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. We will skip here the vendor selection process; we assume and hope you have made the right decision to go with the SAS BI platform!

PLANNING
The implementation of a BI solution can be broken down into three main phases: the planning activities, the actual installation and configuration of the software, and the administration of the software including monitoring usage and training users. The planning phase is the most important of all for a successful implementation. 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. There are also planning steps you can take for the installation, configuration, administration and optimization of the software. Other planning activities such as planning for user training, documentation … are the responsibilities of the BI Program Manager and are out of the scope of this paper.

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. Let’s discuss each of those items in more details.

SOFTWARE SELECTION
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 decisions based on your business needs as well as your internal competencies and corporate standards.

The best approach for this process is to get a good understanding of the SAS BI tools and their capabilities. SAS has done a great job at developing different interfaces for specific audiences. Based on your business requirements, you should identify the different audiences of the BI environment and select which tools they will use based on their needs and/or skills.

This is also the best time to research the SAS Open Metadata Architecture and the SAS Foundation Services. If your organization has specific needs that can’t be served by one of the SAS BI tools, SAS provides an architecture that developers can build on to create applications that are fully integrated in your BI environment.
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. Capacity planning considerations are addressed in the next chapter. Here we focus on the deployment architecture.

The diagram below shows an example of architecture for a SAS BI environment:

In this example, business users will access reports using a Web interface. They will first hit a Web Server where a security realm has been implemented to authenticate users against an LDAP directory. The Web Server also performs load balancing between a couple of clustered application servers where the SAS Web applications reside. The SAS Metadata Server has been isolated on one machine and the other IOM servers have been split between two other machines.

The RDBMS component represents the back-end storage for the content server.

As you can see, important decisions have to be made from the start to provide a robust and scalable platform. Let’s examine in more details the key factors that drove this architecture and the advantages it provides.

Placing an HTTP server in front of the cluster of application servers allows load balancing between the servers to provide high availability, scalability and failover capabilities *. At the same time, the HTTP server can be placed in a DMZ and thus isolate the application servers from a wide range of clients including Internet users, which offers higher security.

The HTTP server can also be used to serve the static content of the SAS Web applications, HTTP servers are typically better suited for this type of activity and it allows the application servers to be dedicated to serving the dynamic content.
In the sample diagram, a security realm has been implemented to authenticate users against an LDAP directory. This allows enabling Web authentication for the SAS Web applications to leverage an existing directory service and make the users and groups administration more manageable for large deployments. This is discussed in more details in the ‘Security’ section.

* Note that in this particular example there is only one HTTP server which presents a point of failure. One can imagine several HTTP servers with a load balancing software in front of them to provide the same failover capability as the cluster of application servers.

As mentioned before, the application servers are in a cluster to provide high availability and failover capabilities. The RDBMS underneath on the diagram represents the back-end storage for the content server. The preferred choice for this server is Xythos which can easily be deployed in the application server as well. The important factor to keep in mind is that the back-end storage for this server is a database and thus can be tuned, backed up ... just like a database serving source data. If performance is poor when retrieving reports, the report data can be stored on the file system, the report metadata is still stored in the database.

The SAS Metadata server has been isolated on a server machine. There is a lot of communications between the different servers and the Metadata server. It is necessary to provide adequate resources for this server, as well as isolating it for medium-to-large deployments.

The Metadata repository is by default a SAS database. Other database systems can be used such as Oracle and DB2 but will not offer the same performance as a SAS database.

The other SAS servers, placed in the larger blue boxes, represent the compute tier. This is where all computational activities will take place. In this example, two Workspace servers have been deployed on two separate server machines to provide load balancing, which provides better performance for running heavy ETL processes. The same servers could be configured for pooling, this would make short-term request faster, such as running reports. Note that load balancing Workspace servers on a single server machine does not increase performance.

Logical Stored Process servers on the other end can be load balanced on one server machine, which is the default configuration when installing the SAS BI software. The load balancing can be configured with many parameters for different algorithms.

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.

**CAPACITY PLANNING**

Planning the hardware resources necessary for a SAS BI environment is key in providing users an optimized and scalable environment. As mentioned before, SAS can make recommendations based on the architecture, a sizing questionnaire and your corporate standards. It is very important to know how much growth is expected on the business side. The type of operating systems and processor influences the sizing as well. While this is not an exact science, here is the approach we use to determine the appropriate hardware resources for a new environment.

The SAS Metadata server is an 'in-memory' server and thus it is best to allocate enough memory to prevent disk swapping. The formula used to estimate the amount of memory is to multiply the maximum number of concurrent users by 32 MB. For instance if the anticipated maximum number of users is 600, at least 19.2 GB of memory should be dedicated to the Metadata server (600 * 32MB).

Determining the number of processors is more complicated, it depends highly on the type of processors. Our approach for this was to find case studies with similar environment; SAS also has published test suites for different operating systems that are very useful.

The approach was similar for the servers that make up the compute tier. It is important to know how many concurrent users you anticipate your system will have, as well as determining what percentage of those users will access dynamic content. Users who retrieve static reports do not put any load on the compute tier and are thus considered light users.

For most Windows-based system, the rule of thumb is to dedicate 1 processor per 8 concurrent jobs. For other platforms such as the different flavors of UNIX, the processors can typically handle more users.
To determine the amount of disk space for the ‘temp’ and ‘cache’ areas, we have two formulas at our disposal. The ‘temp’ area can be estimated by multiplying the number of concurrent jobs by the average size of a query and then by a factor of 3 to 5. For instance one of the environments we worked on estimated the number of concurrent jobs at 200 and was able to determine the average size of a query based on a previous application, at 1.5 GB. We used the formula:

\[200 \times 1.5 \text{ GB} \times 5 = 1.5 \text{ TB},\]

this number ended up being appropriate for the SAS usage.

The cache area used by the Workspace server uses a similar formula, with the exception of the average query size; we replace this by the average query output size.

**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 web 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.

The diagram below depicts what happens during authentication based on the example considered in the architecture section (we have taken out the HTTP server in this case):

![Authentication Diagram](attachment:image.png)

Authenticating against an LDAP directory allows leveraging an existing corporate directory of users and groups and providing a more manageable administration of those entities in large deployments. In our example, a custom security realm for web authentication allows doing more custom authentication, for example external users of the company might need to be authenticated against another LDAP directory or an RDBMS.

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) Denying access to the PUBLIC group for a resource (such as deny ReadMetadata) might prevent any users from accessing the resource unless a specific grant is also set for a group of users. While we recommend starting from a restrictive security system by denying most rights to the PUBLIC group (unless appropriate), each user of the system will need to be a member of a group for which specific grants have been set in order for those users to access resources.

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. It is best to assign permissions at the highest level of the hierarchy.

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.

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.

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/](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: will turn into:

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

**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.

```sas
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;
```
REPORTING
Deciding on how information consumers will tap into the power of the SAS BI platform also requires significant planning. Earlier we addressed architecture, security, capacity planning, optimization and migration. These are important factors to consider when planning your environment from a reporting standpoint but there are also some additional factors to consider.

User Profiles
Understanding the types of users that will be using your platform is important in order to realize a successful roll out. Planning this ahead of time will enable you to implement effective & timely hardware decisions, performance and security configurations and training. This will be very site specific but some common user groups that are found in BI Environments are:

<table>
<thead>
<tr>
<th>SAS BI User Type**</th>
<th>SAS BI Role</th>
<th>SAS BI Client Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrator</td>
<td>User, data, and report security</td>
<td>Management Console</td>
</tr>
<tr>
<td>Designer</td>
<td>Data Modeler, Librarians*</td>
<td>Data Integration Studio</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Information Map Studio</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OLAP Cube Studio</td>
</tr>
<tr>
<td></td>
<td>Stored Process Developer</td>
<td>Enterprise Guide</td>
</tr>
<tr>
<td></td>
<td>Portal Developer</td>
<td>Information Delivery Portal</td>
</tr>
<tr>
<td>Consumer</td>
<td>Business Analyst/Executive</td>
<td>Information Delivery Portal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Web Report Studio</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Add-in for Microsoft Office</td>
</tr>
</tbody>
</table>

* Librarian roles are important in monitoring data and reports to avoid redundant data marts and/or reports being created unnecessarily.

** One person within an organization may be responsible for more than one of the above roles depending on your organization.

SAS BI User Types
Training
Training should be implemented just before or during your software installation and configuration to avoid your BI platform sitting dormant. Customized training at your site is often effective. This training can utilize your data and reporting environment therefore, making the training more relevant.

Understanding how your users prefer to work can really help with planning and roll out. If your users are Excel power users, you will want to focus on the SAS Add-in for Microsoft Office. If your users only understand the business aspects of the data and are more comfortable using a web browser, Web Report Studio and the Information Delivery Portal may be the way to go. Most likely your environment will end up with a blend of tool usage.

Report Graphics
Understanding the types of graphics required for end reports is important for planning the installation and configuration. For example, if reports will need to utilize geographical maps, the SAS/Graph map data sets will need to be installed along with SAS/Graph. Many times, this does not happen by default.

Report Data Architecture
Another important exercise is to define the types of reports needed in your BI environment. This will help immensely in data modeling efforts. Having a plan for reading your transactional data and storing it in a structure that is effective for reporting is very important for performance and the end users having an intuitive understanding of what the data contains.

You will also need to consider the source of your report data to ensure that you purchase the correct SAS/Access products for your needs.

Report Validation
One activity that really generates excitement and gains momentum is to begin matching numbers on legacy reports from the SAS BI reporting tools. It is valuable to find a good representative collection of production reports and emulate them from SAS BI platform reporting tools such as Web Report Studio.

ANALYTICS
Understanding the analytics needs of your information consumers is important in purchasing the right modules along with the SAS BI Platform. A first step is defining what analytics means to your company. If what you are looking for from your business intelligence environment are counts, percentages, sums, averages, and measures of variation, most likely you can get by with the standard SAS BI Platform package. OLAP Cube Studio may also be implemented for drill down type reports. If you are looking for more advanced analytics to be included in your reports, you may also want to consider the following SAS components for your BI environment. This will also play a role in hardware sizing. For example, if you know you will have 10 users running stored processes involving logistic regression, you may want to have a separate compute server just for that purpose.

SAS/Stat
SAS/Stat is the main statistical package in SAS and contains procedures to carry out analyses that include but are not limited to:

- Regression Analysis
- Analysis of Variance
- Categorical Data Analysis
- Survey Data Analysis
- Survival Data Analysis
- Cluster Analysis

Forecasting (SAS/ETS, SAS/HPF & SAS Forecast Server)
SAS contains analytical tools to allow users to create models to forecast events based on time stamped data.

The procedures from SAS/ETS can facilitate analysis which includes but is not limited to:

BI Forum
SAS Global Forum 2007
econometric analysis
- time series analysis
- time series forecasting

The SAS High Performance Forecasting component includes procedures which facilitate choosing and applying the best model to many individual forecast endpoints and display the results back in a hierarchical manner. For example, if your company wants to produce forecasts for 10,000 products across many stores and regions. SAS High Performance Forecasting would go through each product one by one and figure out the best underlying model to use then produce the forecast. It would then reconcile the forecast across the hierarchies (region, store, product type).

SAS Forecast Server is a GUI front end and analytics platform to facilitate building these models across hierarchical data.

**SAS Enterprise Miner**
SAS Enterprise Miner streamlines the data mining process to create highly accurate predictive and descriptive models based on analysis of vast amounts of data from across an enterprise. Taking advantage of the underlying technology benefits of SAS 9, SAS Enterprise Miner identifies trends, opportunities and threats more efficiently and with less risk than ever before and it's integrated, collaborative environment helps statistical modeling, IT and business professionals work together to effectively make key strategic business decisions.

**SAS Text Miner**
SAS Text Miner works with SAS Enterprise Miner to provide a rich suite of tools that have been specifically developed for analyzing and extracting intelligence from multiple text documents.

**SAS/OR**
SAS/OR provides a full set of tools that give companies the knowledge to identify and optimize business processes and management challenges. SAS Operations Research and Management Science delivers answers that help companies effectively make strategic business decisions on a variety of issues ranging from resource allocation and inventory planning to distribution, scheduling and routing.

**SAS/QC**
SAS/QC software delivers a wide range of specialized tools that help organizations identify and understand the sources of process variation, enabling improved product quality, optimized processes and increased customer satisfaction. SAS/QC uses data from across an organization to enable users to discover root causes of problems, and goes beyond basic process control to provide more complex statistical analyses, enabling the creation of more efficient, cost-effective business processes.

**Reporting Client Configurations**
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 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.

**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 installing 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  
Office: (800) 831-6314  
Cell: (269) 760-5716  
e-mail: mvanderlinden@comsys.com

Brian Varney  
Senior Technical Manager  
National SAS Practice  
COMSYS IT Partners  
5278 Lovers Lane  
Portage, MI 49002  
Office: (800) 831-6314  
Cell: (269) 207-7593  
e-mail: bvarney@comsys.com