Nine Frequently Asked Questions about Getting Started with SAS® Visual Analytics

Tricia Aanderud, Ryan Kumpfmiller, Nick Welke, Zencos Consulting

ABSTRACT

You've heard all the talk about SAS® Visual Analytics—but maybe you are still confused about how the product would work in your SAS® environment. Many customers have the same points of confusion about what they need to do with their data, how to get data into the product, how SAS Visual Analytics would benefit them, and even should they be considering Hadoop or the cloud. In this paper, we cover the questions we are asked most often about implementation, administration, and usage of SAS Visual Analytics.

INTRODUCTION

This paper answers the most commonly asked questions by those interested in adding SAS Visual Analytics to their environment. These questions range from *What is it?* to *What is required to implement and administer the tool?* The examples in this document use SAS Visual Analytics 7.3, which is a Flash-based implementation of the tool. While we have made an effort to provide accurate hardware and software information in this paper, users should consult the SAS Visual Analytics website to ensure they understand the most recent requirements prior to making any purchases.[16]

WHAT IS SAS VISUAL ANALYTICS?

Released in 2012, SAS Visual Analytics is the driving product for SAS in the realm of Business Intelligence and Analytics. A web-based application, SAS Visual Analytics gives users the ability to access, prepare, and present their data from anywhere. When it comes to data visualization and data discovery, SAS Visual Analytics is meant to be an all-in-one tool for a customer's analytical needs.

Once an organization begins to store and manage their data, it opens up a world of possibilities to what they could learn about the data. Typically, this journey starts with spreadsheets and then moves to larger data stores such as data marts and data warehouses. When data is centrally located and pulled from multiple sources, an organization can really have a deeper understanding of processes, customers, and even future direction.

ORGANIZATION BI MATURITY

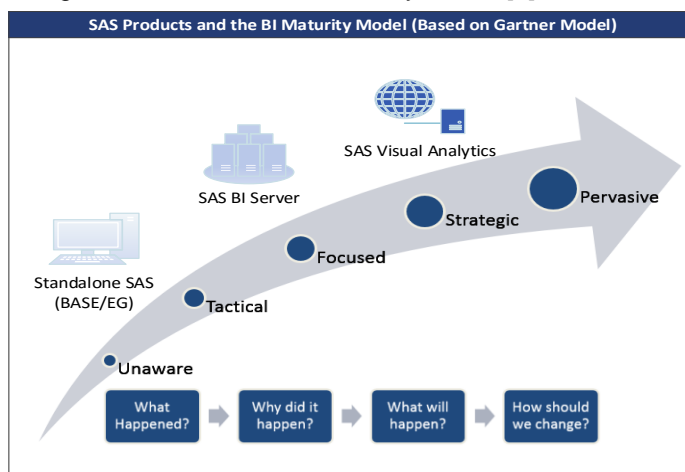In Figure 1, the Gartner BI Maturity Model [5] shows the stages that organizations go through when adopting a data-driven culture. The SAS products are shown above the model to indicate where customers typically add the product during the maturing process. However, customers may have multiple products from different vendors.

Figure 1 Data Driven Organization Maturity Model

The Unaware stage of the maturity model is where you are just beginning to realize that there is value in your data. It includes using spreadsheets for limited analysis, creating one-off reports, and manually collecting data. This stage may be limited to a few employees considered power users because of their in-depth data or process knowledge. The power users are spread through the organization.

The Tactical stage builds upon an organization starting to realize the value from existing data along with adding more data sources. In this stage, you start to realize how much data can enlighten the organization about processes and customers. Power users have most likely left the spreadsheets behind for better data manipulation tools such as Foundation SAS or SAS Enterprise Guide. These tools are great for assembling data, building reports, and completing advanced analysis. However, these tools require programming skills and advanced tool knowledge. During this stage, the organization is still limiting the users to a restricted group of trained employees. There is generally less management support during this stage as managers may still want to trust their guts.

When you shift into the Focused stage, a data-driven culture starts to form within the organization. Power users assemble data to figure out why things are happening. A management stakeholder emerges to lead the organization toward becoming data-driven. Centralization and availability of data is important here since answers to problems could lie in different systems or departments. A data management tool like SAS BI Server allows the organization to centralize data sources and further optimize processing.

Organization may have also centralized data management, reporting, and analysis into Business Intelligence Competency Centers (BICC). These centers are staffed by highly-skilled employees who are responsible for guiding the organization's data and reporting standards. Typically, this is the IT department but can be broadened into business analysts. [15]

In the Strategic stage, the organization has embraced a data-drive culture and is using it to shape business activities. As noted by Eckerson [6], many organizations realize that there are two distinct groups with the organization: Causal and Power users. The Casual users consume the reports from the Power users. However, there is a segment of the Casual users who need more information but lack the skill to use the advanced tools. Eckerson refers to these users as *Super Users* because they understand the business processes but need data to support ad hoc requests for department level concerns. Self-service tools work great for this user and specifically a tool like SAS Visual Analytics.

SAS Visual Analytics allows a deep dive analysis into data to determine why things happened yesterday but also analyze what might occur tomorrow. Being an in-memory application, SAS Visual Analytics can handle drill-downs and report filtering with ease. Super users can get the answers to business questions quickly and show other Casual users how to access information.

In the Pervasive stage, this is where an organization has fully matured and data drives the business strategy. Reaching these high levels of BI maturity is not easy as one of the most crucial parts is having a whole organization on board with a data-driven culture. Successful data driven company's share similar attitudes in that they believe data collection to be an important aspect across all departments, there is buy in from executives at the top, and data should be shared across the company. [11]

SAS VISUAL ANALTYCS APPLICATION OVERVIEW

How is SAS Visual Analytics able to provide all of these capabilities to its end users? SAS has established a unique blend of components under the hood of the application that gives the user control from date source to end analysis or presentation.
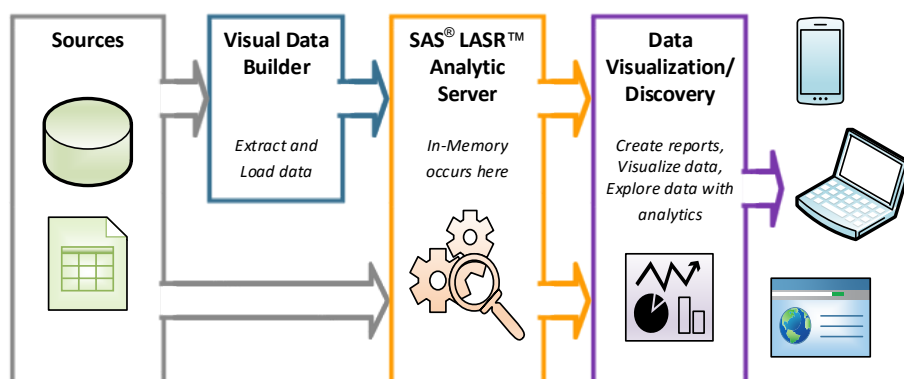

Figure 2 SAS Visual Analytics Overview

As shown in Figure 3, the process from getting data from beginning to end is straight forward. There are multiple ways to get data into the SAS® LASR™ Analytic Server and once data is in the LASR server it can then be used within the web application for data visualization and discovery purposes. The Visual Data Builder and a GUI interface that lets users select, manipulate, and load various sources, is one way to get data into the LASR server. Users can also upload data into LASR server from SAS® Enterprise Guide, Foundation SAS®, or directly from options within the web application. SAS Visual Analytics is also data agnostic. So whether you have text, spreadsheet input, or other types of data, you can load it into the application. With multiple loading options, all with various advantages, the application becomes easy to use for new users even if they have no programming experience.

The LASR server is the key component that makes SAS Visual Analytics so unique. LASR server is an in-memory data storage area that lets users quickly access and combine data. Before LASR server, each time that you queried data or performed a calculation, the computer would have to retrieve data off disk space. This would take up a lot of time, especially if multiple users are on the server at once. With LASR server, all tables are pre-loaded into RAM so the data is easily accessible for various analysis. You also don't need to worry about speed or space running out since additional servers can be added to the system.

Once the data is loaded into LASR server, users can access the data through the web application. Within the application, users can explore data, use advanced analytic tools, and even create reports. There are capabilities that allow users to put additional transformations on their data with calculations, aggregated measures, and hierarchies. Once the data is prepped, there are numerous objects to display the data such as bar charts, bubble plots, geographical maps, etc. All of these objects can also be interactively tied together so that users can filter and drill down into their data.

MAKING ANALYTICS APPROACHABLE

When you hear the industry term *approachable analytics* it makes you think about analytics for anyone. You will always have your normal power users and statisticians, what about your other business analysts? They may not be as mathematically inclined, but know the data better than anyone. What if there was a tool that gave them the statistical power and was user friendly?

That's one of the best aspects of SAS Visual Analytics is the control and ease that it puts into the hands of its users. Whether your end users are looking to play around with their data or build an appealing dashboard, the tools and objects that are available are simple to understand and use. Obviously, the more experience a user has had with similar tools or SAS, the quicker they will pick it up, but even users who have only seen Microsoft Excel are able to use SAS Visual Analytics in a short amount of time.
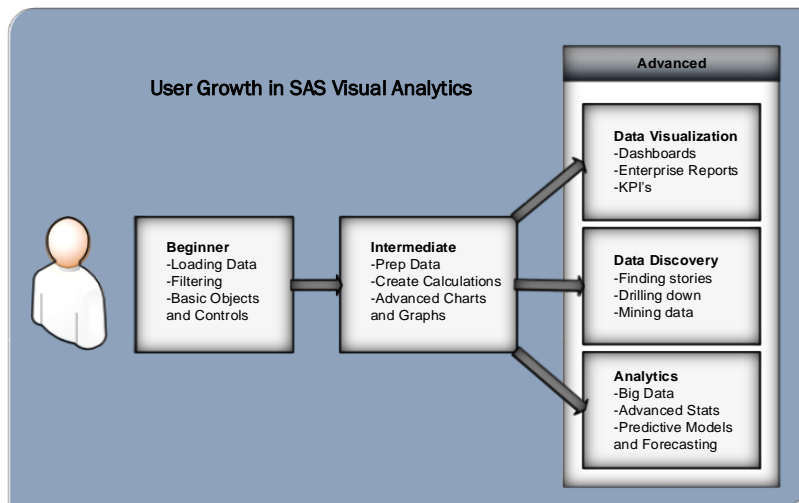

Figure 3 User Growth Cycle in SAS Visual Analytics

As shown in Figure 3, users start by learning how to load data into SAS Visual Analytics and then bring that data into web application where they can use objects and controls to display the data. Once those concepts are developed, users are then able to understand the structure of data that is needed to work with the objects. This is when they'll learn how to prepare the data for the various types of hierarchies, objects, and interactions that they want to use. From here, users can then break out according to their skill level with data and what they are looking to get out of the tool in regards to the data visualization, data discovery, and analytics.

HOW DO USERS ACCESS SAS VISUAL ANALYTICS?

Users can access SAS Visual Analytics in multiple ways. The most popular method is through a Web browser but users can also see the information on a tablet, such as an Apple iPad.

USE A WEB BROWSER

The Web functionality is built into the product and access is controlled by the SAS Metadata server. When users enter the tool they access the SAS Visual Analytics Hub (or homepage), which provides a jumping off point to the other product features. Based on the metadata profile, the user may see some or all reports and access some or all product features. Here's an example of the SAS Visual Analytics Hub that all users can access after logging in – this user is a viewer who can only consume information so they do not have any options to use the data exploration or reporting features. Notice that there are several collections of reports so the user can easily navigate to the content of interest.
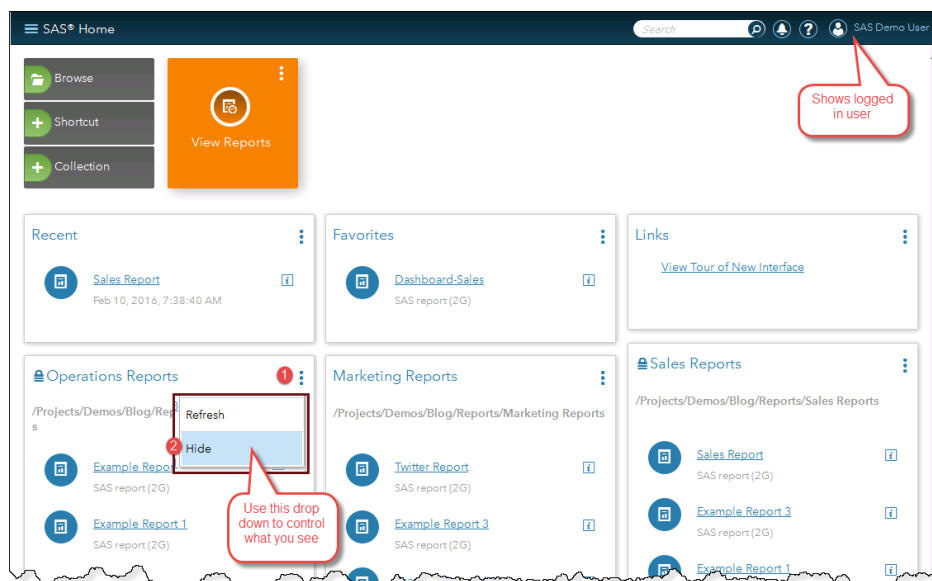

Figure 4 SAS Viewer in the Web Browser

USING THE MOBILE BI APP

User can also access reports from a mobile device. The mobile device management is built-in functionality. You can allow your users to access the portal either within VPN or with just a login. You can adjust the mobile security to ensure the safety of your data. The data is only shown on the device when the user is connected to your network.[9] Figure 5 shows an example of the SAS Mobile BI application.
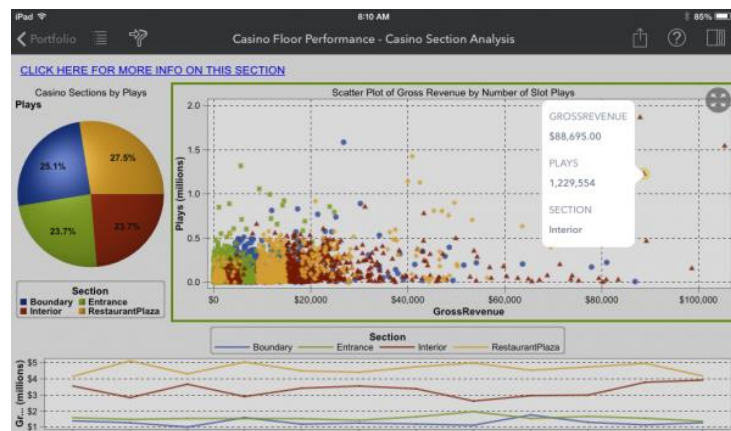

Figure 5 Mobile BI on an Apple iPad

There are several example reports available when you download the free SAS Mobile BI app from iTunes store or from Google Play for Android. In the example, the user's finger is the mouse and by pressing on the screen the user is able to learn more about the individual bubble and can even drill down to get more information.

After learning they can easily distribute reports internally, many users ask if they can share reports with their customers. What a great idea to distribute examples of a report with your customers. Then the customers can also interact with reports

4

you have created using custom data. One marketing agency wanted their customers to see how their advertising dollars were being used. Using SAS Visual Analytics provided the agency project managers had more conversations with the customer about what was happening and how they knew it. It was easier than the older method of using a ton of slide packages or multi-page Excel spreadsheets. SAS Visual Analytics supports this functionality with a tiny bit of finagling and providing a login for the customer.

USE A PUBLIC PORTAL

You might want to make your data available to the public. The public can be anything from your customer base to the entire world. The United Nations has the ComTrade public portal that you can visit. This site allows you to explore the data without any credentials. [19]

You can also use this technique on your intranet where employees can access the reports and use the data discovery techniques without having an additional login. This is built-in functionality that can be configured.

Figure 6 Public Portal Available

WHAT ARE THE DIFFERENT VERSIONS OF SAS VISUAL ANALYTICS?

When looking into buying SAS Visual Analytics, one of the first things to consider is how it is going to fit into your organization's environment. SAS Visual Analytics comes in two distinct versions, distributed and non-distributed.

NON-DISTRIBUTED VERSION

The non-distributed version is a single server deployment of SAS Visual Analytics. This is a simple way to get you up and running with the application. Depending on the data that you are analyzing, this version could be enough for small and midsize corporations. For larger organizations, the non-distributed version can be used as a good test drive before growing out into the larger distributed version. In the following figure, you can see how the single server non-distributed version works.
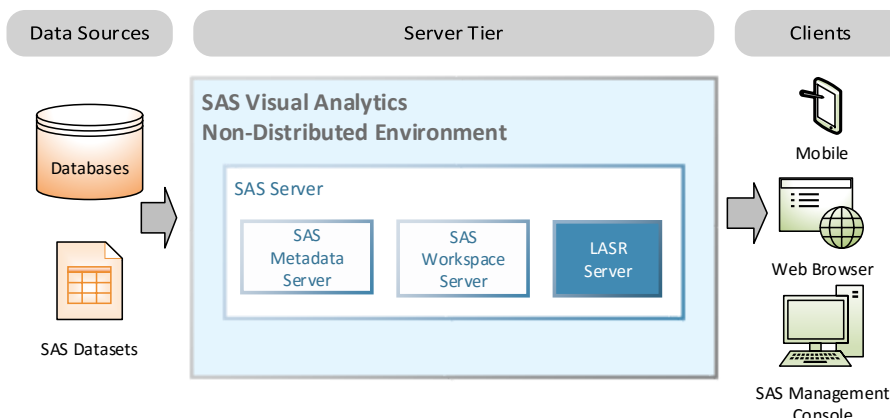

Figure 7 Overview of a Non-Distributed Environment

In Figure 7 you can see that within this deployment there are three parts in the server which all serve a specific purpose. The Workspace Server is in charge of taking incoming data and loading it into the LASR server where it is then stored in memory. The Metadata Server makes sure that user's loading data have the correct permissions.

DISTRIBUTED VERSION

Instead of running on one server, the distributed version spreads the LASR server across multiple nodes. All of the in-memory features of the system stay the same between the versions, the only difference is in how the data is stored in the LASR Server. Figure 8 shows how the distributed environment is set up.
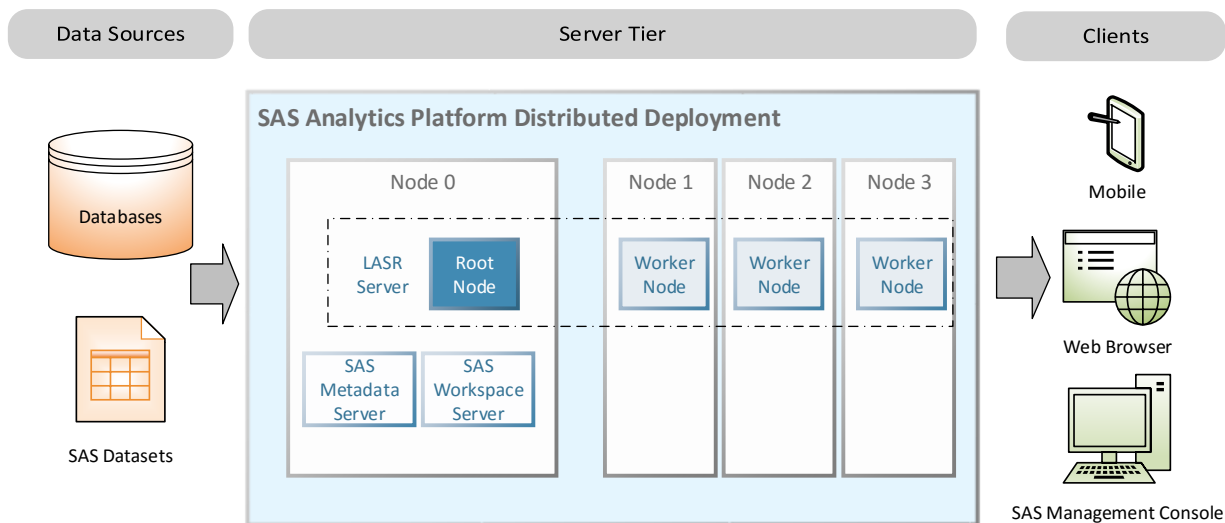

Figure 8 Overview of Distributed Environment

Just like the non-distributed version, all of the components can reside on the main server. The only exception now is that the LASR server is now distributed across a root node and multiple worker nodes. The root node (Node 0) contains no data but can be considered as the controller of the worker nodes (Node 1-3). It is responsible for determining how incoming data should be spread across the worker nodes as well as remembering where all data resides.

With the worker nodes, the distributed version has the benefit of being easy to scale. If you ever need more disk space or processing power, it is minimal effort for the administrator to add another node to the current system.

WHAT IS WITH THE ELEPHANT?

That elephant is referring to Apache™ Hadoop®.[2] Hadoop is an open source computing platform written in Java. Started in the early 2000s as a way to more effectively mine through web pages for search engines, Hadoop has developed over the past decade to become a powerful method of how to store and process large amounts of data. SAS Visual Analytics has an option with the distributed version to use Hadoop's storage capability and depending on the system that's being implemented and data being processed, Hadoop could vastly help out with performance. There are a number of benefits with using Hadoop, but the main reasons to consider come from its low costs, flexibility, and speed of processing data.

HOW DOES HADOOP WORK?

The main point of Hadoop is to store and process data faster than using traditional methods. This is done by the main component of Hadoop, the Hadoop Distributed File System (HDFS). The HDFS works by having a name (parent) node facilitate incoming data through data (child) nodes.
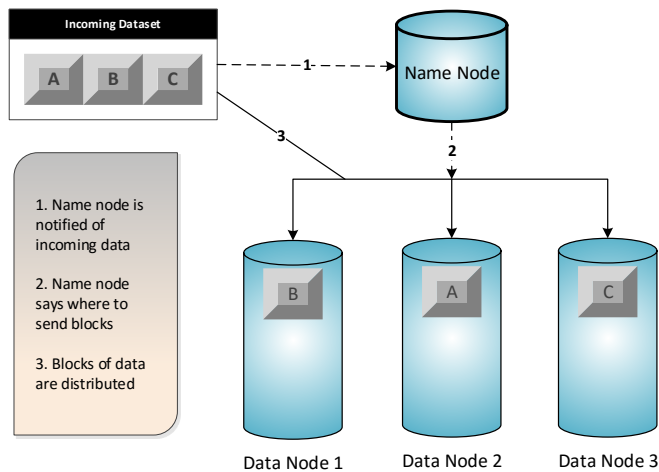
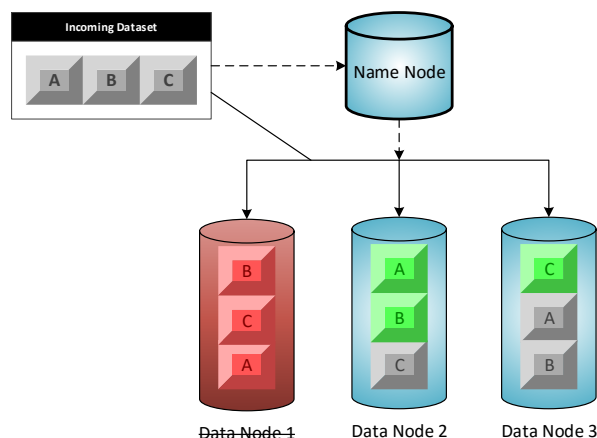*Figure 9 Loading data into HDFS*

In Figure 9, the incoming dataset is split up into blocks, A, B, and C. The Name Node is notified that this data is incoming and then analyzes how it wants to disperse the data blocks. Once it has been decided where the blocks will go, the blocks are then distributed into their respective data nodes from the source.

This is where Hadoop is able to harness its processing power. By having the data spread across multiple nodes and being able to use the computing power of each node, the route to get data in and out of the file system becomes much more efficient since everything is distributed.

Now you may be thinking what happens if one node goes down. Are all of your datasets now going to be missing or incomplete? Fortunately, the HDFS has a plan for that and keeps layers of replicated back up data.

In Figure 10, you can see that now each of the data nodes contains additional copies of the data blocks that were distributed in the first diagram. When the Name Node determines where to distribute the data it also takes care of where to direct the backups as well since it would be illogical to put the back up of block A in the same node as the original block A. Now let's take a look at what would happen if a node went down.
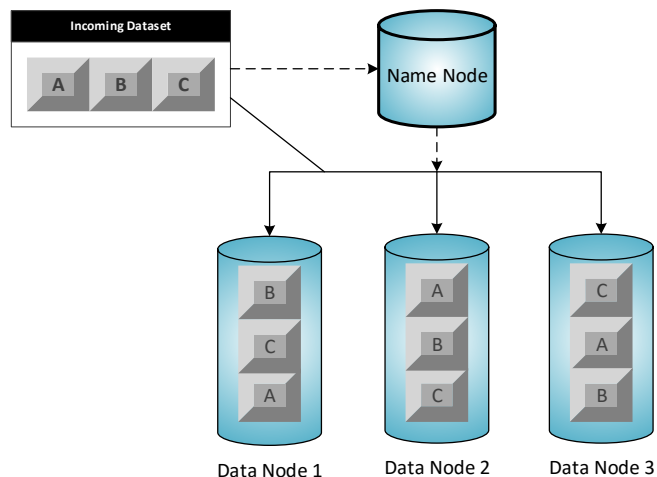
Figure 10 Hadoop data replication

Figure 11 Hadoop Missing a Node

If Data Node 1 went down, then the Name Node would recognize that and reroute any requests for Block B to where its first backup is stored.[4] In this case, that is Data Node 2. Now, all of the blocks in green will be used as the original copy of the data and the Name node will keep track of that. If Data Node 1 comes back up or is replaced by another node, then Hadoop has a balancing utility that can be used to disperse all of the data evenly across the nodes again. With the ability to have multiple copies of data and be able to add and subtract nodes with ease, this is where Hadoop gets its flexibility. It can handle growth by adding nodes and then rebalancing the data and it can also be reliable when parts break down.

The last benefit is the low cost of using Hadoop. For starters, Hadoop is available as open source software so there are many flavors of it out there available for free. There are also some companies that

offer services or their own licensed distribution. The hardware is where organizations can really see a difference in cost. The servers used to house the name and data nodes are all commodity hardware. They are simple servers that can cost typically as low as $3,000 (with limited storage). Therefore, if one went down for good, similar to the example above, your cost to get a new one and bring the system back to its original state is very minimal.

WHERE DOES IT FIT INTO SAS VISUAL ANALYTICS?

Now that we have covered how the distributed version of SAS Visual Analytics works and how Hadoop works, we can look into how they work together to create a robust environment.

The distributed version is capable of adding a co-located data provider to the system. [17] Hadoop is one of the options to use as a data provider. It is important to remember that this is only for the distributed version since Hadoop and the distributed version work in similar manners. The figure below depicts one example of how the environment could be setup.
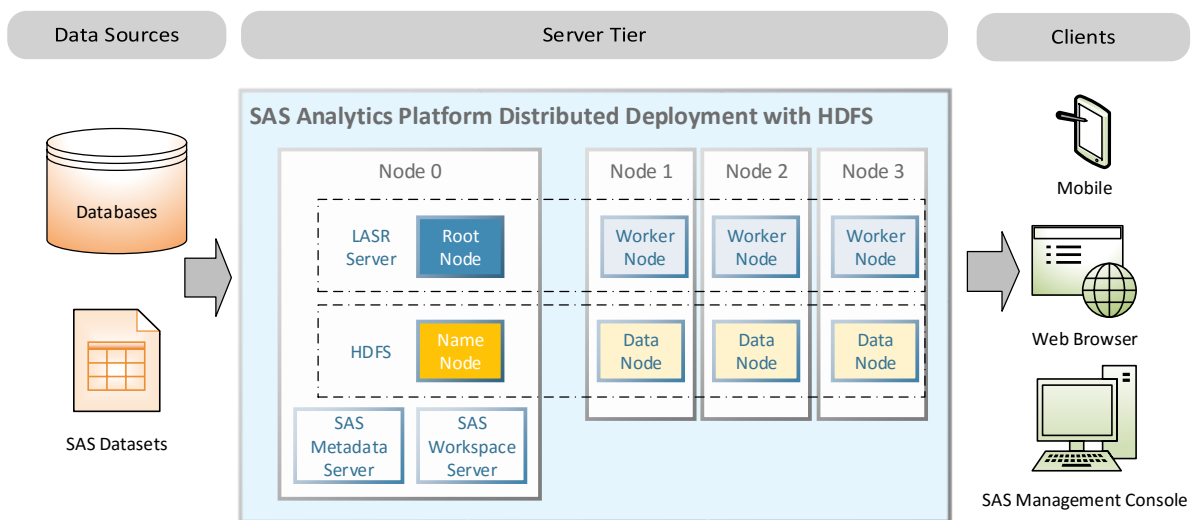

Figure 12 Overview of Distributed Version with Hadoop

When you have the distributed version, HDFS, and co-located storage (Worker and Data nodes on the same machine) up and running you can start using the SASHDAT format for files. When data tables are loaded to the SAS LASR Server they can be stored in SASHDAT format, which is essentially distributed data. The benefit of the SASHDAT file is that the data is laid in the HDFS the way it is looks in memory. This is an optional feature in the system and can only be used when in a co-located environment.

With the HDFS in the system, there are now two different ways that data can be loaded into LASR, serial or parallel. When using just the LASR Server, data is loaded serially into LASR. This means that it goes straight from source, through the Workspace Server, and into LASR. Now that HDFS is an option, data can be loaded in parallel instead. Instead of going straight to LASR, data will go to the HDFS Name Node from the Workspace Server, get distributed among the HDFS Data Nodes, and then get copied into the Worker Nodes of the LASR Server. All data does not have to be processed in parallel, you have the option to use either loading process. For smaller tables that are quick to load, there might not be much of a benefit to load them into the HDFS Server as well, but there can be a significant difference as tables get bigger.

When talking about handling Big Data the parallel loading process is how it can be accomplished efficiently. Remember that the LASR Server is in-memory so when it goes down or gets restarted, all tables in the server need to be reloaded. In the serial process, all tables will have to be reloaded from their source that could be somewhere else in SAS, from a database, stored on a local machine, etc. By using the parallel process, the tables will still be sitting within the system on the HDFS data nodes. They can quickly be copied back over to the LASR Server after any interruption without any need to reload/process data from its source.

HOW DO I GET MY OWN ELEPHANT?

Apache Hadoop is an open source platform that is available as a free download from their site. Also, several companies have created their own supersets of Hadoop that are available for free. In the way that SQL is a standard for querying data and there are various flavors of it, Hadoop is also a standard framework that has been extended by vendors.

SAS provides the SAS High-Performance Deployment of Hadoop which is their version of Apache Hadoop as the default. This is sufficient to get started using Hadoop but there are also other vendors that SAS supports as options to implement and manage your Hadoop instance. Below are the more popular Hadoop commercial vendors supported by SAS: [16]

- Cloudera CDH 4.2.0 and higher
- Hortonworks HDP 2.0 and higher
- MapR distribution 4.0 and higher

Cloudera and Hortonworks were two of the first companies to begin making their own distributions of Hadoop. Both distributions are open source and built on top of Apache Hadoop. In addition to the distributions, each company has software packages to help manage your environment. Hortonworks and Cloudera offer an administration console that makes it easy for users to understand what is happening in the system. This is one feature that makes these two companies a popular choice when using Hadoop.

The other company, MapR, decided to go away from the HDFS (name node to data node) format with their distribution and use their own, MapRFS, file system. [18] With this file system, the name node is distributed across all of the nodes instead of the hub and spoke model shown earlier.

These three companies are just some of the more prevalent versions that we have seen used in the industry. Each one also offers training, support, and consulting to assist in working with Hadoop. Beyond them, there are other companies and versions of Hadoop available to use. If your organization's data storage and management needs are becoming more Hadoop focused, then they would definitely be worth looking into further. However, SAS Visual Analytics works fine with the SAS High-Performance Deployment of Hadoop.

DO I NEED NEW HARDWARE?

Since SAS Visual Analytics relies on memory for speed a power server is a must. Before purchasing any hardware, it's a good idea to discuss your usage and data size with SAS or another professional for proper server sizing.

HARDWARE REQUIREMENTS

The following table provides an overview of the minimal hardware for common deployments according to the SAS Visual Analytics 7.3 System Requirements. [16]

	Non Distributed Deployment Minimum Configuration	Distributed Deployment Minimum Configuration
Server	1 Server - 4 Cores	4 Servers – 16 Cores (4x4 cores)
RAM	64 GB	256 GB (64 GB per server)
Storage	A minimum of 30GB is required for the operating system and application software. The remaining space on the data nodes would need to be sized based on your data requirements.	
Considerations	All software is contained on this machine. The number of user and amount of data are key limiters.	4 servers: This setup uses one server that shares the SAS Server and head node and 3 data nodes.

USING COMMODITY HARDWARE

The SAS Visual Analytics website lists Intel Xeon architecture as preferred, which is available in commodity hardware. [16] Commodity hardware generally refers to inexpensive standardized servers that are easy to purchase from a variety of vendors. These servers do not have a specific purpose and can be used as Windows or Linux based machines.

Some organization prefer to have their development or test environments on commodity machines, while higher-end machines (such as HP or Dell) are used for production environments. The production machines have support contracts attached so any disruptions are kept to a minimum.

Selecting an Operating System

The non-distributed system can be deployed on a Windows or Linux server, but the distributed version is only available for Linux. [16] Of the deployments we have seen, most customers have tended to go with Linux even if they were primarily a Windows-based shop. Linux gives administrators easy control with memory management, no registry to worry about, and native support for components.

It is also possible to use a combination of operating systems. Many customers use Linux for the SAS Visual Analytics backend and a Windows environment for the client applications like the Web browser, SAS Management Console, and SAS Enterprise Guide.

ALTERNATIVE IS A CLOUD ENVIRONMENT

A cloud environment may be a good alternative for your SAS Visual Analytics deployment. Cloud deployments are increasing in popularity and acceptance according to the Gartner 2015 CIO survey.[7] If you are using a service like Evernote, Dropbox, or Instagram then you are already familiar with accessing data from a shared environment hosted on a remote server. Some organizations choose a cloud deployment to keep hardware costs low or just to experiment with the software before making a purchasing decision.

What Kind of Service Do I Need?

Essentially a cloud service is renting space on someone else's service. This means the cloud vendor is responsible for providing the infrastructure (such as networking, hardware, and data storage space). This method is referred to as Infrastructure as a Service (IaaS). Many users prefer to use the cloud service so they avoid capital purchases and have the flexibility to purchase more space as it is needed.

The most obvious advantage to cloud computing is that you only pay for what you need when you want it. Since you are renting the space on someone else's server and you can pay only when the server is active. If you need more space, you just request it. The cloud vendors offer many options base on the server size required and usage period. Refer to the Redpath and Eayrs, *SAS Visual Analytics for the Three Cs: Cloud, Consumerization, and Collaboration* paper for a more detailed explanation of cloud environments. [13]

The only disadvantages of the cloud environments are how you move your data to the environment and you still must maintain the software and database administration piece. Some users are getting around this requirement by using the private cloud from SAS Institute. SAS Cloud provides the administration along with hosting the environment.

Who is the Best Cloud Provider?

Gartner cites Amazon Web Services (AWS) as having "more than 10 times more cloud IaaS compute capacity in use than the aggregate total of the other 14 providers Gartner studied." As more vendors have entered the market the prices have continued to become more attractive. Gartner also noted these other vendors Google, Microsoft Azure, or Rackspace as common choices.[10]

The most important thing to consider when choosing a vendor is if they can meet the deployment requirements for a SAS Visual Analytics implementation. There are other factors to consider, such as services offered to support future growth, the pricing model, and the support model. If your environment is essential to business operations, then consider the service level agreement the vendor supports. This vendor is becoming a part of your organization so a good fit may be essential to your business

operations. [3] Gartner also advised that the market is becoming more competitive and some of the smaller vendors may not survive, so also consider the vendor's financial health and ability to support long-term goals and planning. [7]

The cloud environments are a good choice for short term usage, those experimenting with the tool, or those who need additional IT support.

WHAT LEVELS OF SECURITY EXISTS?

You'll want to create a security model, which involves determining the user permissions within the SAS Visual Analytics system. Most customers have the same basic needs - some users who view reports and some users who prepare the reports. There can be a lot of variation after that point. For instance, in a smaller organization the user may create reports and administer the system, while larger organizations would have designated resources for each role.

MANAGING SECURITY THROUGH METADATA

SAS Visual Analytics uses the metadata to control user access to objects and data. By assigning users to groups and then pre-defined roles to those groups, you can control what each user (as a group member) can view or edit. For instance, everyone may have read privileges to a production report to prevent accidental changes to the report. Thus, even users with advanced edit privileges are forced to save the report to another folder before making changes. This is a simple example and definition of metadata, for a more detailed description refer to *SAS® Intelligence Platform: Security Administration Guide.*

WHAT'S IN A PRE-DEFINED ROLE?

SAS Visual Analytics ships with five roles. Each role is composed of a pre-defined set of capabilities for different things a user or set of users can do, such as consume reports, build reports, and administer the system. The roles follow along what you might expect most organizations would desire for organizing their users: consumer, analyst, report builder, data builder, and administrator.

In Figure 13, notice as that as the role becomes more complex it inherits the capabilities of the roles before it. Each role can be added to a group that creates your security model. Some companies divide the metadata by department (HR, Sales, and so on) while others have logical groups based on product lines or teams. The architecture is flexible enough to support either method.

Figure 13 SAS Visual Analytics Predefined Roles

Applying the SAS Visual Analytics Security Model

When creating a security model, you can assign these roles to groups. This allows you to control what the users are allowed to do within the SAS Visual Analytics system. Notice I didn't say anything about permissions. Roles only say what abilities a user or group has. Permissions are set in ACEs (like folders) or with an ACT.

In Figure 14, you can see an example of a large complex organization where the roles are more defined. The team has the four groups that essentially align with the provided roles. You can use the groups to divide the user by department or logical organization. Refer to *What's in a role?* post for more examples of security models. [1]
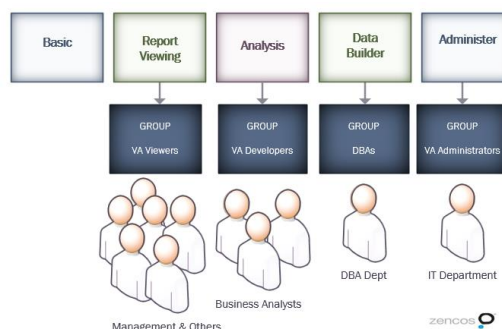

Figure 14 SAS Visual Analytics Roles Applied

HOW DO I LOAD MY DATA INTO SAS VISUAL ANALTYICS?

What customers love about the SAS products is that they are data agnostic. This simply means SAS can read your data no matter if it is in a database or a flat file. SAS Visual Analytics is no exception; you can load a SAS dataset, a text file with delimiters, or an Excel spreadsheet. You can also download social media data from sites such as Twitter or Web data, such as Google Analytics. With one of the ACCESS engines installed, you can connect to any number of data sources. This makes the data you currently have immediately useable to the tool.

HOW DO YOU LOAD DATA?

There are multiple data loading methods available for the SAS Visual Analytics. The most common methods are AutoLoad process, SAS Visual Analytics Import process, SAS Visual Analytics Data Builder, and a coding method. Some methods allow you to load from the HDFS while others have a direct path to the LASR Server. Most organizations support multiple paths to keep the tool access flexible.

Using the AutoLoad Process

Users can move data to a designed server folder to have the data uploaded, appended, or unloaded from the LASR Server. Many sites have users FTP the data to the autoloader. The user does not need a host account for this method or any special metadata permissions. This method supports serial loading.
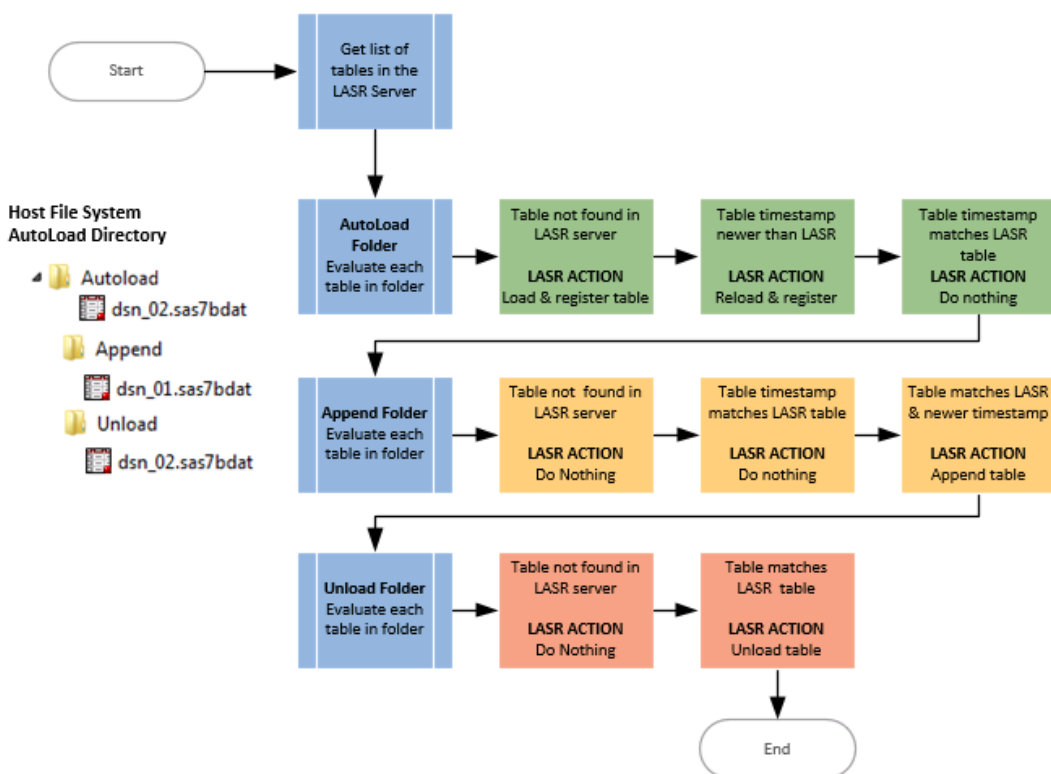
Figure 15 Importing Data with the AutoLoad Process

The AutoLoad feature loads or unloads data from a defined host folder into the LASR server. The data can be a delimited text file, a spreadsheet, or a SAS dataset. The administrator schedules the Autoloader process to run at regular intervals, such as every 10-15 minutes. The Autoloader has three folders: Autoloader, Append, and Unload. Users can place data in each folder based on desired action and the Autoloader process evaluates the data based on its name and its timestamp to determine if it is loaded or unloaded. The following figure shows what happens to the data when placed in each folder.

Things to Consider

- Administrators should consider implementing this method as a user convenience feature.

- Add monitoring to the process so that any issues are quickly identified.

Allowing Users Self Service Import

Users can load data through the SAS Visual Analytics User Interface from a local server, a database connection, another Hadoop instance, or from a social media network – referred to as self-service import or interactive loading. Users require appropriate metadata permissions and a host account. This data is loaded in a serial method.

Some of these methods rely on the SAS Workspace Server and some use the HDFS directly. The HDFS method supports parallel data processing so it's faster. If you have a smaller dataset that takes only a few minutes to load you may not care as much as when you are loading several large datasets that take 15-plus minutes per dataset each time.
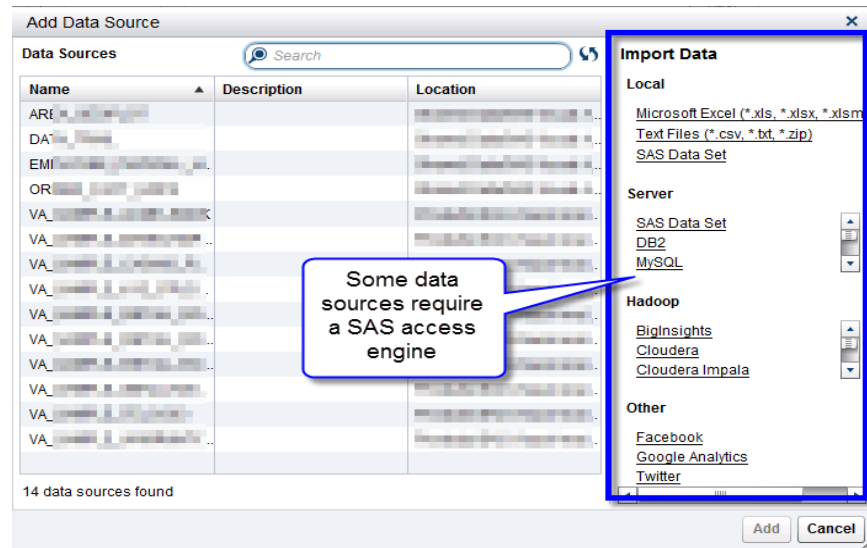
Figure 16 Self Service Import

Things to Consider

- Administrators need a strategy for preventing the LASR server from overloading.

- Users need training to understand how the features work.

Loading with the Data Builder

The Data Builder is a module of SAS Visual Analytics. It's provides a point-and-click interface to the HDFS and LASR server. The Data Builder can create and schedule basic queries. Typically, the feature is limited to data administrators. This method can be serial or parallel depending on how the query is written.

The Data Builder allows the user to query data from a local file, a database, or even the HDFS and save the data in multiple locations – such as HDFS, LASR server, or library. Once the query is written, the user can save it to a file and then re-run as desired. For example, a data query of the COMPLAIN dataset is loaded to the LASR server. When the query is saved, it creates the metadata definition of the dataset in the specified location. When the query is run, the data is physically loaded to the location.

Things to Consider

- Establish a trained group of data administrators who have access to the Data Builder. Depending on the size of your organization, this may be a single person or a group of users.

- Many times it's better to copy the Data Builder code and modify it for a batch process.

Loading Data with the Code Editor

There are several coding methods you can use to load to SAS Visual Analytics based on your deployment. Some of the coding methods are available are DATA step, PROC LASR, and PROC IMSTAT. This code can be run on any editor that has a network connection and permissions to the SAS

LASR Server – which may be SAS Studio or SAS Enterprise Guide. Several other techniques for loading data with code are discussed in the *Tips and Techniques for Efficiently Updating and Loading Data into SAS Visual Analytics* paper. [14] Here's an example of code that loads a dataset from a BASE library to the LASR Server.

```
libname myLASR SASIOLA host="va.myserver.com" port=10031 tag="vapublic";
libname vapublic BASE "//data/mypath";

data myLASR.baseball_american;
set vapublic.baseball;
run;

proc metalib;
  omr (library="Visual Analytics Public LASR");
  update_rule=(noupdate);
  select(baseball_american);
run;
```

Figure 17 Importing Data to the LASR Server with Code

Things to Consider

- While seemingly simple, this method requires the most training. The user needs a complete understanding of the LASR server and metadata management.

How Do You Manage and Monitor the Environment?

There are multiple ways to manage the environment based on what you need to complete. From within SAS Visual Analytics, you can use the Administration tab to manage the data and permission. The SAS Environment Manager allows you to administer and monitor the environment.
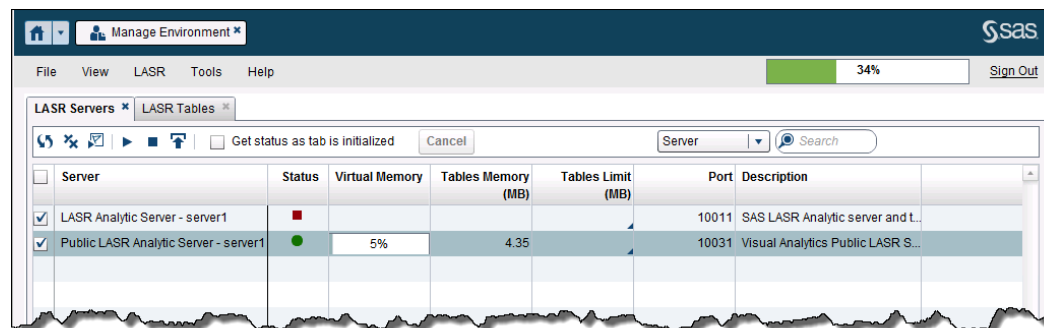
MANAGING THE DATA AND OBJECTS

If you want to manage the data and user access you can use the SAS Visual Analytics Administrator.

Managing the In-Memory Data

The LASR Server can be managed through the SAS Administration interface. This interface shows the status of the LASR Servers along with the memory usage. You can start and stop the LASR server from here. In Figure 18 there are two default servers available: LASR Analytics Server and the Public LASR Analytic Server. You can define your own LASR Server when increased security or organization is required.

When the status is green, it indicates the servers are running. A red block would show the servers as stopped. When a server is stopped, all data is released. This is different from other servers where the data is based on the file system and is only removed by a delete command. This is an easy way to restart the LASR server when needed.

Figure 18 Managing the LASR Servers

In a separate tab called LASR Tables, you can see the content of all the LASR Servers. You can use this tab to register data in libraries, save data to the HDFS, and load data to the LASR server. In Figure 19, you can see a list of the data tables in the Public LASR Server. Remember only data that you have registered in metadata is available. The red boxes indicate that a metadata definition of the tables exists but there is no data available in memory. The green dot indicates the table is registered and has in-memory data available for reports. More details about managing the server can be found in the *SAS Visual Analytics Administration Guide*.

Figure 19 Managing the LASR Server Tables

Managing the HDFS

From the HDFS tab, you can see all of the data loaded in the HDFS along with details about the file, such as block size and owner. This makes it easier to see the contents and delete anything that is no longer needed. You can also load data to the HDFS from the Folder tab (shown in Figure 20Figure 21).

Figure 20 Managing the HDFS

Managing Folder, Data, and Object Permissions

You can manage the user permissions from within the Administrator page. A user with privileges can explore the permissions or make metadata changes on the fly. In this example, the administrator clicked Sales Report folder on the left and selected Authorize. The window shows the metadata permissions for the object and allows the administrator to make adjustments – such as Deny or Grant.
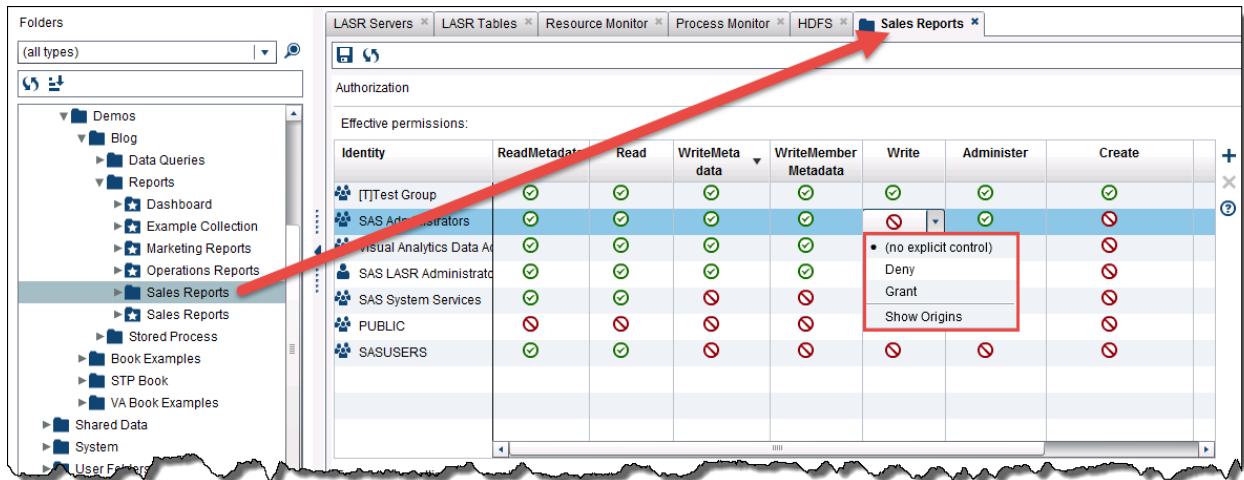

Figure 21 Managing Object Permissions

MONITORING THE ENVIRONMENT

To monitor the environment, SAS Environment Manager provides multiple ways to monitor the SAS platform. The application allows you to administer and monitor SAS resources, collect and chart data about the resources, monitor logs and create alerts. Peters provided an in-depth overview of SAS Environment Manager and its architecture in her *Monitoring 101: New Features in SAS 9.4 for Monitoring Your SAS Intelligence Platform* paper. [12]

Figure 22 contains an example of the tool monitoring the SAS Metadata Server. The charts show the availability and provide some historical data.

Figure 22 SAS Environment Manager provides monitoring [Artwork: SAS Environment Manager Reference Guide]

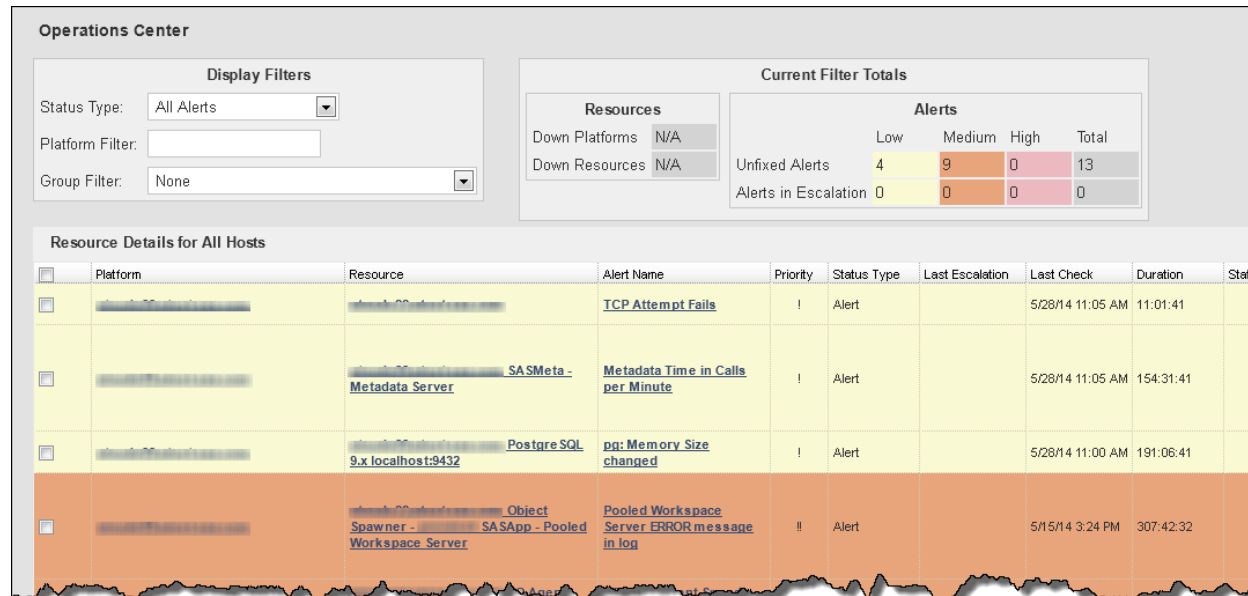You can also view alerts from the Operation Center to assist when troubleshooting issues.

Figure 23 SAS Environment Manager Alerts [Artwork: SAS Environment Manager Reference Guide]

WHAT DO I NEED TO CONSIDER FOR IMPLMENTATION?

If you are about to implement SAS Visual Analytics, the process complexity depends on where your SAS footprint already lies with other SAS products. In most cases, you either are coming from standalone SAS or from the SAS EBI Platform. The EBI Platform is an easier set-up since the framework is similar from SAS EBI Server to the SAS Visual Analytics environment. Standalone SAS is slightly more involved.

COMING FROM STANDALONE SAS

Going back to Figure 1, the natural progression over time would have been to go from Standalone SAS to the BI Server and now to SAS Visual Analytics. The good thing is that if you are still running a standalone version of SAS then you don't have to go to the EBI Server and can go directly to SAS Visual Analytics.

One of the first things to realize when moving on from standalone SAS is that you are moving into a centralized data and reporting structure. No longer is everything going to be stored in siloes on user's desktops. With this unified approach, there is going to need to be resources to manage it. There's the SAS Metadata Server for user and data management, the LASR server for processing and storage, and the web server to host the application, which will all be new to your organization. Nelson provided an excellent overview of the SAS BI Platform in his "Introduction to the SAS® 9 Business Intelligence Platform: A Tutorial" paper. [8]

An administrator is needed to not just take care of these systems but also keep them secure. Since there will be plenty of learning and testing with this approach, it is probably a good idea to start with the non-distributed version to get your feet wet. Many users take this intermediate step with the non-distributed version so they can overcome any data or user hurdles in a smaller environment.

With the need for resources to manage the new environment also comes opportunity and benefits for an organization that is just now starting to mature with their data. It will be easier to incorporate data management policies when everyone is using the same shared services. When it comes to structure, now you can have clearly defined roles for all users and be able to govern what they can do within the system. Lastly, you will now be moving towards a system that is reliable and can scale with the growth of your organization. The servers can always be running to allow access to data and reports at any time and if you need to add more space or power, that's just an added node away.

COMING FROM THE EBI PLATFORM

If you are upgrading from the EBI Platform, then the process is more straight-forward. SAS Visual Analytics integrates with the current environment so if you are running SAS 9.4 TS1M2 then you can just plug it into your existing system. Vitron and Holman in their Considerations for Adding SAS® Visual Analytics to an Existing SAS® Business Intelligence Deployment paper provide more details about the architecture changes.[17]

If you are on an earlier release, then you have some options. You can either upgrade the whole platform to SAS 9.4 or above, or you can implement a dedicated server for non-distributed SAS Visual Analytics to try it out. The Metadata Server can be integrated between the two systems so you will keep all of the users from the EBI platform, however new groups and folders will need to be setup for the SAS Visual Analytics side.

Based on the version of SAS Visual Analytics that you decide to implement you will need additional hardware. If you go with the non-distributed version, the only additional thing you will need to purchase is a node to host the LASR Server. If you upgrade to the distributed version, then the host node as well as the data nodes will need to be purchased and each one would require some additional configuration.

CONCLUSION

In conclusion, SAS Visual Analytics is a powerful tool available to discover, analyze and visualize your data. It is part of a maturing organization needs to have self-service analytics and data discovery. However, it can be a complex tool to implement and we suggest that you don't go it alone – hire an experienced team to assist in assuring your organization gets maximum value from this investment.

REFERENCES

[1] Aanderud, T. "SAS Visual Analytics Administration: What's in a SAS Visual Analytics Role". Zencos Blog. March 2015. Available at: https://www.zencos.com/blog/sas-va-administration-whats-in-a-role

[2] Apache Software Foundation. Available at: http://hadoop.apache.org/

[3] Bystock, C. "What to Consider When Looking for IaaS Providers", Windsor Group Sourcing Advisory, April 2013. Available at: http://www.windzr.com/blog/bid/282955/What-to-Consider-When-Looking-For-IaaS-Providers.

[4] Collum, R. 2015. "The Impact of Hadoop Resiliency on SAS® LASR™ Analytic Server" Proceedings of SAS Global Forum 2015, Dallas, TX: SAS Global Forum. Available at: http://support.sas.com/resources/papers/proceedings15/SAS1760-2015.pdf

[5] Howson, C. "ITScore Overview for BI and Analytics", Gartner, September 2015. Available at: https://www.gartner.com/doc/3136418/itscore-overview-bi-analytics.

[6] Eckerson, W. "The Secrets of Self-Service BI", BeyeNetwork, Janaury 2011. Available at: http://www.b-eye-network.com/blogs/eckerson/archives/2011/01/the_secrets_of.php

[7] Moore, S. "Gartner Says Worldwide Cloud Infrastructure-as-a-Service Spending to Grow 32.8 Percent in 2015", Gartner, May 2015. Available at: http://www.gartner.com/newsroom/id/3055225.

[8] Nelson, G. "*Introduction to the SAS® 9 Business Intelligence Platform: A Tutorial*", Proceedings of SAS Global Forum 2007, Orlando, FL. Available at: http://www2.sas.com/proceedings/forum2007/207-2007.pdf

[9] Nori, M. "How Mobile Changes the BI Experience", Proceedings of SAS Global Forum 2013, San Francisco, CA. Available at: https://support.sas.com/resources/papers/proceedings13/053-2013.pdf

[10] Olavsrud, T. "*Top cloud Infrastructure-as-a-Service vendors*", CIO, July 2015. Available at: http://www.cio.com/article/2947282/cloud-infrastructure/top-cloud-infrastructure-as-a-service-vendors.html

[11] Olavsrud, T. "Data-Driven Companies Outperform Competitors Financially",CIO, March 2013. Available at: http://www.cio.com/article/2387460/business-intelligence/data-driven-companies-outperform-competitors-financially.html

[12] Peters, A (and others). "Monitoring 101: New Features in SAS 9.4 for Monitoring Your SAS Intelligence Platform". Proceedings of SAS Global Forum 2013, Orlando, FL: SAS Global Forum. Available at: http://support.sas.com/rnd/emi/SASEnvMgr/EV_SGF_Paper_Monitoring_101.pdf

[13] Redpath, C and Eayrs, N. 2014. "SAS Visual Analytics® for the Three Cs: Cloud, Consumerization, and Collaboration" Proceedings of SAS Global Forum 2014, Washington, DC: SAS Global Forum. Available at: http://support.sas.com/resources/papers/proceedings14/SAS298-2014.pdf

[14] Rivers, K and Redpath, C. 2015. "Tips and Techniques for Efficiently Updating and Loading Data into SAS® Visual Analytics" Proceedings of SAS Global Forum 2015, Dallas, TX: SAS Global Forum. Available at: http://support.sas.com/resources/papers/proceedings15/SAS1905-2015.pdf

[15] Rouse, M. "Definition: Business Intelligence Competency Center (BICC)", TechTarget Network, July 201. Available at: http://searchbusinessanalytics.techtarget.com/definition/business-intelligence-competency-center-BICC

[16] SAS Institute SAS Visual Analytics 7.3 System Requirements. Available at: http://support.sas.com/documentation/installcenter/en/ikvisanlytofrsr/68904/HTML/default/index.html

[17] Vitron, C. and Holman, J. 2014. "Considerations for Adding SAS® Visual Analytics to an Existing SAS® Business Intelligence Deployment" Proceedings of SAS Global Forum 2014, Washington, DC: SAS Global Forum. Available at: http://support.sas.com/resources/papers/proceedings14/SAS146-2014.pdf

[18] Unknown (Experfy Editor), "Cloudera vs Hortonworks vs MapR: Comparing Hadoop Distributions" Experfy, September 05, 2014 https://www.experfy.com/blog/cloudera-vs-hortonworks-comparing-hadoop-distributions/

[19] United Nations Comtrade Site. Available at: http://www.sas.com/software/visual-analytics/demos/un-comtrade-basic.html

ACKNOWLEDGMENTS

RECOMMENDED READING

- *SAS® LASR Analytic Server Reference Guide,* latest version

- *SAS® Visual Analytics User Guide,* latest version

- *SAS® Visual Analytics Administrators Guide,* latest version

- *SAS® Intelligence Platform: Security Administration Guide*, latest version

- *SAS Environment Manager Administration: User's Guide, latest version*

- Aanderud, T and Homes, M. 2014. "SAS® Admins Need a Dashboard, Too" Proceedings of SAS Global Forum 2014, Washington, DC: SAS Global Forum. Available at: http://support.sas.com/resources/papers/proceedings14/1247-2014.pdf

- Bentley, J. "Metadata: Everyone Talks About It, But What Is It?" Proceedings of SUGI 26. Long Beach, CA, SAS Users Group International. Available at: http://www2.sas.com/proceedings/sugi26/p125-26.pdf

- Kent, P. 2013. "Hadoop and SAS" Proceedings of SAS Global Forum 2013, Orlando, FL: SAS Global Forum. Available at: https://support.sas.com/resources/papers/proceedings13/402-2013.pdf

CONTACT INFORMATION

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

Tricia Aanderud
Zencos Consulting
Cary, NC
taanderud@zencos.com
www.zencos.com

Ryan Kumpfmiller
Zencos Consulting
Cary, NC
rkumpfmiller@zencos.com
www.zencos.com

Nick Welke
Zencos Consulting
Cary, NC
nwelke@zencos.com
www.zencos.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.