

Best practices for implementing a BI strategy with SAS

Mike Vanderlinden, COMSYS IT Partners, Portage, MI

ABSTRACT

Implementing a Business Intelligence strategy can be a daunting and challenging task. SAS has introduced many concepts to us by bringing a comprehensive BI solution to the market. While some of those topics are not new, they can certainly make us think more than before, which is a very good thing. Many companies go ahead with the installation and configuration of the SAS BI platform with very little planning, only to ask themselves many questions afterwards, questions that should have been answered early on. This paper describes the best practices for all phases of a Business Intelligence strategy implementation using the SAS platform. It will cover what needs to be considered during the planning phase such as architecture, security and optimization as well as introduce the different tools available for data collection and reporting and what it means in terms of changes to move from SAS Foundation to the SAS BI platform.

INTRODUCTION

Your organization just created a new project team, the Business Intelligence Group, whose task is going to implement a BI solution from a vendor to improve business processes, reduce maintenance cost, give a clearer view to executives of historical data and possibly make predictions using analytics. Now as the lead of this new initiative, you are faced with many decisions to make and you are probably asking yourself many questions. One question you may even have is: "am I asking myself the right questions?"

This paper will guide you by going through the steps necessary for a successful implementation. I will skip here the vendor selection process; I assume and hope you have made the right decision to go with the SAS BI platform.

PLANNING

This phase is the most important of all. You will need to select the right bundle of tools for your organization, define an architecture for your environment that is optimal for all activities and scalable to accommodate data and business growth, plan for security including deciding on an authentication system as well as plan ahead the different users and groups to create in a way that is not going to be a burden for the SAS administrator to maintain your environment secure. If you are moving from version 8 of the SAS software, you might also need to plan the migration process for the SAS data, programs and applications. This is addressed directly in the migration section.

Selecting the right tools for your SAS BI environment is an activity that will take place with your SAS representative or a partner who can assist you to make the decision based on your business needs as well as your internal competencies.

Every action taken or decision made during the planning phase must be documented. The idea is to gather enough information to be able to develop a comprehensive implementation plan that can be followed by the SAS administrator to make the installation and the configuration of the software as smooth and efficient as possible.

ARCHITECTURE

Once you have selected a SAS BI package you must then define the architecture of your environment. For this purpose the SAS Institute uses a sizing questionnaire, a list of questions related to your hardware, data size and anticipated usage of your new system. The answers allow making recommendations for the servers you will need and how many you will need, along with memory, number of processors and disk space requirements.

When you are considering the purchase of a SAS BI package and would like to get a firm quote in order to make a sound decision, the SAS Institute or one of its resellers first need to work with you to help you define the architecture of your environment, more specifically the deployment architecture which will describe which tools will be licensed and on which servers they will be deployed.

Let's assume that you have purchased the EBI package and the recommendation is to have a three server architecture*, one for the metadata server, one for the compute server which will host the workspace server and the stored process server and one for the web server which will have the web clients on it.

Even if you decide to allocate three powerful servers for your SAS environment, it is important to consider other resource-consuming components such as databases, web servers and any other applications that might be hosted by one of the servers. If for example one of the servers will also be hosting a large database server, you will need to consider increasing the amount of memory available as well as disk space.

It is always a good idea to isolate the metadata server from the rest of the environment. The metadata server keeps the metadata in memory before writing them to disk so it's important to allocate a good amount of memory just for this server. It is also recommended, for performance, to assign a fast drive such as a RAID 5 disk.

Dedicating a server to the web tier will obviously increase performance as well. Regardless of which application server you select, they can all be greedy with memory at times. Additionally, isolating the web tier will also make it more scalable, it will allow you for instance to deploy your own applications on the same server without having to reconfigure the environment or move servers around assuming the server has enough memory to accommodate those new applications.

It is fairly easy to add power to your compute server. For instance if you decided to invest in the Enterprise Data Integration server and start developing time consuming ETL processes, you would just install the necessary software to host a workspace server and define the appropriate metadata for SAS to know about it. While you can add workspace servers on the same machine, load balancing will be useless since it will only have one machine to go to. On the other end, Logical stored process servers can co-exist on the same machine. In the SAS management Console, you can have two logical workspace servers, on different machines, each with its own parameters to make it as efficient as possible (as discussed in the Optimization section), especially if the ETL processes are going to run at the same time the information consumers will be viewing reports it is then highly recommended, if not necessary.

* I purposely left out the details for each server such as operating system, memory, ... This information is very specific to each organization.

SECURITY

Securing your environment requires a lot of planning; typically we encourage organizations to develop a security plan before starting the implementation. The plan will describe your needs for authentication and authorization and how it will be implemented in your SAS environment. The SAS administrator should take an active role in the development of this document because it will also describe the different groups of users and their assigned sets of permissions. It will be his task to keep this document up-to-date as business grows and the number of users of the system increases.

First you need to decide how you users will be authenticated in the SAS environment. You can authenticate by the operating system, an Active Directory or even do Web authentication and let your application server do the work. It is feasible to switch from one form of authentication to another, some organizations decide to start with OS authentication while in testing mode, then switch and leverage their corporate active directory just before going into production.

Then you will need to work on authorization and categorize your user base into different groups and give each group a set of attributes which will determine the set of permissions this group will have. The Authorization Manager plug-in of the SAS Management Console will allow you to give permissions to groups and users. It also has a repository Access Control Template to define default permissions for different groups. Here are a few important things to remember: 1) every user is implicitly a member of the PUBLIC group and every user who has a metadata identity defined is a member of the SASUSER group so great care should be taken when modifying permissions for those two groups in the repository ACT. 2) It is recommended to split your users into two logical groups to start with, those who will create and those who will exclusively consume information. In order for a user to be able to write metadata, such as creating a map or report or any other SAS component, they need write access to the metadata at the repository level, which means you cannot start from a highly restrictive set of permissions and give write access to a few resources for that user, instead you'll need to work the other way around and restrict that user to write metadata for resources he should not be able to modify.

At the time you plan your different groups of users, you will need to review how permissions are evaluated and make sure you understand the precedence rules. They are very well documented in the online help for the SAS Management Console.

Before giving access to data sources, the credentials that are going to be passed to the database must be decided. For maintenance reasons, it is best to allow groups of users to access specific tables with one set of credentials. The User Manager plug-in allows us to create Metadata Identities to which several logins can be assigned. If for instance I want to give access to a group of users to an Oracle database, I will create a group in the metadata, assign that group the necessary login to log on to the database and specify the same authentication domain that was assigned to the Oracle server also defined in the metadata. I will then add the different users and groups as members of this group. When trying to connect to the Oracle database, SAS will try to retrieve a login based on the authentication domain and when it finds one, will use it for the connection.

OPTIMIZATION

Once the software is installed and configured, the role of the SAS administrator is not only to manage permissions to keep the environment secure but also to make it efficient and monitor server activity to ensure it remains optimal. Typically your SAS environment will include the following servers: Metadata server, workspace server, stored process server, OLAP server and an application server. We will assume Tomcat as the application server for this paper.

The metadata server is an easy one, just as mentioned in the Architecture section, it should be on a dedicated machine with the appropriate amount of memory, disk space and if possible a fast disk. For the other SAS servers, you have more options to control them. The workspace server is set as “Standard” by default but can be changed to either “Load Balancing” or “Pooling”.

Load Balancing for your workspace server is not useful if you only have one compute server, the idea is to distribute the work across several machines. On the other end, for Stored Process servers, it is possible to load balance on a single machine, which is the default after you have run the Configuration Wizard as part of configuring your environment. The Object Spawner will take care of sending the work to the least busy server based on information collected from all the object spawners in your cluster of servers. Load Balancing is usually best for ETL processes and other time-consuming jobs that can run in batch mode.

Pooling will allow reusing existing connections to workspace servers, which will save time by not initiating a new connection for each user request. The servers in a pool are also organized by “Puddles”, for which a specific login is assigned in order to authenticate. When a user request a connection, the pool will decide which puddle he can access based on his user id. He must have access to that puddle or be a member of a group that can access it. Pooling is best if you will have many information consumers accessing reports that do not require long jobs to run. Pooling requires you to set up security and should be part of the security planning process early on if possible.

Additionally, you can edit the command line that starts an instance of the workspace server to specify options that will increase performance specifically for Web Report Studio:

- -RSASUSER will set the SASUSER library in read-only mode to make the workspace server faster.
- -work folder_name where folder_name is the directory for the work data library, just like for the metadata server, a fast disk such as RAID 5 is recommended to take advantage of this option.
- -ubufsize, -memsize, -realmemsize, -sortsize and -cpucount allow you to specify settings for memory and processors, the values for these options will be very site specific but should not be overlooked as they can greatly increase performance.

The OLAP server can be tuned for performance also. The simple improvement to make is to modify the temporary file path to point to a fast disk. There are numerous options for the OLAP server in order to optimize it, too many to go over each one in this paper but the important thing is that The Performance Logging option will allow you to generate a file with information on usage of the server. From this information you can then go through each option and set an appropriate value.

The Tomcat server starts with several options to control the amount of memory available, ... SAS provides a set of values for a given architecture and a predefined numbers of users, which is a good starting point. The SAS administrator will need to monitor the performance of the application server and modify those options as needed. Note that the values to be set for those options are not trivial and sometimes requires a trial-and-error approach.

CLIENTS

Once your environment is installed and you have been through the basic configuration steps, you need to do some configuration for the SAS clients. We will consider here Information Map Studio, Web Report Studio and Stored Processes.

Information Map Studio

Information Maps by default will be stored in the Maps folder in the BIP tree; this is where Web Report Studio will look for them. It is recommended to create a folder hierarchy that can relate to the different groups you have created, not every user will be able to use all maps. This needs to be thought out before creating your reports because if a map is moved to a different folder, the reports that rely on it will not be functional anymore. Groups of users who need to create reports based on a map will need the Read and ReadMetadata permissions for the map, while users who need to be able to create or modify a map will also need the WriteMetadata permission, all other permissions are not effective on maps.

The decision as to how many maps should be created is a matter of design and maintenance. You could create one map that represent a whole database, queries for that map will be optimized based on the variables selected for a report, but the goal to create a map is to make it simple for business users to create their own reports and thus a compromise must be reached. If you create just one map, then you will only need to maintain that map but everyone will have access to the same data and the map might be confusing for users. On the other end, creating too many maps will require more maintenance for both the SAS administrator and the users creating the maps. In most cases, it is best to create a map that can serve 3 or 4 reports to start with.

Web Report Studio & Stored Processes

Just like Information Maps Studio, you need to create a folder structure to store reports, a structure that can easily be managed by the SAS Administrator given the different groups of users.

Users who can create reports will need Read and ReadMetadata on the appropriate maps, but will also need WriteMetadata in the repository and in specific folders where they should store their reports. A good approach I have recommended to several organizations is to let users who can create reports store them in a staging area where they wait for formal validation. Once a report is ready to be consumed it is then moved by the administrator to the appropriate folder and given the right set of permissions. This gives the administrator more control to keep the environment clean and secure.

Keep in mind that in order to see a report, users will not only need access to the report but also to the underlying objects associated with the report including the map and the data sources.

All this applies if a report or part of it is based on a stored process, users need read access to it. Web Report Studio will look for stored processes in a specific location that should be controlled by the administrator, users should store the stored processes in a staged area as well.

MIGRATION

From SAS 8 or earlier to SAS 9

Moving from SAS 8 or earlier to 9 has more implications than if you are already on version 9. You will need to go through a migration process that involves upgrading most of the SAS components.

The first step will be to take an inventory of all the files that need to be migrated, including SAS data, programs, applications, ... SAS did a great job at documenting these in the Migration Community at <http://support.sas.com/rnd/migration/>. You can use PROC MIGRATE to upgrade the SAS files. Note that if you decide not to upgrade your SAS data, you will take advantage of the 64 bit processing nor will you be able to write to those data sets. You will also need to consider third-party tools and evaluate which ones need to be upgraded to support SAS BI.

From SAS Foundation to SAS BI

If you already are on version 9, you can simply install the BI software on top of SAS Foundation. I want to mention a couple of things you can do to make existing reports available from the BI tools very rapidly:

Convert a SAS program to a stored process:

SAS programs can easily be converted to Stored Processes by simply enclosing the output part of your programs with the macros %STPBEGIN and %STPEND. These macros will take care of the ODS statement for you while you still have control using ODS options.

Parameters will be defined in the metadata and declared as global in the SAS program before the *ProcessBody statement (this statement is only necessary if the stored process will run on the workspace server).

Note that macros and SCL entries cannot be run directly by the stored process server.

For example the following macro:

```
%macro test(parm1, parm2);  
[Data step]  
[Ods statement]  
[Proc report]  
%mend;
```

will turn into:

```
%global parm1 parm2;  
[Data step]  
%STPBEGIN;  
[Proc report]  
%STPEND;
```

Write directly to the repository:

Another way to reuse a report from a SAS program to be available in the repository is to modify the ODS statement and write to the repository directly.

For example the following program will take the output of the PROC PRINT and create a report in the repository name testReport.

```
filename tgt sasxprp "SBIP://Repository/Bip Tree/ReportStudio/Users/userName/Reports"  
userid="userName" password="userPass" domain="userDomain"  
OMRHost="metataserver" OMRPort="portNumber" OMRUser="connectUser"  
OMRPassword="connectPass" OMRReposName="Repository"  
;  
option noovp;  
ods sasreport file="testReport.xml" path=tgt;  
proc print data=sashelp.shoes;  
run;  
ods sasreport close;
```

CONCLUSION

The most important aspect of a BI implementation is planning. This paper is meant to give you a basic approach for an implementation, allow you to ask the right questions from the start and provide you with a list of considerations that need to be reviewed in detail prior to install the software if possible. There are many other components to the SAS BI suite but the topics covered here are enough to give you an environment that is going to be operational from the start and yet be easily scalable. I have seen many companies jumping into BI without much planning, accepting the fact that “they will learn as they go” too easily. This might work for a small organization that doesn’t anticipate too much growth in the near future but it is definitively best to invest more time in planning and documenting your environment early on. There will inevitably be a learning curve for both the SAS administrators and SAS users and while experience working with the tools is necessary, getting a clear picture of where you want your environment to be and how you will ensure its growth is invaluable.

RECOMMENDED READINGS

SAS Intelligence Platform Administration Guide
SAS Intelligence Platform Installation Guide
SAS Intelligence Platform Security Administration Guide
Online Communities at <http://support.sas.com/rnd/intro.html>

CONTACT INFORMATION

Mike R. Vanderlinden
Senior Technical Manager
National SAS Practice
COMSYS IT Partners
5278 Lovers Lane
Portage, MI 49002
(269) 7605716
mvanderlinden@comsys.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 registered trademarks or trademarks of their respective companies.