Architecture and Deployment of SAS® Visual Analytics 7.2 with a Distributed SAS® LASR™ Analytic Server for Showcasing Reports to the Public

Vimal Raj Arockiasamy, and Ratul Saha,

Kavi Global, Barrington, IL

ABSTRACT

This paper demonstrates the deployment of SAS® Visual Analytics 7.2 with a distributed SAS® LASR™ Analytic Server and an internet-facing web tier. The key factor in this deployment is to establish the secure web server connection using a third-party certificate to perform the client and server authentication. The deployment process involves the following steps: 1) Establish the analytics cluster, which consists of SAS® High-Performance Deployment of Hadoop and the deployment of high-performance analytics environment master and data nodes. 2) Get the third-party signed certificate and the key files. 3) Deploy the SAS Visual Analytics server tier and middle tier. 4) Deploy the standalone web tier with HTTP protocol configured using secure sockets. 5) Deploy the SAS® Web Infrastructure Platform. 6) Perform post-installation validation and configuration to handle the certificate between the servers.

INTRODUCTION

SAS Visual Analytics is primarily a web-based application that takes advantage of in-memory analytics and also includes several components that provide a range of functionality from data exploration and reporting to data preparation and administration. This In-Memory approach allows quick response times to the customer even with huge data volumes. An environment for multiple customers has the requirement to scale with the customer demands. MPP (massive parallel processing) configuration is used as scope of this paper to demonstrate distributed deployment model utilizing 5 servers (4 Node Analytics Cluster and 1 Standalone Web Server as shown in Display 1.), each with multiple processors and large memory allocation. The knowledge about distributed deployment of SAS® Visual Analytics 7.2 will help the readers to understand this paper better.

ESTABLISH THE ANALYTICS CLUSTER

An analytics cluster is a high-performance environment that is characterized by massively parallel processing (MPP) used to perform analytic tasks on big data residing in a distributed data storage appliance or in a Hadoop cluster. There are two options to categorize the analytics cluster.

- Analytics Cluster co-located with the data store.
- Analytics Cluster located remotely from the data store.

In this paper we are focused on co-located analytics cluster for the SAS® Visual Analytics 7.2 Deployment. Setting up the analytics cluster requires SAS High-Performance Analytics Environment. The data store is the Hadoop File system (HDFS), which is installed parallel to the LASR Server on every data node. The LASR server has its own Hadoop file format, called SAS HDAT, this format is optimized for fast loading of data into the LASR server.

SAS High-Performance Analytics Infrastructure consists of the following:

- SAS High-Performance Analytics Environment
- SAS High-performance Computing Management Console
- SAS High-Performance Deployment of Hadoop
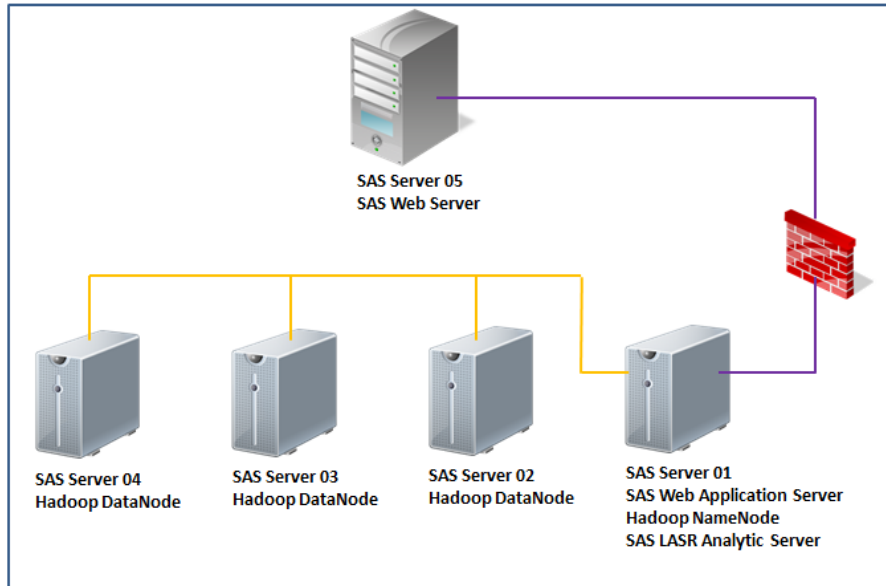
Display 1. SAS® Visual Analytics Distributed Server Architecture

DEPLOY THE SAS HIGH-PERFORMANCE ANALYTICS ENVIRONMENT

Before deploying the SAS High-Performance Analytics Environment the system has to be prepared for the successful deployment. The recommended system setting has to be performed in all the machines in the analytics cluster. Please refer the "SAS High-Performance Analytics Infrastructure: Installation and Configuration Guide" to prepare the system. Then follow along to deploy:

- The software that is needed for the SAS High-Performance Analytics environment is available from within the SAS Software Depot

  Depot/standalone_installs/SAS_High-Performance_Node_Installation/2_94/Linux_for_x64

- Copy "TKGrid_Linux_x85_64.sh" and "TKTGDat.sh" to the /tmp directory of the root node of the cluster.

- Log in to the root node of the cluster with a user that has necessary permissions.

- Change directories to the installation location, (i.e. /opt)

- Run the TKGrid shell script in this directory and respond to the prompts from the shell script.

- Run the TKTGDat.sh in the same directory and respond to the prompts from the shell script.

DEPLOY SAS HIGH-PERFORMANCE COMPUTING MANAGEMENT CONSOLE

Before deploying the SAS High-Performance Management Console the system has to be prepared for the successful deployment. The SAS High-Performance Computing Management Console is deployed on the machine where the SAS High-Performance Analytics environment is deployed. In our scenario it was installed on "SAS Server 01". Below are the installation steps:

- Log in to the target machine as root

- Locate the below path to copy "sashpcmc-2.6.tar.gz" file to the location where the management console need to be installed.

  Depot/standalone_installs/SAS_High-Performance_Computing_Management_Console/2_6/Linux_for_x64

- Enter the following command to extracts the contents into a directory called sashpcmc.

  rpm -ivh sashpcmc*.rpm

DEPLOY CO-LOCATED HADOOP

SAS High-Performance Deployment for Hadoop provides the option of data store to be co-located with the SAS High-Performance Analytic environment on the analytic cluster. The Hadoop that is used here for the deployment is an Apache Hadoop distribution that could be easily configured for use with the SAS High-Performance Analytics environment. Alternatively it could also use the pre-existing supported Hadoop environment. Below are the deployment steps:

- The software that is needed for SAS High-Performance Deployment for Hadoop is available from within the SAS Software Depot similar to the following location.

  /standalone_installs/SAS_High-Performance_Deployment_for_Hadoop/2_8/Linux_for_x64

- Prepare the system for deploying co-located Hadoop.

- Log on to the Hadoop NameNode machine as root.

- Create a directory for installation of Hadoop if it doesn't exist.

  mkdir Hadoop

- Copy the sashadoop.tar.gz file to a temporary location and extract it.

  cp sashadoop.tar.gz /tmp

  cd /tmp

  tar xzf sashadoop.tar.gz

- Above step will create a directory named **sashadoop.**

- Change directory to the **sashdoop** directory and run Hadoop installation command.

  cd sashadoop

  ./hadoopInstall

- Respond to the prompts from the configuration program.

After installing Hadoop on the NameNode it finishes the process by installing High-Performance Deployment of Hadoop on the other machines in the cluster.

GET THE THIRD-PARTY SIGNED CERTIFICATE AND THE KEY FILES

Transport Layer Security (TLS) and Secure Sockets Layer (SSL), are cryptographic protocols that are designed to provide communication security over the Internet. These certificates have a key pair namely a public and a private key. These keys work together to establish an encrypted connection. The certificate also contains what is called the "subject," which is the identity of the certificate/website owner. In our environment we have used SSL certificate. We use these protocols to ensure the network data privacy, data integrity, and authentication.

To get a certificate, one must create a Certificate Signing Request (CSR) for the Web Server. This process creates a private key and public key on Web Server. The CSR data file that you send to the SSL Certificate issuer (called a Certificate Authority or CA) contains the public key. The CA uses the CSR data file to create a data structure to match your private key without compromising the key itself. The CA never sees the private key.

The SAS entities that use the certificates are SAS Deployment Agent, SAS Environment Manager Server and Agent, SAS Information Retrieval Studio, SAS Web Server and SAS Web Application Server. Before

deploying the server tier, middle tier and web tier the site-signed or third-party-signed certificate and key files should be obtained to proceed further.

DEPLOY THE SAS VISUAL ANALYTICS SERVER TIER AND MIDDLE TIER

Install and configure SAS LASR Analytic Server, SAS Visual Analytics server, and SAS Visual Analytics middle-tier software using the SAS Deploying Wizard.

Deploying the SAS Visual Analytics server tier consists of installing and configuring the following components on the server-tier machine:
SAS LASR Analytic Server Monitor
SAS Workspace Servers
SAS Pooled Workspace Servers
SAS Stored Process Servers

Deploying the SAS Visual Analytics middle tier consists of installing and configuring the following components on the middle-tier machine:
SAS Visual Analytics Home Page
SAS Visual Analytics Explorer
SAS Visual Statistics
SAS Visual Analytics Designer
SAS Visual Data Builder
SAS Visual Analytics Viewer
SAS Visual Analytics Administrator
SAS LASR Authorization Service
Visual Analytics Transport Service
Search Interface to SAS Content
SAS Remote Services

Deployment plan used for our scenario is custom plan because of the distributed environment with the separate web server. The installation using the SAS Deployment Wizard is a multilevel deployment as follows
1. SAS Server and Middle Tier (Initial) on SAS Server 01
2. Metadata Server Node (Optional)
3. SAS Web Server on SAS Server 05
4. SAS Server and Middle Tier (SAS Web Infrastructure Platform 9.4) on SAS Server 01
5. Middle Tier Node (Optional)
6. Clients

SAS Server and Middle Tier is a two stage deployment with initial installation and configuration on SAS Server 01" for the above listed components of server-tier machine and middle-tier machine. The second deployment is performed after the deployment of the SAS Web Server, to configure the SAS Web Infrastructure Platform 9.4 on the same server (SAS Server 01) as we deployed for step1. Deployment of the SAS Web Server is covered in the next topic.

DEPLOY THE STANDALONE WEB TIER WITH HTTP PROTOCOL CONFIGURED USING SECURE SOCKETS

After deploying the SAS Visual Analytics Server Tier and Middle Tier is the deployment of the standalone web tier. In our environment the web tier is a standalone server. In the SAS Web Server deployment, the SAS Deployment Wizard does not include an option to configure SAS Web Application Server for HTTPS. The communication path between SAS Web Server and SAS Web Application Server uses HTTP.

SAS Web Server is installed and could be automatically configured with SAS Deployment Wizard. "**Configure SAS Web Server automatically"** check box is enabled on the "SAS Web Server: Automated or Manual Configuration Option" page of SAS Deployment Wizard to configure the SAS Web Server automatically. It is strongly recommended to choose to automatically configure the SAS Web Server
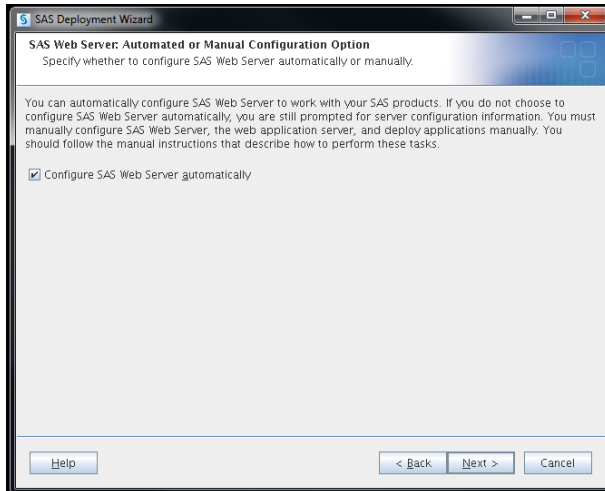
**Display 2. SAS Web Server: Automated or Manual Configuration Option**

Before starting SDW to install and configure the SAS Web Server SSL know the location of the server certificate and its private key. Also make sure the Common Name (CN) field of the server certificate matches the machine name where the SAS Web Server is deployed. During the SSL handshake, the CN of the server certificate is checked against the server machine name for certificate validation.

The page following the "SAS Web Server: Automated or Manual Configuration Option" is the "SAS Web Server: Configuration" page where we have to select "HTTP Protocol using Secure Sockets" and click "Next" from "Configured Protocol" drop down.
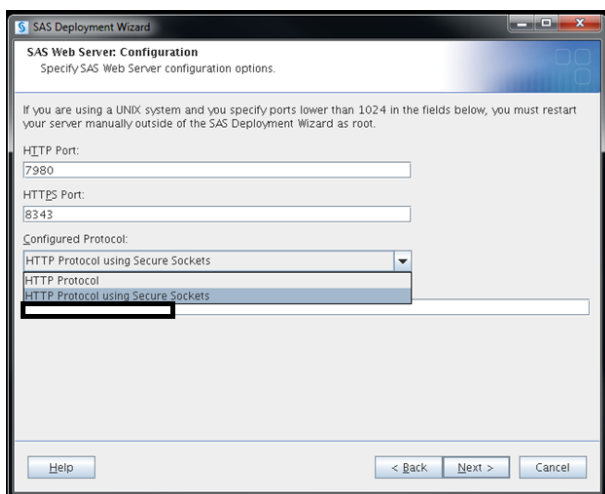
Display 3. SAS Web Server Configuration

On Clicking "Next" from the previous window shown is Display 3, brings up the "SAS Web Server: Location of X.509 Certificate and RSA Private Key" window. Specify the location of the third-party signed certificate file and server key file in this window as shown in Display 4.
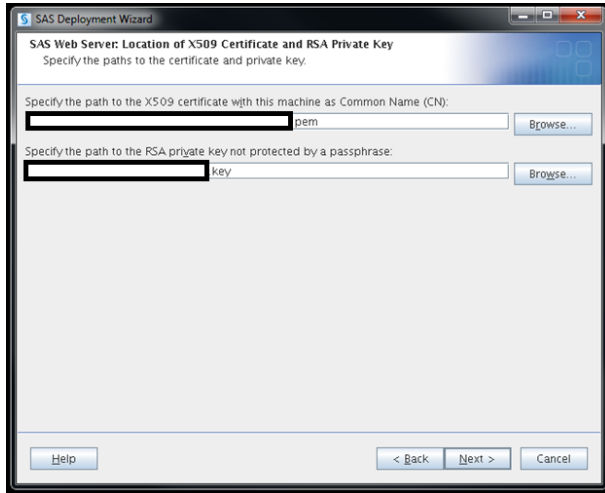
**Display 4. SAS Web Server: Location of X.509 Certificate and RSA Private Key**

The specified certificate and key file in the above display is copied to SAS Web Server configuration location as "/opt/sas/config/Lev1/Web/WebServer/ssl".

DEPLOY THE SAS® WEB INFRASTRUCTURE PLATFORM

Install SAS Web Infrastructure platform using the SAS Deployment Manager on SAS Server 01. Start the SAS Deployment Wizard on SAS Server 01 with following command and follow the screens.

./setup.sh -record -deploy -responsefile /opt/sas/resources/config2.txt

On the "Select Deployment Step" screen select the "Step 4: SAS Server and Middle Tier (SAS Web Infrastructure Platform 9.4)" to configure SAS Web Infrastructure Platform 9.4.
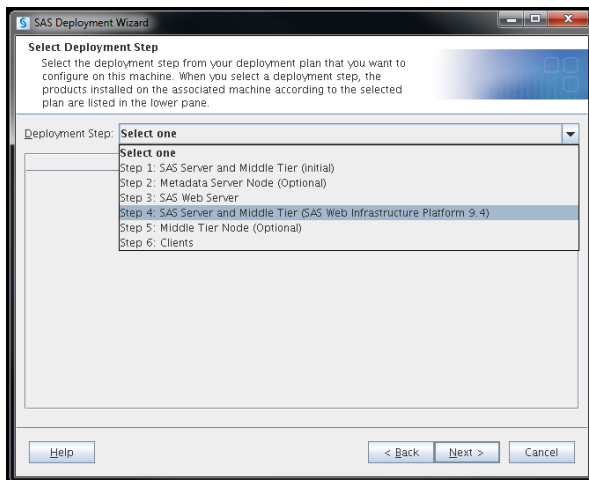

Display 5. Select Deployment Step

On the "SAS Visual Analytics: Allow Guest Access" screen select the "Guest Access Permitted" to allow the guest to access the visual analytics application at a guest URL with the anonymous account.
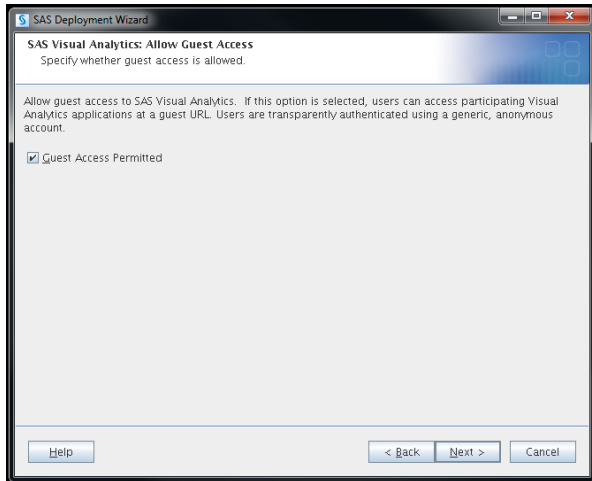
Display 6. SAS Visual Analytics: Allow Guest Access

Below screen select the "HTTP Protocol using Secure Sockets" and mention the SAS Web Server's host name in the Host "Name" field
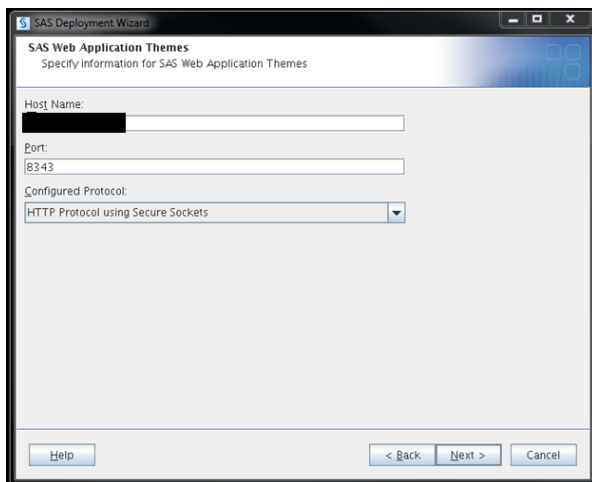
Display 7. SAS Web Application Themes

PERFORM POST-INSTALLATION VALIDATION AND CONFIGURATION TO HANDLE THE CERTIFICATE BETWEEN THE SERVERS

SAS WEB SERVER

Post installation and configuration of the SAS Web Server (SAS Server 05) perform the following steps

1. Import the certificate file to the keystore and put it in opt/sas/sashome/SASPrivateJavaRuntimeEnvironment/9.4/jre/lib/security using below command

   /opt/sas/sashome/SASPrivateJavaRuntimeEnvironment/9.4/jre/bin/keytool -importcert –file ssl_certificate.crt -keystore /opt/sas/sashome/SASPrivateJavaRuntimeEnvironment/9.4/jre/lib/security/cacerts -storepass changeit -alias clientca

SAS WEB APPLICATION SERVER

Post installation and configuration of the SAS Web Application Server (SAS Server 01) perform the following steps

1. Convert all the 3 crt files (ssl_certificate, IntermediateCA & thawte_Primary_Root_CA) to PEM using the below command

    openssl x509 -inform PEM -in IntermediateCA.crt -out IntermediateCA.pem

    openssl x509 -inform PEM -in ssl_certificate.crt -out ssl_certificate.pem

    openssl x509 -inform PEM -in thawte_Primary_Root_CA.crt -out thawte_Primary_Root_CA.pem

2. Create the cacert.pem by concatenating all the 3 files (ssl_certificate, IntermediateCA & thawte_Primary_Root_CA) and copy the file to /opt/sas/config/Lev1/certs

    cat IntermediateCA.pem > cacert.pem

    cat ssl_certificate.pem >> cacert.pem

    cat thawte_Primary_Root_CA.pem >> cacert.pem

3. Import the certificate file to the keystore and put it in opt/sas/sashome/SASPrivateJavaRuntimeEnvironment/9.4/jre/lib/security using below command

    /opt/sas/sashome/SASPrivateJavaRuntimeEnvironment/9.4/jre/bin/keytool -importcert –file ssl_certificate.crt -keystore /opt/sas/sashome/SASPrivateJavaRuntimeEnvironment/9.4/jre/lib/security/cacerts -storepass changeit -alias clientca

CONCLUSION

SAS Visual Analytics with SAS® LASR Analytic Server provides a robust environment for a wide range of business intelligence and analysis activities as it uses the concept of Hadoop Distributed File System (HDFS) as a co-located data provider. The SAS Visual Analytics has a built-in Web server that needs to be installed to create the reports and publish the URL. It opens the door and allows greater flexibility to publish reports using a secured URL to access a VA report. Publishing the reports to the public is securely encrypted and cannot be intercepted. Post-Deployment Tasks for Certificates has to be followed carefully in order to successfully configure the secured URL links for the reports. Incorporating the 'SSL handshake' requires a basic understanding of HTTPS protocol, purchasing the Certificate and establishing the SSL Certificate.

REFERENCES

SAS® Visual Analytics 7.2 Installation and Configuration Guide (Distributed SAS® LASR™)

https://support.sas.com/documentation/onlinedoc/va/#va72

SAS High-Performance Analytics Infrastructure: Installation and Configuration Guide

http://support.sas.com/documentation/solutions/hpainfrastructure/

SAS® High-Performance Deployment for Hadoop 2.8

http://support.sas.com/documentation/solutions/hpainfrastructure/

Managing SAS® Web Infrastructure Platform Data Server High-Availability Clusters

http://support.sas.com/resources/papers/proceedings15/SAS1776-2015.pdf

SAS® 9.4 Intelligence Platform Installation and Configuration Guide Second Edition

http://support.sas.com/documentation/cdl/en/biig/69172/PDF/default/biig.pdf

ACKNOWLEDGMENTS

CONTACT INFORMATION

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

Vimal Raj Arockiasamy
Kavi Global
(847) 387 6760
vimal.raj@kaviglobal.com
http://www.kaviglobal.com

Ratul Saha
Kavi Global
(847) 387 6760
ratul.saha@kaviglobal.com
http://www.kaviglobal.com