Paper 023-2012

The Emergence of Patterns in SAS® Usage and Infrastructure

Chuck Kincaid and Mike Vanderlinden, Experis, Kalamazoo, MI

ABSTRACT

It used to be that there was one particular look and feel to a SAS® environment. Then platforms changed, editors changed, but for a long time the infrastructure remained the same. With the introduction of a comprehensive Business Intelligence Platform, the expansion of SAS across the enterprise, and its integration with IT departments, new usage patterns have emerged. More infrastructure choices are available to satisfy the diverse roles required by companies who aim for analytics maturity. This paper will present our analysis of the more prevalent patterns to date, and the pros and cons of each. The interested audience could include IT and business managers who are interested in the possible ways to deploy SAS and what fits best for their environment.

INTRODUCTION

The reader may be familiar with *patterns* as tools for creating things used in many areas – architecture, software, and of course, knitting. The reader may also be familiar with *types* as tools for classifying things – software, sociology, anthropology. In some instances, such as urban planning and architecture, it seems like both are used to describe the same thing.

The purposes for both concepts, patterns and types, are similar and helpful. For example, German sociologist Max Weber defines ideal types are "idea-constructs that help put the chaos of social reality in order."At the same time, Gamma, et al. say that software design patterns "capture these solutions [to specific problems] in a succinct and easily applied form. Both statements bring out important aspects of what this paper is about.

In our business, we see many ways that SAS, and more broadly, analytics has been implemented within organizations. By accentuating those characteristics that are "readily observable from a distance," we have identified five patterns or types that may help to put some order to the chaos, help other people to understand and implement the "best practices" for analytic infrastructure, and help us as a community to communicate better by giving us a common language.

WHAT'S SUP?

Point of view

SAS Usage Patterns are patterns (or could have been types, but SUT wasn't as funny) that describe the way that companies use SAS. Your point of view is important when determining what is a pattern and what's not. A user (programmer, statistician, developer) is going to see it differently than someone from IT or someone from the business. For this paper, we tried to look at SAS from the business' point of view. IT already has their own patterns for architecture, so we did not take their point of view. We also thought that the user both has a narrow view of the infrastructure (not always, but a lot of the time), and is familiar enough with SAS that they would be able to adapt the business' point of view to their own.

Benefits

The other consideration regarded the benefits of the paper. Being able to speak at MWSUG is a good enough reason to write the paper, but we also would be pleased if others received value from it. As Analytics, and certainly SAS, makes greater and greater inroads into companies, more and more people will be using SAS who have never seen its power. It is hoped that this paper will help the audience learn about other ways of using SAS so that they can increase the return on their investment.

Conformity

From our experience, there are five patterns to the way that organizations use SAS. This does not imply that everyone uses SAS the same. They do not. Nor does it imply that these are the only ways to use SAS. It is not. Instead, we have grouped SAS installations into similar styles and brought out the key characteristics – almost like statistical clustering, though certainly nothing nearly as formal – that we think are important from a business manager's point of view. As we discuss below a future interest is to objectively understand the prevalence and variability of the patterns in the real world.

Maturity

The patterns that we describe below can be building blocks for an analytics infrastructure. They are not meant as a maturity model. Each pattern itself has different levels of maturity, and we've begun that definition in this paper. Com-

Paper #023-2012, continued

bining patterns does indicate an increase of maturity in an organization's Analytics Infrastructure. We discuss this later.

Another aspect of maturity to note, is that often all companies have most, if not all, of these patterns, but they implement some of them with other software or tools (Excel being very prevalent). This increases the overhead of the analytics infrastructure, requiring different groups for development, maintaining, training, etc. Another form of maturity, in our admittedly biased opinion, is when companies implement these patterns in SAS taking advantage of the unified platform. [reference]

ANALYTICS INFRASTRUCTURE

Pattern Template

A typical pattern in the world of software development (a.k.a. *design patterns*) is described by its name, the problem it addresses, its solution, and the consequences of applying the pattern. While we aim at describing patterns in this form in the future, our patterns are more observation based given our years of experience providing consulting services to the SAS community and thus fit the definition of the aforementioned *Ideal Types*. Those are usually described by the characteristics and elements of a particular situation and may not be comprehensive or perfect.

To help the understanding and discussion of patterns, we have created a template that includes the name, description, usage, infrastructure, participants, collaboration, and analysis. Below is a description of each of these sections of the template.

Item	Description
Name	Convey the pattern's general nature. Over time, as patterns are adopted by more organizations, the names of the patterns create a common vocabulary to reference them.
Description	The pattern is described in terms of general use and applicability.
Usage	The typical business usage or situation is depicted in this section with the intent to evolve to a clear problem statement when the pattern matures.
Infrastructure	The different architectures are discussed along with the tools licensed and software products utilized.
Participants	The functional roles encountered in the situation the pattern describes are listed along with the typical required skill set associated with them. In some cases, one person may take on multiple roles.
Collaboration	The interactions between the active participants of the pattern and the required support entities are discussed, and their respective responsibilities in order to optimize the usage of the pattern.
Analysis	The pros and cons of the situation described by the pattern are weighed and typical transitions from this pattern to another or the addition of a new one are described.

The patterns we depict in this paper are situational by nature. They describe a specific usage of a SAS configuration and should not be confused with traditional design patterns encountered in software development. As more information and feedback is collected, they will evolve and so will the template used to describe them.

SAS USAGE PATTERNS

Pattern #1: Advanced Analytics, aka The Statisticians' Lair

Description

By definition, a lair is a secluded or hidden place, much like a secret retreat in the context of statisticians. They are never more at ease than when given an isolated environment where they can independently perform the tasks they do best: statistical analyses. In this configuration, their work is similar to a black box, they are given data, and spill out
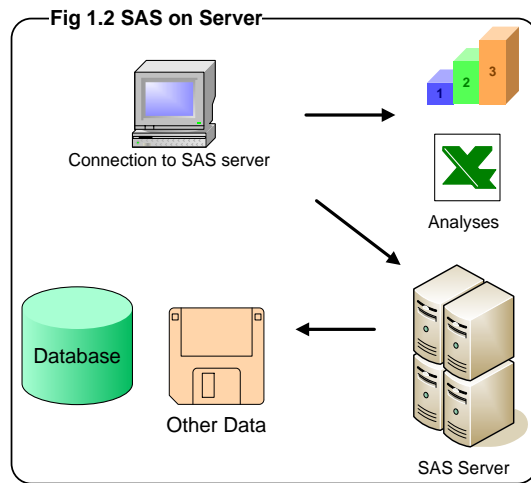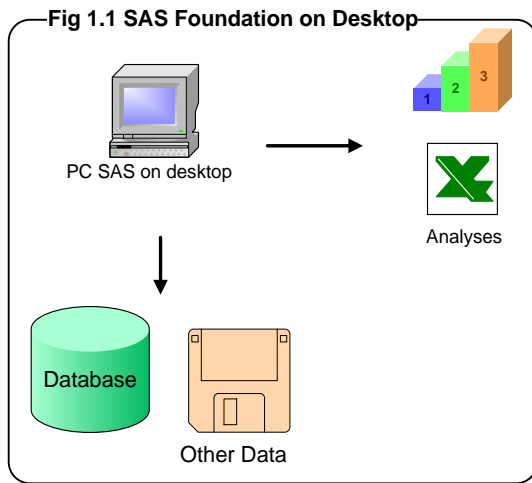
Paper #023-2012, continued

results but no one dares asking how they got there. It is very much an ad hoc environment to answer one-off questions from their management or business partners.

Usage

In the SAS world, statisticians have a myriad of tools at their disposal to perform all sort of analytical activities. They may use the SAS Foundation products such as STAT, ETS, OR and QC that allows them to work from SAS on their PC or SAS Enterprise Guide and build statistical models on their own. Or for more specialized tasks they may use an interface provided by SAS such as Enterprise Miner for data mining, Forecast Studio for forecasting, JMP for data visualization and analytics, or Text Miner for mining information out of free form text. The end product is typically the results of the analysis. For Research & Development, it may include content for publication. For the business, it could be a presentation, spreadsheets or an analysis write-up.

Infrastructure

The infrastructure is typically one of the following: (1) SAS Foundation (or Enterprise Guide) on desktop; the software is solely on their desktop/laptop. New releases are installed through SMS packages prepared by the IT group. (2) SAS on a server (Windows, Unix, Mainframe) and the user connects to it through a remote login or with SAS/Connect, running SAS programs via remote submit commands to perform their analyses.

Fig 1.1 SAS Foundation on Desktop

PC SAS on desktop

Analyses

Database

Other Data

Fig 1.2 SAS on Server

Connection to SAS server

Analyses

Database

Other Data

SAS Server

Participants

The skills required to work in this configuration include a good knowledge of the SAS language and several years of programming experience. They typically have an advanced degree in Statistics and keep up to date with the latest features of the SAS/STAT, SAS/QC, SAS/ETS and/or SAS/OR modules. In some cases, a good knowledge of SAS/GRAPH is either required or useful, though the newer Statistical Graphics procedures are becoming the tool of choice. They may use other statistical packages in collaboration with the SAS software.

Role	Skill Set
Statisticians	Statistics, SAS programming,
Data Provider	Knowledge of the source data provided
SAS Administrator	SAS administration, metadata management

Collaboration

Users of this environment require little support since they can rely on SAS Tech Support or the SAS documentation to address specific issues or answer any question about the SAS language.

On the software side, basic, part-time SAS administration skills are necessary to update and upgrade the environment as needed. As long as statisticians are given plenty of storage space, they usually have everything they need.

And because statisticians are always looking for needed data, a good relationship with the data providers, whether it's a database administrator or someone producing and passing along a spreadsheet, is key to not only get access to the data, but also to get a basic understanding of how it is organized and where it came from.

Analysis

On the plus side, an ad hoc environment is ideal to perform statistical analyses. Statisticians enjoy working independently; accessing the data they need wherever they find it. From a software perspective, this is typically a cost-

Paper #023-2012, continued

efficient solution since only SAS Foundation products are licensed and, unless large amounts of data are involved, do not necessitate to be deployed on a large server.

The drawback with an environment that offers such freedom to its users is that it becomes very difficult to keep a single version of the truth. From our experience, data sets are duplicated over and over again without any auditing or history to document the changes made. In this case, a lack of data governance can create the risk that the results of a statistical analysis, that was supposed to provide insight into an anomaly on a simple report, cannot necessarily be trusted.

Also, by duplicating data sets, it is not uncommon to have space management issues. Statisticians will try to get their hands on any data they can find and validated processes to make this newly found data available to them are rarely implemented.

Model management is another area that can benefit from more rigorous processes. While the ad hoc nature of developing statistical models calls for an ad hoc environment, it is still important to have the right set of procedures to manage the lifecycle of the models.

Pattern #2: Data Management, aka. The Data Party Store

Description

Ever since people started making decisions, they have been gathering data in one form or another. Companies have lots of data that needs to be cleaned, modeled, and merged, and while Data Warehousing has been part of our daily lives for a long time, managing data in any one organization can still be quite a challenge. The analogy to a Data Party Store is quite appropriate. When walking into a party store, I can find all kind of good things to make the weekend more interesting, but if I haven't been there before, chances are I will spend a great deal of time looking for items before heading to the cashier.

When companies use SAS for this purpose, their pattern implementations range from ad hoc data preparation for one-off requests all the way to Enterprise Data Warehouses. At the "immature" end of that spectrum little management or governance of the data exists, However, at the other end, Data Warehouse experts, data modelers and ETL developers ensure that our party store soon turns into a high-end designer boutique. In order to do so they use one of the two Data Integration platforms (Data Integration (DI) Server or Enterprise Data Integration (EDI) Server) from SAS that offer tools for Data Acquisition and Integration, Job Deployment and Scheduling, OLAP Structures Building, Data Storage and Master Data Management.

Usage

The typical output of this pattern is data delivered to another party, such as a SAS data set, Excel file, OLAP cube, or tables in an Enterprise Data Warehouse. The primary activities in the realm of data management with SAS include creating ETL and data quality programs (ETL^q). At the mature end, it includes executing a comprehensive Master Data Management strategy across the organization. SAS is uniquely positioned in the Data Quality area in that its power of analytics on the reporting side can be leveraged in ETL processes to resolves complex data quality issues.
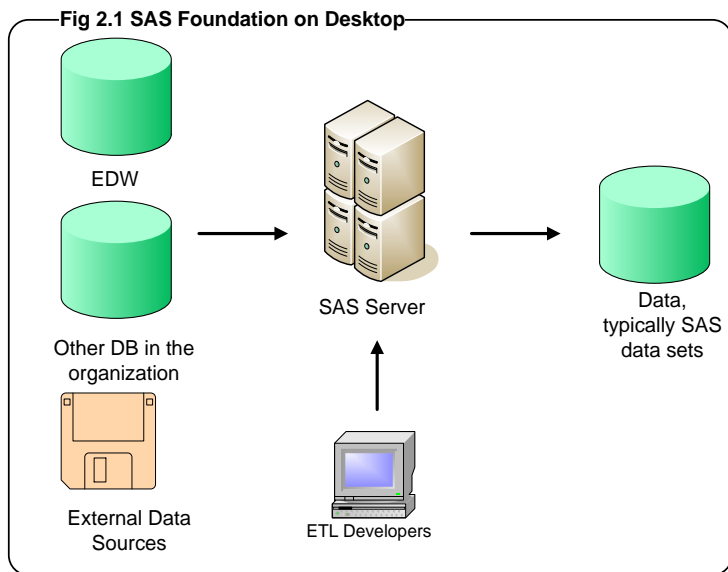
SAS programmers (in this case ETL developers) utilize Base SAS or Data Integration Studio to gather and manipulate data from source systems and create target data sets and OLAP cubes which together consist of new data marts. In the case where Base SAS is utilized, it is not uncommon to find a set of SAS programs, or better yet SAS macros that have been developed over time to promote code re-use and facilitate repetitive tasks. However, companies that wish to mature their data management program usually invest in the licensing of one of the Data Integration platforms, which give SAS developers much needed tools to manage the complexities of ever increasing data preparation activities.

With Data Integration Studio in their toolkit, SAS developers can collaborate easier in creating (and re-using) processes for extracting, transforming and loading data into target marts. When pre-defined transformations are used, and thus metadata leveraged, it allows the developer to quickly identify and estimate the impacts of changes to the overall system.

SAS now offers a comprehensive Data Management platform that goes well beyond using Data Integration Studio to create ETL jobs and OLAP structures. It includes tools for managing data quality issues and performing Master Data Management. When the DataFlux or qMDM tools are available, SAS developers focus on resolving data quality issues or integrating data into core entities to disseminate throughout the entire organization.

Infrastructure

Paper #023-2012, continued

Fig 2.1 SAS Foundation on Desktop

EDW

Other DB in the organization

External Data Sources

SAS Server

ETL Developers

Data, typically SAS data sets

The simplest infrastructure is one where the SAS Foundation products are used to create ETL jobs.

This is not uncommon in departmental units where a need arose to create their own data mart to pre-summarize data, be able to quickly integrate new data sources, or create a model that is more utilizable than the EDW by their report writers, analysts and statisticians.

This type of ETL is what usually leads to a data party store as data is not centralized and data integration processes are quickly and informally designed.

When DI or EDI server is licensed, the high level logical architecture doesn't change that much; what is different though is that SAS developers have a formal ETL tool to design and create ETL and data cleansing routines.

Whether or not the full capabilities of the tools are leveraged is a matter of skill set and training. Too often, SAS developers plug in existing code into their ETL flows; while it allows them to quickly convert SAS programs to a data Integration Studio job, it does not take advantage of all the features of the tool including the richness of the metadata one would expect to find in an ETL process.

For larger organizations, or, to be more accurate, for organizations that have heavy ETL loads, the SAS server has several Workspace servers that are load balanced to provide efficiencies. Platform Computing is also utilized not only to schedule DI jobs, but also to share the load across multiple servers.

Participants

The table below lists the roles that are encountered in this pattern. It is important to note that there is no such role as "SAS Developer" in this

Fig 2.2 SAS Data Integration Studio

Data Sources

External Data Sources

SAS EDI Server

ETL Developers

EDW or Data Marts

list. The reason is that designing and creating ETL processes usually requires a different mindset than one of a SAS programmer. The ETL developer must be familiar with ETL development techniques and have some understanding of data modeling. For instance, if the Kimball methodology is followed, a good understanding of its key concepts and ETL and data modeling techniques is required.

Role	Skill Set
Database Administrator	Database administration
Data Architect/Modeler	Data architecture and data modeling
ETL Developer	SAS programming, SAS tools (DI Studio, DataFlux, qMDM), knowledge of data sources, ETL techniques and data modeling
Business Analyst	Business requirements and processes
SAS Administrator	SAS administration, metadata management

Collaboration

ETL developers rely heavily on the database administrator and the data architect to gain a good understanding of the data sources and of how the data is to be loaded into the warehouse or marts. Along with the business analysts who must clearly convey business requirements in order for the new data store to be utilized, they form the core team of any ETL project.
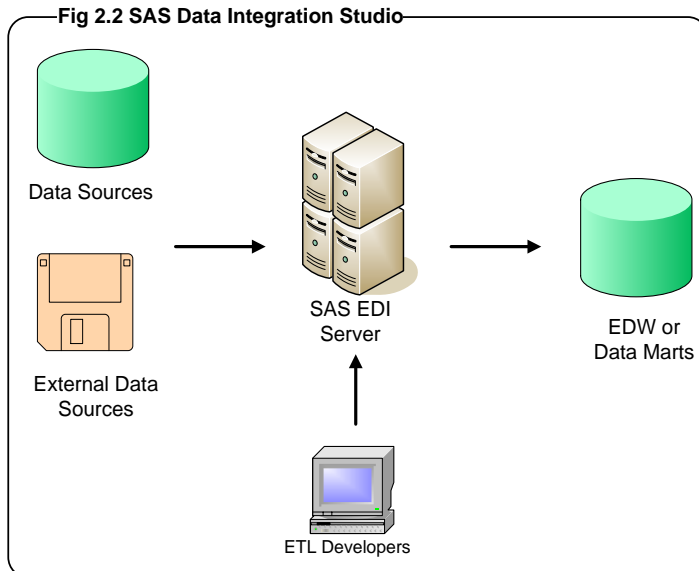
Paper #023-2012, continued

The SAS administrator ensures that the SAS Metadata server is fully leveraged and that ETL jobs are easily deployed and scheduled. While the efficiency of the ETL code is the ETL developers' responsibility, the administrator may also assist in the execution of jobs within the SAS environment.

Analysis

Our experience has been that pre-defined transformations are not used enough and, instead, SAS programmers make heavy use of the User-Written transformation to re-use existing code and simply plug it into a DI Studio job. While it provides a pseudo-documented ETL job, it is far from leveraging the full features of the tool and the metadata associated with an ETL job. Proper training on the SAS tools such as Data Integration Studio is important to avoid the situation where the tool is just a wrapper for existing ETL code.

DataFlux and qMDM, the tools for Data Quality Management and Master Data Management respectively, are slowly being adopted by organizations. When they are present they do necessitate the same software training that relevant for new Data Integration Studio users.

Pattern #3: Reporting & Analysis, aka. The Reporter's Desk

Description

In general, Report Writers have a myriad of tools to choose from when deciding on the technology to use to develop and publish reports. It is the same in the SAS world, since, regardless of which SAS products are licensed, there are many options to create reports and propagate them to business users. From a simple infrastructure using SAS Foundation where developers create reports with the Output Delivery System, to a SAS BI platform where developers can leverage an extensive publishing framework and business users enjoy self-service reporting through Web Report Studio.

Usage

SAS programmers use the different procedures available to them to produce outputs. The reports can include analytics, but they typically do not without Pattern #1 Advanced Analytics. Programmers love the ease of use of the Output Delivery System and start using tag-sets more frequently for specialized tasks. Although hardcore SAS programmers are reluctant to move to Enterprise Guide, the migration of using SAS on the desktop to a more centralized architecture using Enterprise Guide to connect to a server is on the rise.

If a Stored Process server is available to them, they are quick to convert their programs to stored processes where they can take advantage of the prompting framework and the ability to execute stored processes from different client interfaces. They create templates in SAS Web Report Studio to allow business users to create their own reports, or they create the report themselves. They are starting to leverage the SAS Publishing Framework more and more, although older delivery systems (manual and e-mail) are still prevalent in most organizations.
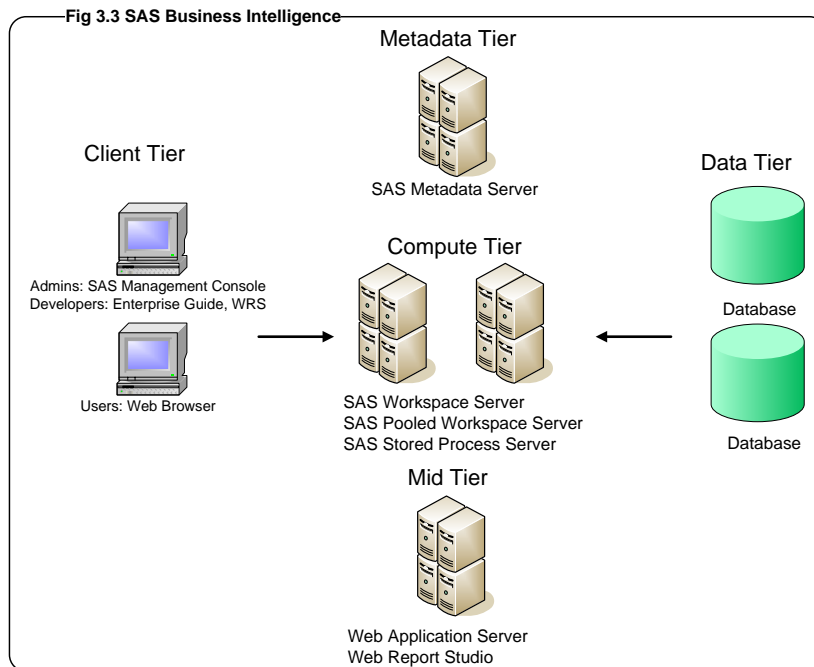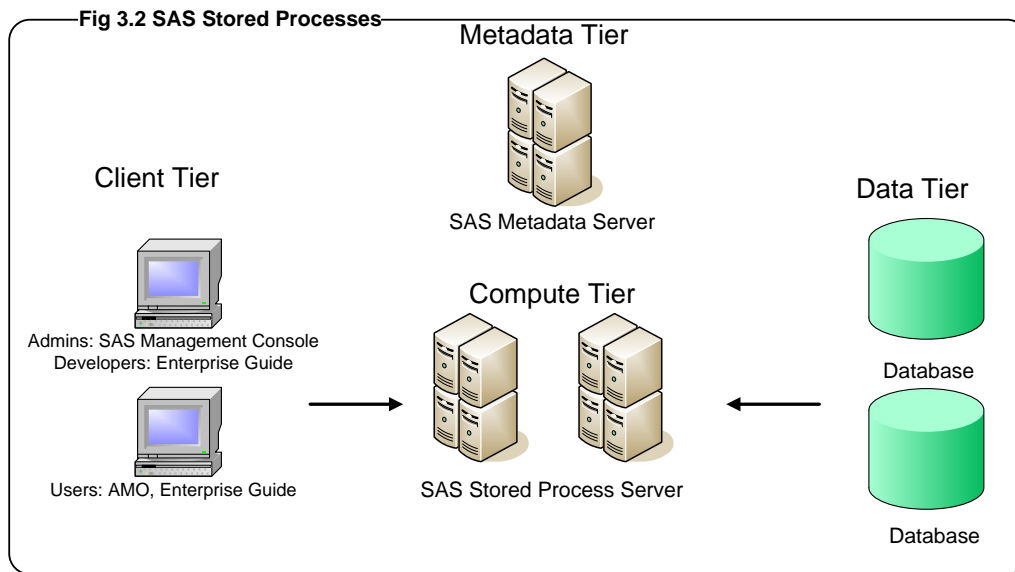
Infrastructure

Which software is licensed and how it is deployed depends on the size of the reporting activities, the maturity of the reporting system, and of course the budget for software and hardware. It is interesting to note that while the next three diagrams do not constitute a reporting maturity model, they do seem to represent a logical progression in formalizing a reporting platform and giving better access to reporting to business users.

The first architecture depicts an environment where only SAS foundation products are utilized, most likely in conjunction with SAS ODS to create reports in HTML or PDF, which are then e-mailed or moved to a central repository for consumption (which could be a local filesystem or one like Sharepoint).

Fig 3.1 SAS Foundation
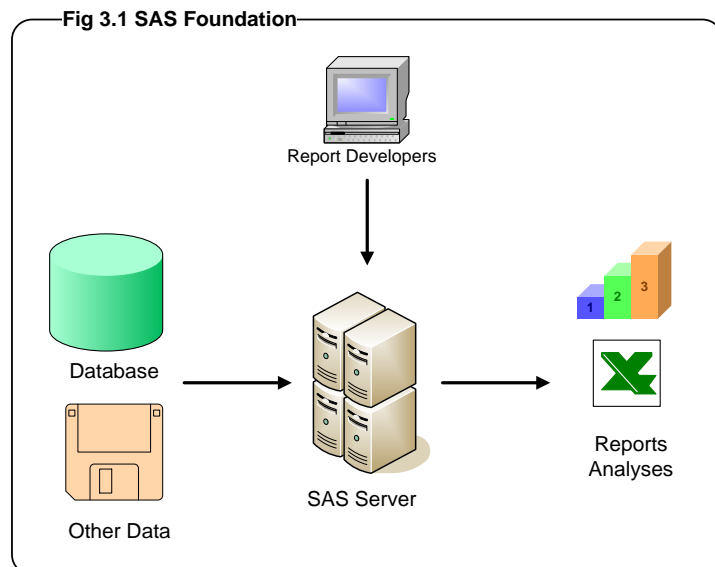
In the second scenario below, SAS developers create reports as SAS Stored Processes which are then consumed from a SAS client such as Enterprise Guide or the Add-in for MS-Office. Reports are thus centralized and more likely to represent a single version of the truth.

Paper #023-2012, continued

Fig 3.2 SAS Stored Processes

Metadata Tier

SAS Metadata Server

Client Tier

Admins: SAS Management Console
Developers: Enterprise Guide

Users: AMO, Enterprise Guide

Compute Tier

SAS Stored Process Server

Data Tier

Database

Database

Fig 3.3 SAS Business Intelligence

Metadata Tier

SAS Metadata Server

Client Tier

Admins: SAS Management Console
Developers: Enterprise Guide, WRS

Users: Web Browser

Compute Tier

SAS Workspace Server
SAS Pooled Workspace Server
SAS Stored Process Server

Mid Tier

Web Application Server
Web Report Studio

Data Tier

Database

Database

The third goes one step further by moving the report creation downstream, enabling business analysts to create their own reports with Web Report Studio. SAS developers still create stored processes but, in this pattern, their greater contribution is to create a framework for self-service reporting, with information maps and Web Report Studio templates.

Participants

The participants in this pattern includes business analysts who can run pre-defined reports or create their own from pre-built report templates, SAS report developers who create the reports and the templates, and in some configurations (such as #2 and #3 above) a SAS administrator who manages the SAS metadata.

Role	Skill Set
Report Developer	SAS programming, Output Delivery System, Stored Processes, Web Report Studio
Data Provider	Knowledge of data sources, their organization and meaning
SAS Administrator	SAS administration, metadata management
Business Analyst	Business requirements and processes

Paper #023-2012, continued

Collaboration

The report developers collaborate with the business analysts to understand the business requirements and create mockups of reports. Technical specifications or, more specifically, the technology used to create the reports is chosen by the developers to best meet the requirements.

The technical specifications also include data source mapping, which is an exercise that the developers go through with the data providers (typically someone on the ETL team) in order to find the right data sources for the reports.

The SAS administrator contributes by making the necessary metadata available to developers and putting in place the right structure for delivery.

Analysis

As mentioned earlier, the three configurations depicted in the Infrastructure section give somewhat of a natural progression of accumulating software and skills. However even with only SAS Foundation products there are some improvements that organizations can make in managing their reporting platform. For one, moving SAS programmers to Enterprise Guide has the big advantage that, if available, SAS Metadata can be utilized to manage the environment and promote more code and process re-use. It also has the benefit of being a teaching tool for newcomers or even for seasoned SAS programmers who need to use some aspects of the SAS language sporadically.

Other benefits of Enterprise Guide are faster report development, and its integration with Web Report Studio for one-button publishing.

Most organizations have realized the power and flexibility of using Stored Processes, centralized SAS programs that can be run from many other applications, especially the suite of MS-Office products. However the implementation of a true self-service reporting system, which entails the careful design of Information Maps and report templates, seems to still be slow in adoption.

Pattern #4: Integration Technologies, aka. The SPI (SAS Programming Interface)

Description

Whether caused by budget constraint or high confidence in building custom applications that serve their needs, some organizations decide to opt against the licensing of one the BI platforms. After all, SAS Integration Technologies exposes a rich set of Application Programming Interfaces to interact with the different SAS servers, and it is this very same set of APIs that was leveraged by SAS R&D to build the SAS Web Applications such as SAS Web Report Studio or the SAS Portal. This pattern deals with custom interfaces to SAS servers in lieu of licensing of the BI or EBI platforms. Note that some organizations have built quite complex applications using SAS/IntrNet, and they fit this pattern as well since they are building an interface to the SAS engine.

Usage

In this scenario, custom (web) applications are built, typically in the Java language but not limited to it, that interface with the SAS IOM servers to leverage built-in features such as security, session management, and load balancing. The most common application is a Web portal for business users to run pre-defined reports, however some organizations have used the same technologies to build applications to meet their very specific needs.

In other cases, interfaces to the SAS platform have been developed using SAS/IntrNet or the SAS Stored Process Web Applications. They are almost identical in functionalities and in fact, the latter is the most recent version of two very similar products.

One interesting application of this pattern we would like to point out, was the development by one of the authors (Vanderlinden) of APIs that supplement the SAS Java APIs and provide a bridge to the SAS Metadata and all objects defined by it through simple HTTP requests that follow the REST architectural style. This allows a zero-footprint connection from other portals into the SAS world.

Infrastructure

Paper #023-2012, continued

The infrastructure is not much different than the last scenario of the Reporting & Analysis pattern. Instead of using the BI clients and Web applications supplied by the BI/EBI platform, they use custom applications for the desktop or the Web which have been developed internally.

Fig 4.1 Custom Applications
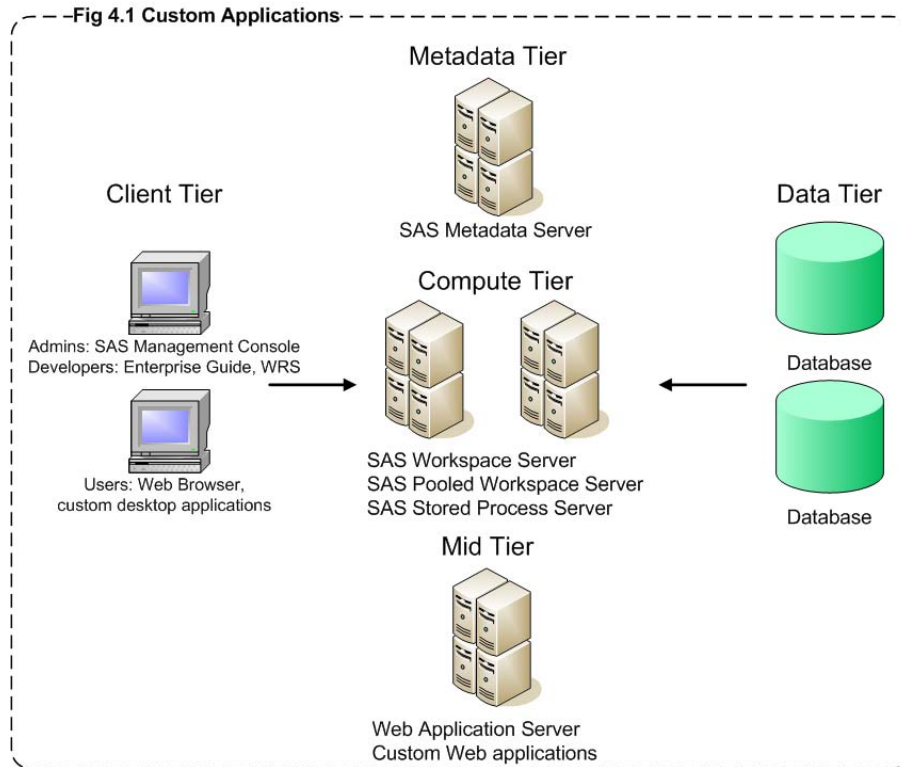
Participants

Role	Skill Set
Business Analyst	Business requirements knowledge
Application Developer	Integration Technologies, SAS APIs, Java or .NET
SAS Developer	SAS programming, Output Delivery System, Stored Processes
SAS Administrator	SAS administration, SAS metadata management

Collaboration

In this case, the application developer must not only know how to design and build applications with his/her preferred technology, but must also become familiar with the SAS APIs to communicate with the different SAS IOM servers; the SAS metadata and how it is organized; and the overall architecture of a SAS intelligence platform. SAS developers still create SAS routines, programs and stored processes that will be used from the custom application.The SAS and application developers work hand-in-hand design the application architecture.

The business analysts provide the business requirements for the application and acts as a liaison between development and the business for testing and training and adoption.

The SAS administrator typically has the skill set to deploy and maintain the application in his/her environment, and assists developers by providing knowledge on the SAS metadata.

Analysis

Developing custom desktop or Web applications with Integration Technologies ultimately depends on relationships. One must have the confidence that they understand the business requirements from end users well enough to provide a custom built interface that they will be responsible for, which includes its efficiency, validity and usability.

Also, since application development skills in Java or .NET technologies are not necessarily present within a SAS organization, a good relationship with a third party firm or an internal IT group must also exist in order to ensure a strong collaboration between the application developers and the SAS folks.

Paper #023-2012, continued

While this is not the most common pattern, the organizations who decided to go this way have done so because they had the confidence and skills to deliver custom applications, which resulted in a highly customized solution to their needs and in most cases, a lower cost in the short-term than licensing the SAS BI or EBI platform. The drawback here is that custom applications need to be maintained and updated internally.

Pattern #5: Portal & Dash-boarding, aka. The Executive Sweet

Description

High level executives need to make decisions based on succinct information that allows them to quickly identify trends and anomalies. Portals and Dashboards are the presentation layers of choice when it comes to deliver business information to our decision makers. This pattern deals with those of us who are fortunate to enjoy the benefits and ease of publishing of the SAS Information Delivery Portal and the BI Dashboard. All the components of the BI platform are included but the focus of this pattern is on the delivery of information and the leveraging of an enterprise publishing solution such as the SAS Portal.
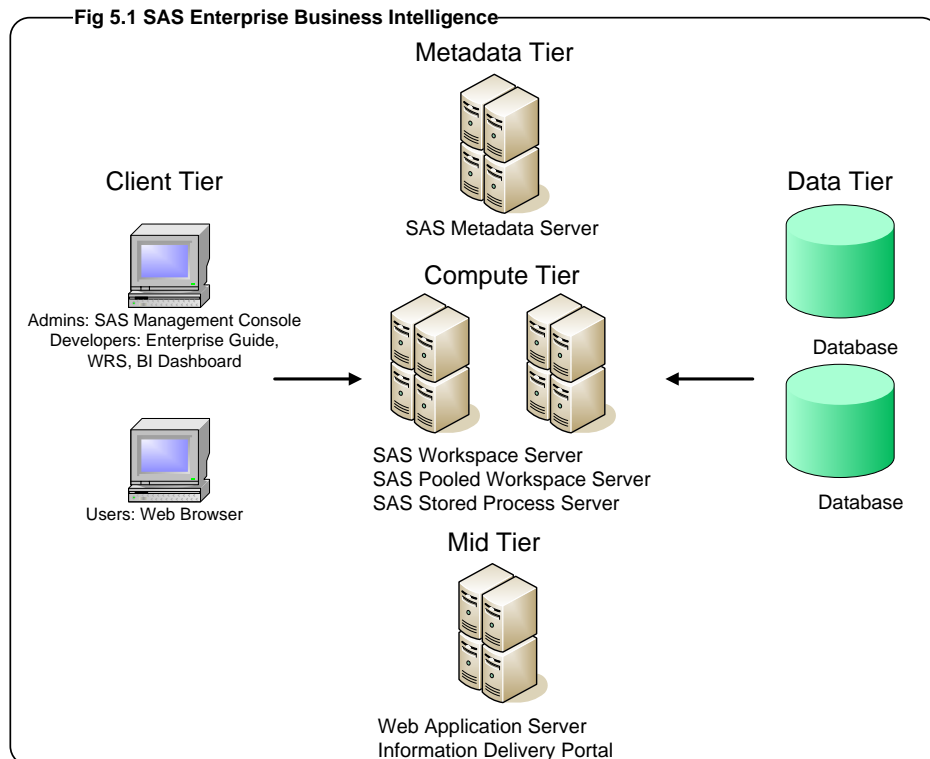
Usage

In this context reports, documents and all supporting documentation to make business decisions or to explain report anomalies are centralized in one location, they benefit from an appealing presentation layer to surface reports, and leverage a comprehensive publishing framework to share the information across the enterprise.

BI developers develop reports in one of the technologies available through SAS EBI and surface them in the Information Delivery Portal, they build custom dashboards that give quick insights on key metrics in their organizations, they also create custom pages for end users, arranging content to make it easily available to them.

Note that building custom portlets, in our opinion, falls in the Integration Technologies pattern. As we'll see in the next section, patterns are very often combined within an organization or even within a SAS environment.

Infrastructure

Again the architecture is not much different than the previous pattern. What has changed here is the type of end users consuming information and where they go to get it. We expect end users in this pattern to solely use a Web browser or Microsoft Outlook to retrieve reports, charts, and dashboards, and while they may be interactive, these users do not create new content in our SAS environment.

Fig 5.1 SAS Enterprise Business Intelligence

Metadata Tier

SAS Metadata Server

Client Tier

Admins: SAS Management Console
Developers: Enterprise Guide,
WRS, BI Dashboard

Users: Web Browser

Compute Tier

SAS Workspace Server
SAS Pooled Workspace Server
SAS Stored Process Server

Data Tier

Database

Database

Mid Tier

Web Application Server
Information Delivery Portal

Paper #023-2012, continued

Participants

Role	Skill Set
Business Analyst	Business requirements knowledge
SAS BI Developer	SAS programming, Output Delivery System, Stored Processes
SAS Administrator	SAS administration, SAS metadata management

Collaboration

In general, the SAS EBI platform may include ETL developers, portlet developers, and report developers, but the focus of this pattern is on the BI developer who is creating portal pages and dashboard for higher level executives. They obviously interact with all the aforementioned personas to figure out where the information is located but assuming they keep up to date with their environment, their main interactions are with business analysts to make sure requirements are fully understood, and with SAS administrators to manage page and because confidential information is more prevalent with high level executives, manage the security for all the portal content.

Analysis

Our blunt observation of organizations that are fortunate to have a comprehensive SAS EBI platform is that is it more often than not, very underutilized. There are several reasons for this.

One, the lack of knowledge about the tools and the functionalities they offer. Sales presentations on the SAS EBI tools are very impressive but while they clearly showcase the tools, they may fail to demonstrate how they would address your specific business needs.

Two, considerations for a new platform such as SAS EBI needs to include not only the technical infrastructure point of view but also the human factor: who is going to manage the environment, who is going to support our users, who is going to train developers, users, etc. If these critical areas are overlooked, tool adoption is typically low.

There is also a cultural challenge in adopting the SAS EBI tools. Your SAS users are not your typical users. Quite often, they have gained an arsenal of technical skills over the years that they may be reluctant to give up even for better and improved tools. It is a matter of demonstrating to them that they can work better, still utilizing the technical skills they have but much more efficiently.

PUTTING IT ALL TOGETHER

As you read each of these patterns, you may notice that your organization applies more than one and possibly all of them. For instance, an internal statistical analysis group might apply the first pattern which allows them to work independently and to be highly responsive to their customers. At the same time, the organization may have a large SAS EBI platform that serves ETL developers, Report writers and Dashboards developers in the different fashions described in the patterns. Another group within the organization may decide to leverage SAS by building their own interface to the very same server others use.

One of the reasons we split the patterns as we did was to illustrate the different needs in configuration, skills, collaboration and support for each of the usage pattern.

Obviously not everyone is fortunate to start out with all the SAS tools. Most organizations/departments start small and evolve or transition to a different pattern either to scale up their operations, or out of necessity to accommodate the strategic directions of their organizations, such as the centralization or decentralization of technical operations and/or infrastructure.

A transition may not be necessarily to a different pattern (after all our patterns are quite arbitrary at this point). The Reporting and Analysis pattern illustrates this very well. By acquiring better tools and adjusting its associated processses, an enterprise is able to move from ad hoc reporting to a more formalized reporting system that eventually will lead (we hope) to a well managed and controlled self-service reporting environment.

SURVEY

While all of this is very interesting, the obvious question is how well do these patterns reflect reality? We are confident that the reader will find them very familiar, but since one of the authors is a statistician (Kincaid), we understand that it is more important to have real data to quantify their prevalence. To that end, we are soliciting input from SAS users regarding how SAS is used in their organization.

Paper #023-2012, continued

The anticipated questions are

For each pattern which you are familiar, please answer the following questions.

1. Advanced Analytics
    a. Does your company implement this pattern with SAS?
    b. Are there significant differences in your implementation that we did not capture? If so, what?
    c. If not in SAS, then in what technology?
2. Data Management
    a. Does your company implement this pattern with SAS?
    b. Are there significant differences in your implementation that we did not capture? If so, what?
    c. If not in SAS, then in what technology?
3. Reporting and Analysis
    a. Does your company implement this pattern with SAS?
    b. If not in SAS, then in what technology?
    c. Are there significant differences in your implementation that we did not capture? If so, what?
4. Integration Technologies
    a. Does your company implement this pattern with SAS?
    b. Are there significant differences in your implementation that we did not capture? If so, what?
    c. If not in SAS, then in what technology?
5. Portal and Dashboarding
    a. Does your company implement this pattern with SAS?
    b. Are there significant differences in your implementation that we did not capture? If so, what?
    c. If not in SAS, then in what technology?
6. Please tell us about your organization.
    a. What industry is your organization in?
    b. How large is your organization?
    c. What group within your organization are you part of?
    d. How long have you been using SAS?
    e. How long has your group within your organization been using SAS?

The survey set up on Zoomerang and announced on some discussion forums. It might be found here, http://www.zoomerang.com/Survey/WEB22EFFVCDX97, depending on how far into the future you read this paper. Combining results from a regional conference with the online results, we have 41 responses. For now we will not try to generalize the results to the larger community, but we will continue to collect data.

The results we have so far are as follows. The respondents were from a variety of industries with none standing out. Therefore, no analysis of response by industry will be performed. Similarly for technologies used other than SAS. When these were answered, there was no consistency in the responses. As more responses are gathered, this may change. Finally, 50% of the respondents reported themselves as *SAS Programmer (power user)*, with the others scattered through the other choices.

The univariate response percentages are shown in the dot plot in Figure 1.
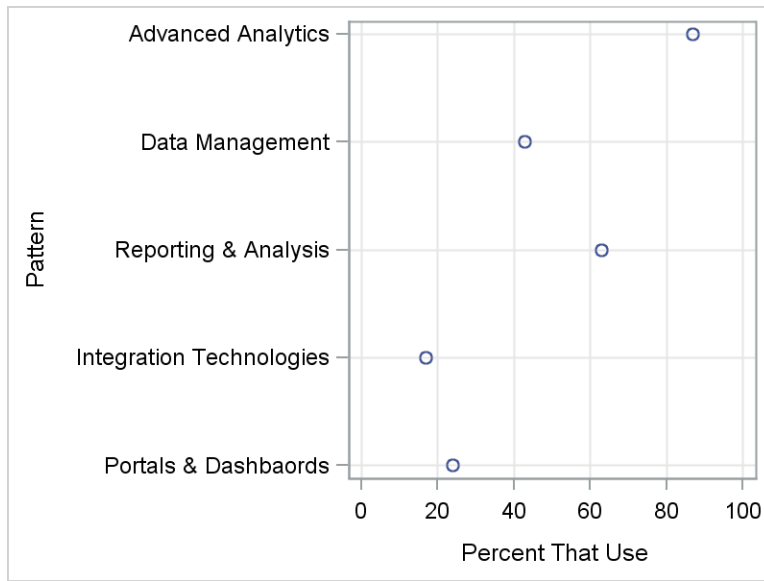
Paper #023-2012, continued

<div align="center">**Figure 1 Percentage of Respondents that use each Pattern**</div>

We also looked at the data considering the first three patterns as *Basic* and the second two as *Advanced*, and grouped the responses (Y or N) accordingly. For example, if a respondent's company used Advanced Analytics only, then their response was coded as "Y-N-N", and if they used Patterns 1, 2 and 3, then their response was coded "Y-Y-Y". The Advanced patterns were coded in a similar manner. With this explanation, the data gives us the following graphs.
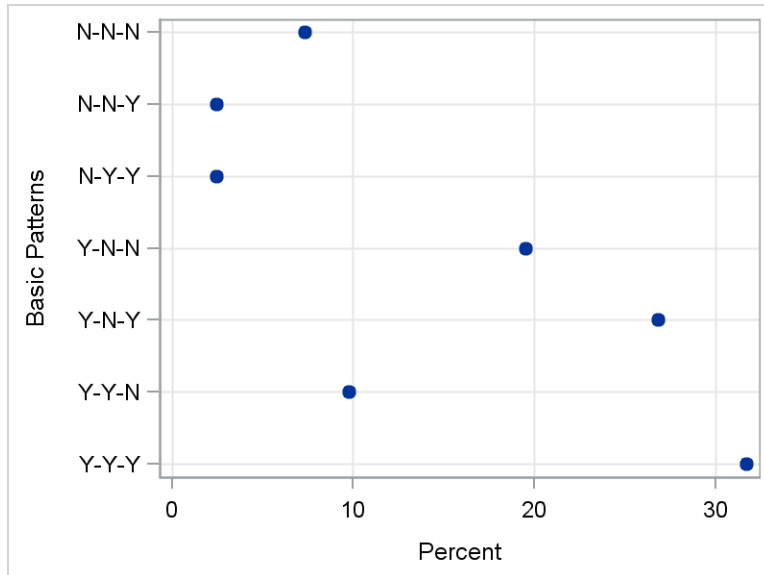

<div align="center">**Figure 2 Basic Pattern Usage**</div>

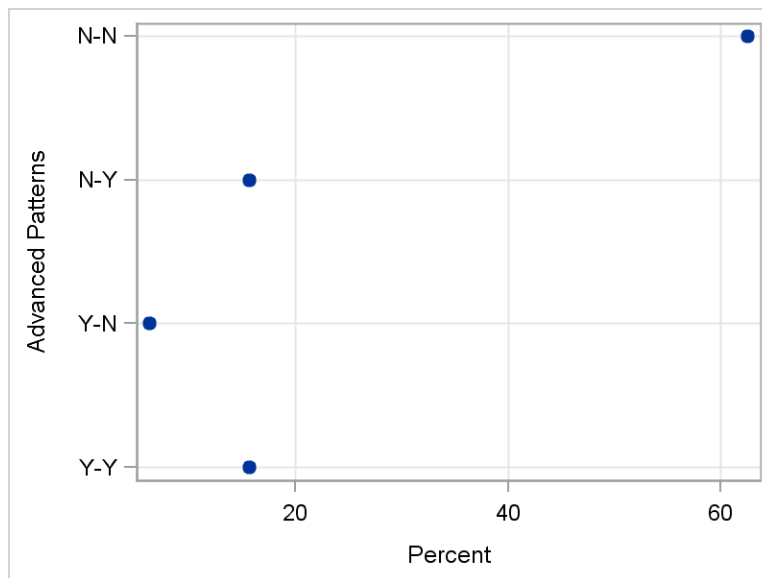Paper #023-2012, continued

<div align="center">

Figure 3 Advanced Pattern Usage

</div>

Looking at the survey results, it is not surprising to see that most organizations implement a combination of the three basic patterns (statistics, data management and reporting). While the SAS software is renowned for its advanced analytics capabilities, it also provides extensive features and tools to perform data management and reporting activities. As mentioned in the description of the first pattern, statisticians, when they work independently, will need to acquire and manipulate data before they can perform their analyses. These activities can easily become ETL processes if they are routinely run.

Another interesting observation in our survey data was the lack of presence of the SAS Portal and the SAS BI Dashboard. Because of the small sample size we hesitate to draw any conclusions from this observation. Whether most organizations have selected other tools or our survey did not capture a representative sample, we won't say. Companies that have selected other tools have chosen to forgo the single platform approach. Clients often ask us to integrate SAS with these other tools and systems and, as such, we understand the investment in time and money to achieve a smooth (a.k.a. fully automated) integration.

With respect to *how* these patterns are implemented one person noted the debate in their organization between using Enterprise Guide to implement these patterns and traditional interactive SAS. We see that a lot in organizations and the observant Global Forum attendee will recognize that many presentations are designed to provide a compromise between the two.

CONCLUSION

People use SAS in many ways. With the growing footprint of SAS across the enterprise in more and more companies, not everyone will know all the ways that SAS can be used, or how best to use it. The patterns proposed in this paper are the first steps in helping develop a consistent language for SAS implementation. We anticipate that a business manager looking for ways to increase their investment in SAS will be able to leverage the information provided to do just that.

We also welcome feedback, comments and discussion on these patterns to enhance the descriptions, and make them more meaningful to a wider audience of new SAS users.

REFERENCES

Ideal Type, http://en.wikipedia.org/wiki/Ideal_type

Gamma, Erich, Richard Helm, Ralph Johnson, John Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley. 1995.

ACKNOWLEDGEMENTS

We'd like to thank Jack Full, Brian Varney, Chris St. Peter, David Bosak, Chuck Binninger and Cindy Gardner for their valuable comments that improved the paper tremendously.

Paper #023-2012, continued

CONTACT INFORMATION

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

Mike Vanderlinden
Experis
5220 Lovers Lane
Portage, MI 49002
Phone: 269-553-5100
E-mail:mike.vanderlinden@experis.com

Chuck Kincaid
Experis
5220 Lovers Lane
Portage, MI 49002
Phone: 269-553-5140
E-mail:chuck.kincaid@experis.com