workload.sas.com/class=compute:NoSchedule |
| Connect | The connect pods will have a preference to be scheduled on nodes that are labeled: workload.sas.com/class=connect  The default tolerations will be used, the connect pods will tolerate the following taints:   * workload.sas.com/class=connect:NoSchedule |

### SAS Viya Production:

Stateless and Stateful might become one pool(Services) as AD004.

|  |  |
| --- | --- |
| Pod type | Affinity and Tolerations |
| Stateless | As the aks-sasviya-prod environment will use HA for the stateless pods, the stateless pods will have a toleration for both the stateless and stateful taints to enable the pods to be scheduled across both the stateless and stateful node pools.  The stateless pods will have a preference to be scheduled on nodes that are labeled: workload.sas.com/class=stateless  The default tolerations will be used, the stateless pods will tolerate the following taints:   * workload.sas.com/class=stateless:NoSchedule * workload.sas.com/class=stateful:NoSchedule |
| Stateful | The stateful pods will have a toleration for the stateful and stateless taints to enable the pods to be scheduled across both the stateless and stateful node pools  The stateful pods will have a preference to be scheduled on nodes that are labeled: workload.sas.com/class=stateful  The default tolerations will be used, the stateful pods the tolerate the following taints:   * workload.sas.com/class=stateful:NoSchedule * workload.sas.com/class=stateless:NoSchedule |
| CAS | The cas pods will have a preference to be scheduled on nodes that are labeled: workload.sas.com/class=cas  The default tolerations will be used, the cas pods will tolerate the following taints:   * workload.sas.com/class=cas:NoSchedule   The default CAS resource requests (the CAS auto-resources option) will be enabled for the CAS pods. |
| Compute | The compute pods will have a preference to be scheduled on nodes that are labeled: workload.sas.com/class=compute  The default tolerations will be used, the compute pods will tolerate the following taints:   * workload.sas.com/class=compute:NoSchedule |

### Workload Placement Summary

**Pre-Production**

Figure 7 provides a summary of the workload placement plan, showing the result of the node taints and pod tolerations for the SAS environments (as discussed in the previous sections).

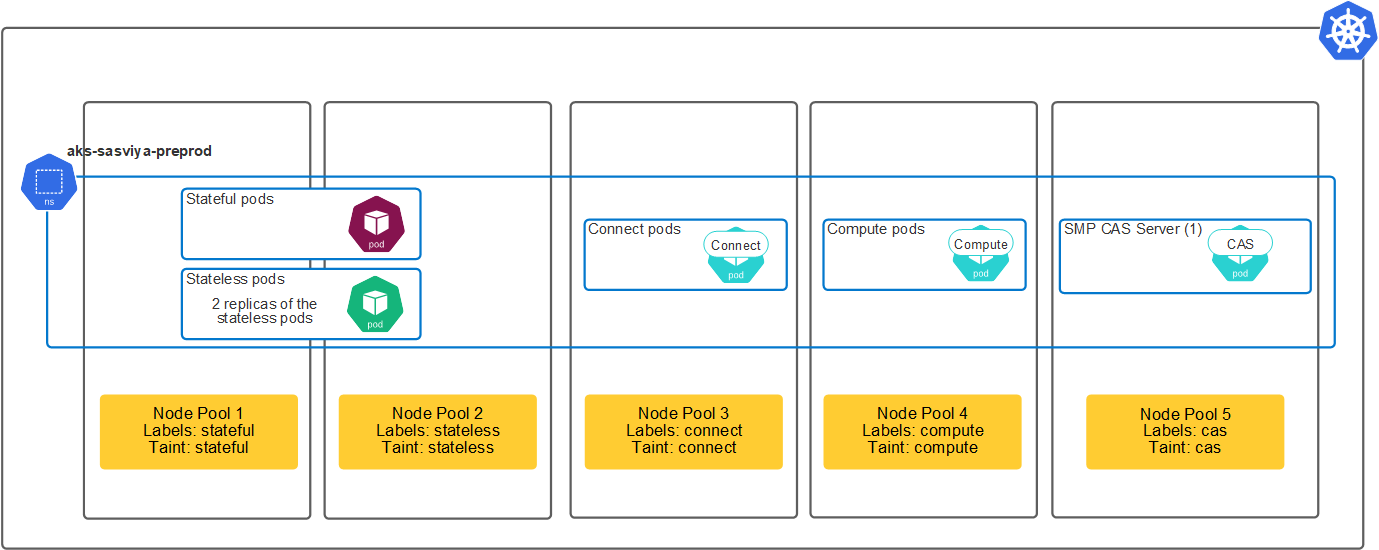


Figure 7. Workload placement summary for PreProd

This configuration means that:

* Only pods that have a toleration for the “stateful” taint will be scheduled onto the stateful node pool (Node Pool 1) nodes.
* Only pods that have a toleration for the “stateless” taint will be scheduled onto the stateless node pool (Node Pool 2) nodes.
* Only pods that have a toleration for the “connect” taint will be scheduled onto the connect node pool (Node Pool 3) nodes.
* Only pods that have a toleration for the “compute” taint will be scheduled onto the compute node pool (Node Pool 4) nodes.
* Only pods that have a toleration for the “cas” taint will be scheduled onto the cas node pool (Node Pool 5) nodes.

The SBSA Viya PreProd environment will use a SMP CAS server.

**Production:**

Figure 8 provides a summary of the workload placement plan, showing the result of the node taints and pod tolerations for the SAS environments (as discussed in the previous sections).

Diagram

Description automatically generated

Figure 8: Workload placement summary for production

This configuration means that:

* Only pods that have a toleration for the “stateful” taint will be scheduled onto the stateful node pool (Node Pool 1) nodes.
* Only pods that have a toleration for the “stateless” taint will be scheduled onto the stateless node pool (Node Pool 2) nodes.
* Only pods that have a toleration for the “compute” taint will be scheduled onto the compute node pool (Node Pool 3) nodes.
* Only pods that have a toleration for the “cas” taint will be scheduled onto the cas node pool (Node Pool 4) nodes.

The SBSA Viya Production environment will use a SMP CAS server. The sizing exercise recommended MPP CAS, however the SBSA team has opted to start with SMP CAS to managed costs and grow the environment over time.

# Environment and Infrastructure Details

This section provides details of the Kubernetes cluster and related infrastructure that will be used to run the SAS Viya environment(s).

## AKS Cluster Details

**Pre-Production:**

As discussed previously, the AKS cluster will be implemented with five ‘user’ node pools for the SAS Viya software and the default (system) node pool will be used for the required third-party applications. One namespace will be implemented for the SAS Viya environment.

The following table details the AKS configuration options.

Table 9. AKS cluster details for pre-prod

|  |  |
| --- | --- |
| Attribute | Details |
| Azure Resource Group name: | SAN-DataServices-PreProd-SAS |
| Kubernetes cluster name: | aks-sasviya-preprod |
| Azure Region: | South Africa (North) |
| Kubernetes version | 1.20.7 |
| Node Pool names and node specification: | See Table 11 (on page ) |
| Container registry: | sasMirrorRegistry |
| Tags: | Name: Vendor Value: SAS |
| Infrastructure resource group | MC\_SAN-DataServices-PreProd-SAS\_aks-sasviya-preprod\_southafricanorth |
| API server address | aks-sasviya-preprod-c765d010.privatelink.southafricanorth.azmk8s.io |

**Production:**

AKS cluster will be implemented with four ‘user’ node pools for the SAS Viya software on Production and the default (system) node pool will be used for the required third-party applications. One namespace will be implemented for the SAS Viya environment.

Figure 9: AKS Cluster details for production

|  |  |
| --- | --- |
| Attribute | Details |
| Azure Resource Group name: | SAN-DataServices-Prod-SAS |
| Kubernetes cluster name: | aks-sasviya-prod |
| Azure Region: | South Africa (North) |
| Kubernetes version | 1.20.7 |
| Node Pool names and node specification: | See 1 (on page ) |
| Container registry: | sasMirrorRegistry |
| Tags: | Name: Vendor Value: SAS |
| Infrastructure resource group | MC\_SAN-DataServices-Prod-SAS\_aks-sasviya-preprod\_southafricanorth |
| API server address | aks-sasviya-prod-c765d010.privatelink.southafricanorth.azmk8s.io |

### Nodes and Node Pools

**Pre-Production:**

The following approach to sizing the node pools will be used.

Table 10. Node pool capacity estimates for pre-prod

|  |  |
| --- | --- |
| Node pools | Capacity estimates |
| Stateless | As the environment will use HA for the stateless pods there will be 3 replicas of all the stateless pods.  A minimum of two nodes will be used. |
| Stateful | Two nodes will be used for the stateful node pool. |
| CAS | One node will be used. |
| Compute | One node will be used. |
| Connect | One node will be used. |
| Default (system) | One node will be used. |

**Production:**

Figure 10: Node pool capacity estimates for production

|  |  |
| --- | --- |
| Node pools | Capacity estimates |
| Stateless | As the environment will use HA for the stateless pods there will be 3 replicas of all the stateless pods.  A minimum of two nodes will be used. |
| Stateful | Two nodes will be used for the stateful node pool. |
| CAS | One node will be used. |
| Compute | One node will be used. |
| Default (system) | One node will be used. |

The following table provides details for the Node Pools and nodes that will be used to support the SAS Viya pre-prod environment.

Figure 11: Node Pool and node specification for pre-production

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Node Pool | Stateless | Stateful | Connect | Compute | CAS | Default |
| Node Pool name | stateless | stateful | connect | compute | cas | default |
| Description | Used for Stateless pods and for the stateful pods. | Used for Stateful pods and for the stateless pods. | Dedicated to connect pod. | Dedicated to compute pod. | Dedicated to CAS pod. | Used for system pods, 3rd Party applications and other required software. |
| Node specification | Standard\_D8s\_v4 (8 vCPU’s, 32GB) | Standard\_D8s\_v4 (8 vCPU’s, 32GB) | Standard\_D8ds\_v4 (8 vCPU’s, 32GB) | Standard\_E8ds\_v4 (8 cores, 64GB) | Standard\_E16ds\_v4 (16 cores, 128GB) | Standard\_D8s\_v4  (8 cores, 32GB) |
| Node Pool size | Static: 2 node | Static: 1 node | Static: 1 node | Static: 1 node | Static: 1 node | Static: 1 node |
| Availability zones | None | None | None | None | None | None |
| Node specific storage | OS 128 GB | OS 128 GB | OS 128 GB | OS 128 GB | OS 128 GB | OS 128 GB |
| Node specific storage | N/A | N/A | 300GB of instance temporary SSD to be used for SASWORK. | 300GB of instance temporary SSD to be used for SASWORK. | 600GB of instance temporary SSD to be used for CAS disk cache. | N/A |
| Shared Storage | N/A | N/A | Azure Files PVC  Breakdown in table 12 below | Azure files PVC  Breakdown in table 12 below | Azure files PVC  Breakdown in table 12 below | N/A |
| Node Labels: workload.sas.com/class= | stateless | stateful | connect | compute | cas | N/A |
| Taints: workload.sas.com/class= | stateless | stateful | connect | compute | cas | N/A |

The following table provides details for the Node Pools and nodes that will be used to support the SAS Viya production environment.

See here and [here](https://learn.microsoft.com/en-us/azure/virtual-machines/ebdsv5-ebsv5-series#ebdsv5-series) for more details.

Figure 12: Node Pool and node specification for production

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Node Pool | Stateless | Stateful | Compute | CAS | Default |
| Node Pool name | stateless | stateful | compute | cas | default |
| Description | Used for Stateless pods and for the stateful pods. | Used for Stateful pods and for the stateless pods. | Dedicated to compute pod. | Dedicated to CAS pod. | Used for system pods, 3rd Party applications and other required software. |
| Node specification | Standard\_E16dsv4(16 vCPU, 128GB) | Standard\_E20dsv4 (20 vCPU, 160GB) | Standard\_E64bds\_v5(48 vCPU, 512GB) | Standard\_E64bds\_v5(48 vCPU, 512GB) | Standard\_D8s\_v4  (8 cores, 32GB) |
| Node Pool size | Static: 2 nodes | Static: 1 node | Static: 1 node | Static: 1 node | Static: 1 node |
| Availability zones | None | None | None | None | None |
| Node specific storage | OS 128 GB | OS 128 GB | OS 128 GB | OS 128 GB | OS 128 GB |
| Node specific storage | N/A | N/A | 300GB of instance temporary SSD to be used for SASWORK. | 900GB of instance temporary SSD to be used for CAS disk cache. | N/A |
| Shared Storage | N/A | N/A | Cloud Volumes ONTAP  Breakdown in table 12 below | Cloud Volumes ONTAP | N/A |
| Node Labels: workload.sas.com/class= | stateless | stateful | compute | cas | N/A |
| Taints: workload.sas.com/class= | stateless | stateful | compute | cas | N/A |

### Storage

The following table provides details of the storage required for the SAS environment.

**Pre-Production**

Table 12. Storage specification for pre-prod

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Storage volume | Type | Share location | Used by | Pod mount point | Access mode |
| User data 512 GB | Shared FS | /azure\_files PVC/sasdata | CAS, Compute & Connect pods | /sasviya/  /sasviya/data | ReadWriteMany |
| User home directories  64 GB | Shared FS | /azure\_files PVC/homes | CAS, Compute & Connect pods | /home | ReadWriteMany |
| Open-source companion software  64 GB | Shared FS | /azure\_files PVC/opensource | CAS, Compute & Connect pods | /bin  /bin/pvcsviyapython  /bin/pvcviya\_r | ReadWriteMany |
| CAS\_DATA\_DIR | Shared FS PVC “cas-default-data” |  | CAS, Compute & Connect pods | /data | ReadWriteMany |
| persistent volumes provisioned using the filesystem | Shared FS |  | Any pods as needed | /pvs | ReadWriteMany |
| CASPERMSTORE | Shared FS |  | CAS pods | /permstore | ReadWriteMany |
| Backup | Shared FS | / azure\_files /backup | CAS and any other pods as needed | /backups  /backups/cas  /backups/common | ReadWriteMany |
| QKB | Shared FS | / azure\_files /qkb | CAS, Compute & Connect pods | /quality-knowledge-base | ReadWriteMany |
| Consul  Default PVC 1GB |  |  | Consul pod(s) |  | ReadWriteOnce |
| RabbitMQ  Default PVC 2GB |  |  | RabbitMQ pod(s) |  | ReadWriteOnce |
| Geode  Default PVC 2GB |  |  | Geode pod(s) |  | ReadWriteOnce |
| Postgres (internal)  Default PVC 128GB |  |  | Postgres pod(s) |  | ReadWriteOnce |

**Production**

Table 13. Storage specification for Production

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Storage volume | Type | Share location | Used by | Pod mount point | Access mode |
| User data 512 GB | Shared FS | /cvo/sasdata | CAS, Compute & Connect pods | /sasviya/  /sasviya/data | ReadWriteMany |
| User home directories  64 GB | Shared FS | /cvo/homes | CAS, Compute & Connect pods | /home | ReadWriteMany |
| Open-source companion software  64 GB | Shared FS | /cvo/opensource | CAS, Compute & Connect pods | /bin  /bin/pvcsviyapython  /bin/pvcviya\_r | ReadWriteMany |
| CAS\_DATA\_DIR | Shared FS PVC “cas-default-data” |  | CAS, Compute & Connect pods | /data | ReadWriteMany |
| persistent volumes provisioned using the filesystem | Shared FS |  | Any pods as needed | /pvs | ReadWriteMany |
| CASPERMSTORE | Shared FS |  | CAS pods | /permstore | ReadWriteMany |
| Backup | Shared FS | /cvo/backup | CAS and any other pods as needed | /backups  /backups/cas  /backups/common | ReadWriteMany |
| QKB | Shared FS | /cvo/qkb | CAS, Compute & Connect pods | /quality-knowledge-base | ReadWriteMany |
| Consul  Default PVC 1GB |  |  | Consul pod(s) |  | ReadWriteOnce |
| RabbitMQ  Default PVC 2GB |  |  | RabbitMQ pod(s) |  | ReadWriteOnce |
| Geode  Default PVC 2GB |  |  | Geode pod(s) |  | ReadWriteOnce |
| Postgres (internal)  Default PVC 128GB |  |  | Postgres pod(s) |  | ReadWriteOnce |

## Namespaces

details the Kubernetes namespaces to be implemented. Along with the namespace for the SAS Viya environment (**sasviya-preprod**) there are several namespaces for the supporting third-party applications.

Table 13. Kubernetes namespaces for pre-prod

|  |  |
| --- | --- |
| Environment or application | Namespace |
| Viya 4 PreProd Environment | sasviya-preprod |
| NGINX Ingress Controller | ingress-nginx (default – set by Deployment Tool) |
| Cert-manager | cert-manager (default – set by Deployment Tool) |
| Monitoring applications | monitoring (default – set by Deployment Tool) |
| Logging applications | logging (default – set by Deployment Tool) |
| EFS Provisioner | efs-provisioner (default name – set by Deployment Tool) |

Table 14 details the Kubernetes namespaces to be implemented. Along with the namespace for the SAS Viya environment (**sasviya-prod**) there are several namespaces for the supporting third-party applications.

Table 14. Kubernetes namespaces for production

|  |  |
| --- | --- |
| Environment or application | Namespace |
| Viya 4 Prod Environment | sasviya-prod |
| NGINX Ingress Controller | ingress-nginx (default – set by Deployment Tool) |
| Cert-manager | cert-manager (default – set by Deployment Tool) |
| Monitoring applications | monitoring (default – set by Deployment Tool) |
| Logging applications | logging (default – set by Deployment Tool) |
| EFS Provisioner | efs-provisioner (default name – set by Deployment Tool) |

## Networking

provides an overview of the network sessions for systems and clients that will access the SAS Viya environments and the monitoring and logging applications. As illustrated, several DNS aliases need to be configured to support the system access.

**Pre-Production:**

The figure shows the connections for the SAS Viya visual interfaces ([**https://sasviya-preprod.standardbank.co.za**](https://sasviya-preprod.standardbank.co.za)) and the connections for the following:

* The SAS Viya CLI (the administration command-line interface)
* Access to the CAS Server (both the RESTful and BINARY APIs)
* SAS/CONNECT sessions, and
* The monitoring (Grafana, Prometheus and AlertManager) and logging (Kibana) services.

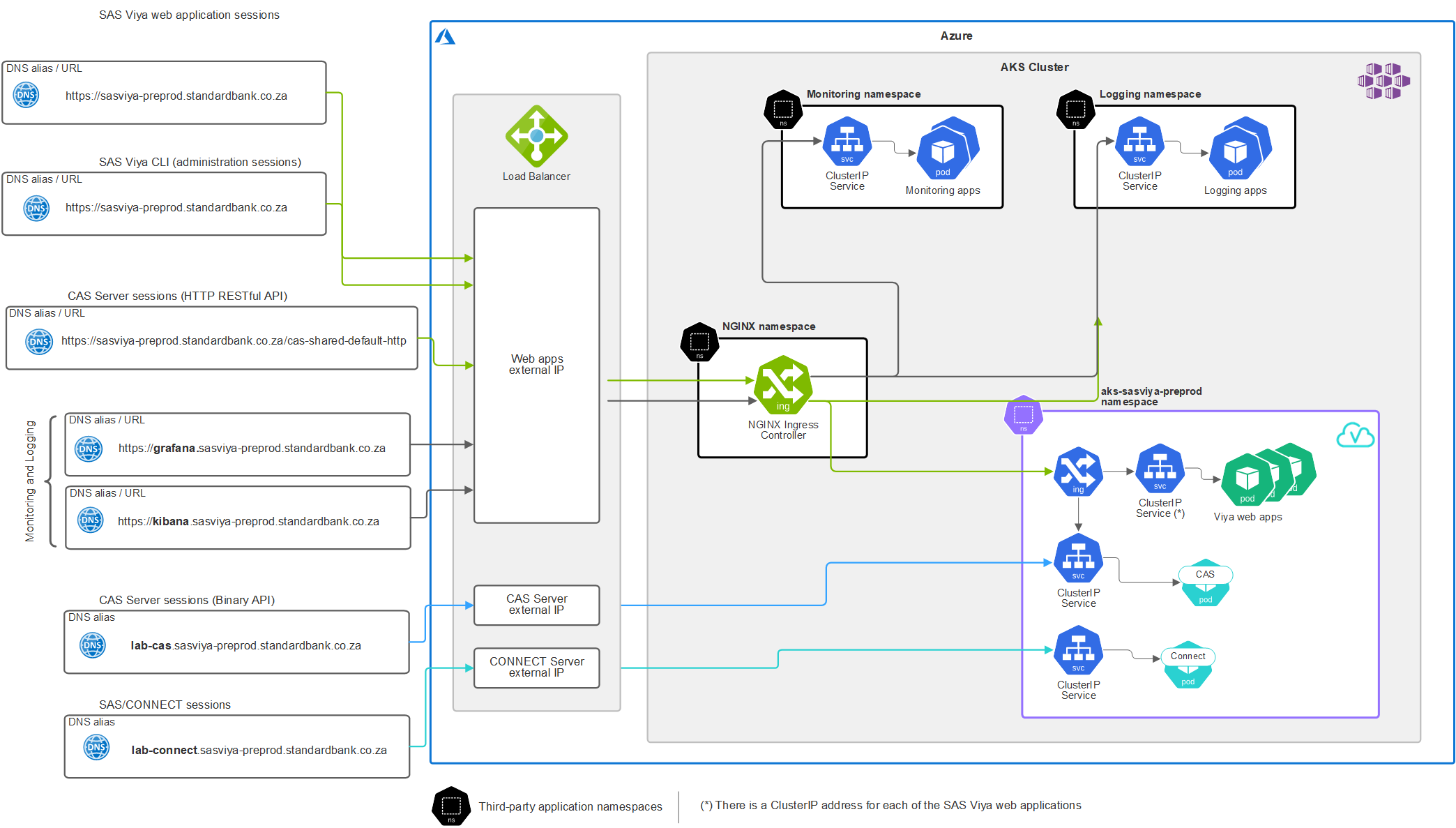


Figure 13. Networking overview for pre-prod

**Production:**

The figure shows the connections for the SAS Viya visual interfaces ([**https://sasviya-prod.standardbank.co.za**](https://sasviya-prod.standardbank.co.za)) and the connections for the following:

* The SAS Viya CLI (the administration command-line interface)
* Access to the CAS Server (both the RESTful and BINARY APIs)
* SAS/CONNECT sessions, and
* The monitoring (Grafana, Prometheus and AlertManager) and logging (Kibana) services.

Diagram

Description automatically generated

Figure 14: Networking overview for Production

Note, the figure only depicts the namespaces that relate to the session flows and/or clients detailed on the left of the figure. For example, the ‘cert-manager’ namespace isn’t shown.

The ARM template scripts we customized to create a Load Balancer with private routing tables to prevent the use of Public IP Address. The ExpressRoute CIDR will be used in the ARM template scripts to create the Network Security Group (NSG) rules to secure access to the SAS environment.

The ExpressRoute network peering will also have to be updated to allow access to the SAS environment using either Azure Public or Microsoft peering.

SBSA are responsible for the configuration and testing of the connectivity and firewall rules, including the Azure Network Security Groups (NSG).

## Database Drivers and Client Software

The database drivers and client software must be installed on a folder in a persistent volume so that they can be mounted to the Compute Server and CAS pods.

Cloud Volumes ONTAP will be used as NetApp is not yet available in Azure South Africa.

It is a SBSA responsibility to provide and install any required database drivers, driver managers and other client software to support the data access.

## Approach to Kubernetes Administration

SBSA & Microsoft will provide K8s application administration

## Private Registry Information

The SAS Mirror Manager can be used to copy the images from the registry SAS hosts to a local Private Docker Registry. SBSA have made the decision to implement a local registry, see decision AD007. The Azure Container Registry (ACR) will be used.

### Registry Update Approach

SBSA will implement an automated batch process to nightly pull the SAS images to ACR. This process will run from the Jump Host in the management VLAN. This will ensure that SBSA have latest, or a daily image update to match any Deployment Assets that will be used in the future.

The definition of the batch process is beyond the scope of this document.

Manual trigger is recommended.

## Deployment Assets

SAS Viya Orders is a command-line interface (CLI) that calls the appropriate SAS Viya Orders API endpoint to obtain the requested deployment assets for a specified SAS Viya software order. This can be used to automate the download of the deployment assets.

[SAS Viya Orders CLI GitHub project](https://github.com/sassoftware/viya4-orders-cli)

# SAS Viya Configuration Details sasviya-preprod

This section details the SAS Viya configuration and additional customizations for the SBSA aks-sasviya-preprod environment. The environment will implement the following SAS software.

|  |  |
| --- | --- |
| SAS Software / Offering | Order number |
| SAS Visual Data Science plus Risk Modelling Add-on | <customer order> |

## SAS Viya Deployment Approach

The default method is to perform the deployment using Kustomize and kubectl commands. The SAS Deployment Operator can also be used.

|  |  |
| --- | --- |
| SAS Viya Deployment Operator |  |
| Will the SAS Viya Deployment Operator be used? | No for Pre-Prod  Yes for Production |
| Will the [SAS Viya 4 Automated Deployment tool](https://github.com/sassoftware/viya4-deployment) be used? | Yes |

## Namespace Information

Each SAS environment is installed into a dedicated namespace. This architecture only depicts the Viya PreProd Environment.

The following namespaces name will be used:

**sasviya-preprod**

**sasviya-prod**

## Directory Structure

The following directory structure will be used for the deployment assets. The Deployment Assets can be manually downloaded from the SAS portal (<https://my.sas.com>).

The deployment assets will be downloaded to the JumpHost and stored under the following directory structure: ‘**/sas/deploy/<environment>/**’

For the pre-prod environment this will be: /sas/deploy/preprod/

For the prod environment this will be: /sas/deploy/prod/

## SAS Viya Configuration

The following sections detail the SAS Viya specific configuration.

### Ingress Host and SAS Services information (Pre-Prod)

The following details the Ingress host and SAS Services URL information.

|  |  |
| --- | --- |
| Kustomization | Value |
| INGRESS\_HOST= | sasviya-preprod.standardbank.co.za |
| SAS\_SERVICES\_URL= | https://sasviya-preprod.standardbank.co.za |

Note, The SAS\_SERVICES\_URL variable is used to provide the value for the SERVICESBASEURL option in compute server sessions. It should point to the host name and port of the cluster's Ingress for this deployment.

### Ingress Host and SAS Services information (Prod)

The following details the Ingress host and SAS Services URL information.

|  |  |
| --- | --- |
| Kustomization | Value |
| INGRESS\_HOST= | sasviya-prod.standardbank.co.za |
| SAS\_SERVICES\_URL= | https://sasviya.standardbank.co.za |

Note, The SAS\_SERVICES\_URL variable is used to provide the value for the SERVICESBASEURL option in compute server sessions. It should point to the host name and port of the cluster's Ingress for this deployment.

### Security Aspects

The following sections provide the details for the authentication and identity provider configuration that will be used for the SAS environment.

#### Authentication: SAS Logon Manager

SBSA will use the following authentication methods for the SAS Logon Manager:

* OpenID Connect with single sign-on enabled.

#### Identity Provider

There is the option to pull identity information from an LDAP provider.

The following information will be used to configure the LDAPS, to configure SAS Viya as an enterprise application.

ldaps://ldapadprd.sbicdirectory.com:3269

ldaps://ldapgcprd.sbicdirectory.com:3269

The SBSA team is testing this option for multiple domains. A final decision will be documented later.

#### TLS configuration

The SAS Security Certificate Framework uses cert-manager, which is an open-source Kubernetes tool, to manage certificate pairs (private key and signed server identity certificate). By default, SAS Viya generates a self-signed Certificate Authority which is used to sign all internal certificate pairs.

A choice must be made for the certificate pair used by the Ingress Controller. The Ingress Controller certificate pair can be signed by the internal cert-manager issuer or provided by the customer. Customers can either provide a PEM format certificate pair or their own configured cert-manager issuer.

The following configuration will be used:

* The SBSA certificate will be used for all external (inbound) access to the SAS environment.
* The SBSA certificate will be used for all internal sessions.
* Cert Manager will be used to connect to Venafi.

#### SSSD Configuration (optional)

The use of SSSD isn’t required for the platform to function. If SSSD is enabled for SAS Cloud Analytic Services, the customer has the option to use the default configuration that is generated or to use a custom configuration.

|  |  |
| --- | --- |
| SSSD configuration |  |
| Will SSSD be enabled? | Yes |
| If yes, will the default SSSD configuration be used or is a custom SSSD configuration required? | Default |

### SAS Infrastructure Data Server

SAS Infrastructure Data Server stores SAS Viya user content, such as reports, authorization rules, selected source definitions, attachments, audit records, and user preferences. SAS Viya uses PostgreSQL for SAS Infrastructure Data Server.

The default configuration will deploy an instance of PostgreSQL inside Kubernetes namespace, which is described as an internal instance. Alternatively, you can provide your own PostgreSQL (from Azure Postgres), which is described as an external instance.

As per decision AD006, SBSA have decided to an Internal instance of PostgreSQL.

### CAS Server mode and configuration

The CAS Server will run in SMP mode. The following table details the CAS server configuration.

Autoscaling – manual overnight

|  |  |
| --- | --- |
| CAS configuration |  |
| CAS mode (SMP or MPP): | SMP |
| The Backup Controller is required? (MPP mode only): | No |
| Number of CAS Workers (MPP mode only): | N/A |
| Adjust CPU Resources | Yes – the default settings will be used |
| Adjust RAM Resources | Yes – the default settings will be used |

#### CAS Disk Cache Configuration

The CAS Disk Cache is to be configured to use the temporary (ephemeral) disk on the Kubernetes nodes, rather than the default emptyDir settings. The **hostPath** configuration will be used for the CAS Disk Cache.

#### Optional Connectivity

Additional configuration is required to support the following connections:

* Connections from SAS 9.4, SAS Viya 3.5, or other binary clients.
* Connections to CAS from SAS Data Connect Accelerators.

The SBSA Viya PreProd environment will NOT have any BINARY client connections and will not use the SAS Data Connect Accelerators.

#### Additional CAS Configuration

No additional CAS configuration is required. Autoscaling to be considered. But manually configured.

### SASWORK Configuration

The SASWORK is to be configured to use the temporary (ephemeral) disk on the Kubernetes nodes, rather than the default emptyDir settings. A **hostPath** configuration will be used for SASWORK.

### Configure SAS/CONNECT Settings

The SBSA Viya PreProd environment will not use SAS/CONNECT.

### SITEDEFAULT.YAML Configuration

SITEDEFAULT.YAML will not be used to pre-load any configuration settings.

The SAS Viya CLI can be used for scripting updates to the SAS Viya configuration. This approach (using the SAS Viya CLI) can be used with CI/CD build processes that SBSA might implement in the future.

The SAS Environment Manager can also be used for interactive updates.

### High-Availability Configuration

The default configuration provides 1 replica of the stateless pods. A Kustomize transformer enables HA for the stateless microservices.

|  |  |
| --- | --- |
| HA configuration |  |
| HA for the SAS Viya stateless microservices is required? | Yes |
| Is there a need to configure the stateless services availability with additional replicas?  The default HA configuration provides 2 replicas. | No |

### Mail Service Configuration

SBSA to provide the details for the Mail Service configuration, during config stage

TBC - Tumelo

### Workload Placement Implementation

See section for details.

# Systems Management

This section details the system management approach and implementation.

## Logging and Monitoring

As per decision AD009, the decision was made to use the SAS Viya Monitoring for Kubernetes GitHub project. <https://github.com/sassoftware/viya4-monitoring-kubernetes>

## Backup and Recovery Recommendations

*A Cold DR environment will be deployed in a different Availability Zone.*

A recovery period of up to 4 hours is allowable by SBSA.

*SBSA will take daily backups of the environment.*

# Installation and Configuration

Installation and configuration of the software will take place following SAS’ standard installation engagement process. A pre-installation workshop or meeting will be held to ensure that all installation prerequisites are finalised, and all parties are aligned as to how each of the components will be installed and configured.

The output of the workshop will be a plan detailing steps and responsibilities of the installation, configuration and testing of SAS software prior to hand-over to .

At the time of installation, SAS will document and provide an installation report. This document provides a summary of the configuration of SAS software, but does not include details on installation steps. SAS strongly recommend that installation only be performed by trained SAS installers, and do not provide detailed screenshot-by-screenshot installation guides.

## SAS Viya PreProd Summary

The following table provides a summary of the key aspects of the installation for the SAS installation specialist.

Table 14 Installation summary – SAS Viya PreProd

|  |  |
| --- | --- |
| Item / Task | Description |
| Customer site number | TBA |
| Cloud or on-premises deployment (If using a Cloud Provider state the provider being used) | Azure AKS |
| SAS software cadence | Stable 2020.1.5 |
| SAS Viya product offering(s) | SAS Visual Analytics |
| Environments to be installed | SAS Viya PreProd |
| Will a shared Kubernetes cluster be used? | No |
| Will the Deployment Operator be used | No |
| Will the [SAS Viya 4 Automated Deployment tool](https://github.com/sassoftware/viya4-deployment) be used? | Yes |
| Is a High Availability (HA) configuration required for the stateless services? | No |
| Will the CAS Server run in SMP or MPP mode? | SMP |
| If using CAS MPP, is a backup Controller required? | N/A for SMP |
| Internal or External instance of PostgreSQL? | Internal |
| Will the customer’s TLS (X.509) certificate be used for the Ingress? | Yes |
| What authentication method will be used? | LDAP |
| What Identity providers will be used? | SCIM |
| Will a custom SSSD configuration be used? | No |
| Is external access to binary ports required? | No |
| Will SAS Viya Monitoring for Kubernetes be used? | Yes |
|  |  |

1. Pre-Installation Checklist

The following checklist details the activities that should be completed prior to the SAS installation specialist starting the install and configuration of the SAS software. It also includes the high-level tasks to configure and deploy the SAS Viya software.

| ID | Task | SBSA | SAS | Status / Notes |
| --- | --- | --- | --- | --- |
|  | Download the Deployment Assets for the selected SAS Viya cadence version from the my.sas.com portal | ü |  |  |
|  | Create service and SAS administrator accounts | ü |  |  |
|  | Create any new user groups and required accounts | ü |  |  |
|  | Provide any required X509 Certificates | ü |  |  |
|  | Download the SAS Viya Administration Resource Kit from GitHub <https://github.com/sassoftware/viya4-ark> | ü |  |  |
|  | **Kubernetes environment setup** |  |  |  |
|  | Azure administration & subscription | ü |  |  |
|  | Provide functioning AKS (K8s) cluster for SAS Viya 2021.1.5  (For Production: SAS Viya 2022.1) | ü |  | Will SAS RMS be used? |
|  | Provide AKS (K8s) cluster administration for SAS Viya 2021.1.5  (For Production: SAS Viya 2022.1) | ü |  | Will SAS RMS be used? |
|  | Required client software and tools have been installed on a jump host or the workstation to be used for the installation activities |  |  |  |
|  | Create the base directory structure that will be used for the deployment assets | ü |  |  |
|  | Create the Kubernetes Cluster for the SAS Viya environment(s) | ü |  |  |
|  | Configure the NSG and firewalls rules required to support communication and access to the SAS environment |  |  |  |
|  | Label and taint nodes to support the Workload Management Plan | ü | ü | SAS to assist as required. |
|  | Install the Kubernetes logging and monitoring tools | ü | ü | SAS to assist as required. |
|  | Create a Mirror Registry | ü | ü | SAS to assist as required. |
|  |  |  |  |  |
|  | Data Sources |  |  |  |
|  | Install required database client software and drivers on persistent volume | ü |  |  |
|  |  |  |  |  |
|  | **SAS Viya Deployment** |  |  |  |
|  | Configure High Availability (optional) |  | ü |  |
|  | Configure TLS |  | ü |  |
|  | Configure PostgreSQL (internal instance) |  | ü |  |
|  | Configure PostgreSQL (external instance) | ü | ü |  |
|  | Configure CAS Server type and settings |  | ü |  |
|  | Configure Kerberos (if being used) | ü | ü |  |
|  | Configure SSSD (optional) – customer to advise if SSSD will be enabled and to provide any custom SSSD settings (optional). See [here](https://go.documentation.sas.com/?cdcId=itopscdc&cdcVersion=v_001LTS&docsetId=dplyml0phy0dkr&docsetTarget=n08u2yg8tdkb4jn18u8zsi6yfv3d.htm&locale=en#n0pshy2wfr9aw0n1g9p5sbbhzyqr) for details. | ü | ü |  |
|  | Persistent storage is required by multiple SAS Viya components. Verify that a storage class has been defined. See the [System Requirements: Persistent Storage Volumes, PersistentVolumeClaims, and Storage Classes](https://go.documentation.sas.com/?cdcId=itopscdc&cdcVersion=v_001LTS&docsetId=itopssr&docsetTarget=n0ampbltwqgkjkn1j3qogztsbbu0.htm&locale=en#n1egh9hqndi6lin13w58nozc7vco)  Also see [Additional PVC Requirement for Microsoft Azure](https://go.documentation.sas.com/?cdcId=itopscdc&cdcVersion=v_001LTS&docsetId=itopssr&docsetTarget=n0ampbltwqgkjkn1j3qogztsbbu0.htm&locale=en#n1k4kmuzevhojkn1j47wdfbeuupk) |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  | **Post Installation Tasks** |  |  |  |
|  | Configure authentication / Identity provider |  | ü |  |
|  | Perform platform validation tests |  | ü |  |
|  | Arrange for a full backup of the persistent storage and the SAS Viya environment configuration, for the purposes of using this as a restore point in the event of a disaster to a known working state | ü | ü | SAS to assist as required. |
|  | Assign customer resources for Knowledge Transfer session(s) | ü |  |  |

1. Definitions of Terms and Glossary

| Term | Description |
| --- | --- |
| AD | Active Directory |
| AD DS | Active Directory Domain Services |
| BI | Business Intelligence (reporting related) |
| BU | Business Unit |
| Data Aggregation | Data Aggregation is a type of data migration where data may be drawn from several different source databases and consolidated on a single target database. |
| Data Constraints | Data Constraints describe the constraints for data items. For example, one data item must be numeric while another must be alphanumeric and match the ISBN format. |
| Data Dictionary | Data Dictionary is a catalogue listing of the entities, their attributes, data types, constraints, and descriptions. Input is provided from an initial analysis of the legacy data sources and the Physical Data Catalogue produced earlier. However, the data dictionary will be aligned to the requirements provided by the Functional team and the business. The data dictionary will also be aligned to the different process areas as identified by the functional team. |
| Data Mapping Activity | Data Mapping Activity is the activity that defines the mapping or translation of records and fields as used by one system to those as used by another. This activity is necessary for successful data migration and data integration. |
| DBMS | DBMS is a Database Management System and usually represents an industry standard database such as Oracle or Sybase SQL Server. |
| Dispatch Window | A dispatch window specifies one or more time periods during which batch jobs are dispatched to run on hosts. Jobs are not dispatched outside of configured windows. Dispatch windows do not affect job submission and running jobs (they are allowed to run until completion). |
| Domain Model | Domain Model is a data model that represents the overall logical structure of a database, independent from any software or data storage structure considerations. It is based on an understanding of the legacy data sources as well as the new data requirements. The Domain Model will provide a high-level view of the new enterprise data requirements and logical organisation of that data. |
| DR | Disaster Recovery |
| EDW | Enterprise Data Warehouse |
| ELT | Extract, Load, Transform |
| Entity Relationship Diagram | Entity Relationship Diagram (ERD) is a graphical method for showing the entities and relationships between them. For example, what is the relationship between a ‘company’ and ‘employee’? A company may have many employees, but an employee can only belong to one company. |
| ETL | Extract, Transform, Load |
| FTP | File Transfer Protocol |
| GB | GigaByte |
| HA | High Availability |
| Logical Data Model | Logical Data Model models the overall logical structure of a database, independent from any software or data storage structure considerations. This may be an existing SAS pre-defined Data model. |
| SBSA | Standard Bank South Africa |
| SCIM | System for Cross-domain Identity Management (SCIM) is a standard for automating the exchange of user identity information between identity domains, or IT systems. |
| SSO | Single Sign-On |
| Static/Reference Data | Static/Reference Data is data that does not change often and is used to reference common data items. For example, currencies are used in different places, but the choice of currencies is always the same. It is often used to populate user choices for the front end. |
| Vocabulary | Vocabulary is used to refer to a collection of data items that have the same meaning but may be called by different names, depending on the system that references the data item. There should be a common agreed vocabulary for the enterprise data model. This will ensure consistency and avoid confusion. |