

Paper 051-2008

From Tiers to Intelligence: The SAS® Enterprise Intelligence Platform and What It's All About

Sandra Sayers, Amadeus Software Ltd, UK

ABSTRACT

Have you ever wondered what a servlet container is and why it's required to run SAS® Web Report Studio? Or maybe wondered what the Object Spawner is and exactly what it does? Or even wondered what a repository is and how it relates to the SAS® Metadata Server? And by the way, what is a SAS Metadata Server?

This paper is aimed at those of us whose answer to any of the above is 'yes'. It seeks to summarise and clarify the terminology associated with the SAS9 Business Intelligence offering.

The paper will contain a general overview of what the SAS Enterprise Intelligence Platform is. It will discuss the SAS9 multi-tiered environment and give an overview of the various SAS@9 applications, including SAS® Management Console, SAS® Data Integration Studio and SAS® Web Report Studio.

It is, in short, a summary of the SAS Enterprise Intelligence Platform, and what it's all about.

INTRODUCTION

The idea for this paper came about a while ago, when I had to start performing installations myself, with my new laptop being my guinea pig. A colleague recommended that I install SAS Enterprise BI Server with Xythos, so that I could familiarise myself with the SAS Web Applications. Just hearing this left me cold, because a) I wasn't quite sure what was included in the SAS Enterprise BI Server bundle, b) I had no idea what Xythos was and c) I couldn't use SAS code to do it. So I started talking to people, and ploughing my way through the mountain of documentation provided by SAS. From there the idea of an overview of the SAS Enterprise Intelligence Platform was born, to clarify some of the terminology and to help you (and me) understand what it is and what it consists of.

The paper will explore the SAS9 multi-tiered architecture, and highlight the most important areas from each tier. I will also be giving an overview the different client and web applications available to the platform, and discuss their capabilities.

Anyone who's seen the amount of documentation that comes with SAS9 will know that there is a lot to know. This paper is designed to provide an overview of the most important concepts of SAS' business intelligence offering.

First let's set the scene and start by looking at what Business Intelligence is.

WHAT IS BUSINESS INTELLIGENCE?

Type this question into any popular search engine and you will get a variety of responses. It is basically all about providing the right information, to the right people, at the right time, in the right format via a set of software tools and applications. It will enable business users to interact with their company data in an easy, efficient and effective manner, meaning that the analysts can spend more time analysing and the decision makers can spend more time making decisions.

SAS AND BUSINESS INTELLIGENCE

SAS has developed an end-to-end business intelligence solution named the SAS Enterprise Intelligence Platform. It enables you to extract and transform your data into a fully integrated reporting warehouse, leading to more useable business data and more effective decision making.

Because of the way the platform is structured, it means that the data is stored only once, and can be accessed by all of the different client applications, leading to consistent answers. Seems pretty obvious, but how many times have you tried to calculate some figures and your answer differs depending on who you speak to / which type of data extract you used / the lunar phase etc? The SAS EIP is designed to stop all of that.

Figure 1 below highlights the structure of the platform.

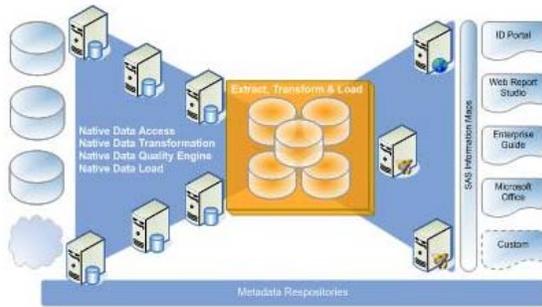


Figure 1: The SAS Enterprise Intelligence Platform

The SAS Enterprise Intelligence Platform is sold in many different “bundles”, depending on your business requirements. For example, SAS® Enterprise Business Intelligence Server, which is an end-to-end product or SAS® Data Integration Server, which focuses on the area of integrating data.

METADATA

Metadata is often referred to as “data about data”. In the context of SAS, metadata is information about all of the resources used in the provision of BI. For example it will be;

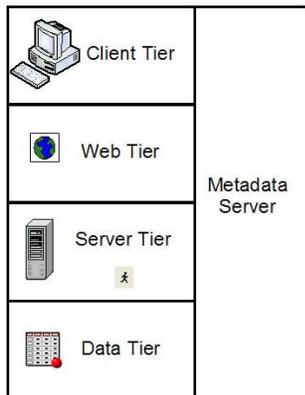
- the name of a column rather than the actual values stored within it
- information about how to read in data from a non-SAS location
- a description of the report content, rather than the actual report itself
- details of the access rights of all users
- a description of the source and target tables used in data integration jobs, as well as details of all of the transformations.

SAS stores its metadata in central repositories and this metadata is accessed and managed by the SAS Metadata Server, which I will talk about in more detail later.

All of this is supported by the SAS® Open Metadata Architecture. This allows the different SAS applications to share the same metadata, achieving consistency in the data, and in interpretations of the data.

MULTI-TIERED ARCHITECTURE

The SAS Enterprise Intelligence Platform is split into different parts, or tiers. The metadata sits alongside these tiers and is the glue that sticks them all together. There are 4 different tiers; Data, Server, Client and Web.



These 4 tiers can be installed on separate machines or all on one. It is common to have the data, server and client tiers all on separate machines, and maybe the web on the same machine as the server tier.

Figure 2: Multi-Tiered Architecture

DATA TIER

The data tier is where the enterprise data sources reside. The data storage options within the SAS Enterprise Intelligence Platform are detailed below in Table 1:

Table 1: Data storage options

| Format | Description |
|-----------------------|--|
| SAS Data Sets | SAS' format for data storage. |
| OLAP Cubes | Multi dimensional data cubes. |
| SAS SPDS and SAS SPDE | Parallel storage. |
| RDBMS data tables | SAS/ACCESS is merely the way in which SAS can talk to other data structures, e.g. Oracle, in order to make the data usable in SAS. |
| ERP systems | Using the SAS Data Surveyor products, ERP systems such as SAP and some Oracle applications can be directly accessed. |

SERVER TIER

This is where the SAS Servers are installed. SAS Servers are similar to windows services, in that they run in the background, servicing requests from client applications.

SAS Servers can be (and often are) spread over many machines. They can be on the same or a different machine to the client applications

SAS SERVERS

Table 2 lists and describes all of the Servers in the Server tier.

Table 2: SAS Servers

| Server | Description |
|---------------------------|--|
| SAS Workspace Server | This is where the SAS code, generated and submitted by the client applications, is sent to be executed. There is an individual process for each user. It is equivalent to a SAS Display Manager Session or the execution of a SAS batch job. Client applications can access SAS libraries, execute SAS code and output the results. |
| SAS Stored Process Server | This is where all SAS Stored Processes are executed. SAS Stored Processes are centrally stored SAS programs that can be accessed from the different SAS client and web applications. |
| Object Spawner | This is used to start the SAS Workspace Servers and SAS Stored Process Servers. It connects to the SAS Metadata Server |
| SAS OLAP Server | This is used to access, not create, OLAP cubes. It can be used to read in OLAP cubes from other non-SAS applications. Processes MDX queries and returns data from cubes. Workspace server builds the cubes, OLAP server simply queries them. |
| SAS/CONNECT Server | Can be used to submit code from one machine to be executed on another. Provides Remote Library Services, update data on other servers as if it was the host. Data Transfer Services allows for transfer of data between machines. Needed for SAS Promotion and Replication. Can be used by DI Studio. Plays a role in grid computing |
| SAS/SHARE Server | SAS/SHARE gives access to data files for multiple users at the same time. It will also allow users to avoid using separate SAS/CONNECT remote log ins when accessing SAS files on a server |
| Platform Suite for SAS | This is used for scheduling all elements of the BI process. |
| Batch servers | These are used to execute SAS code. They associate code with program that will execute it and also allow for scheduling. |
| Grid Server | Allows SAS/CONNECT sessions to be used by Platform LSF. |

The most important of the above are the SAS Workspace Server, SAS Stored Process Server, Object Spawner and SAS OLAP Server.

A Server, in this instance, is a program which can be requested from client applications. These server programs fulfil requests for data or services and can either run on the same or a different machine to the client applications. For example, a user of the client application, SAS Enterprise Guide, may run a job which produces a summary report from a dataset. The request for this data is sent ultimately to the SAS Workspace Server, where the actual SAS code is run, and the output is then sent back to the user.

The way this works can be seen below in Figure 3.

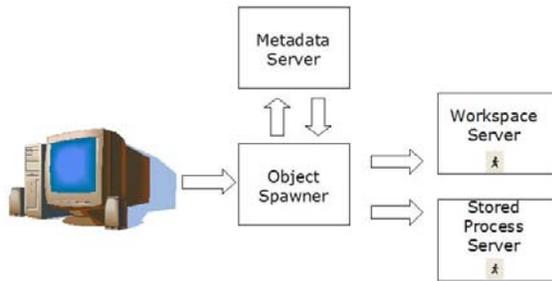


Figure 3: SAS Servers

The Object Spawner listens for requests from client applications. When a request is received for either a SAS Workspace Server or a SAS Stored Process Server, the Object Spawner will turn to the SAS Metadata Server. The SAS Metadata Server then in turn checks the Metadata Repositories for the next available server and lets the Object Spawner know. The Object Spawner can then go the relevant server and start it, so that the code/stored process from the requesting client application can execute.

SAS APPLICATION SERVER

So, what is a SAS application server? In this instance, an application server is a way of grouping together different logical servers, which leads to the question of what is a logical server?

A logical server is a group of one or more servers. For example, if you install one or more SAS Workspace Server, they will all be grouped together and called the logical Workspace Server. The Stored Process Server will be grouped together as the logical Stored Process Server. It is not possible to install multiple instances of all servers, for example is only possible to have one OLAP server within the logical OLAP server.

The logical servers belonging to the SAS Application Server are as follows:

- SAS Workspace Server
- SAS Stored Process Server
- SAS OLAP Server
- SAS/CONNECT Server
- Batch Server
- SAS Grid Server

All of these server and grouping definitions are stored within the SAS Metadata Repository as Metadata Objects. I will come back to SAS Metadata Repositories a little later on.

All application servers, logical servers and actual servers can be viewed in the Server Manager Plug-In of the SAS Management Console, as can be seen in Figure 4:

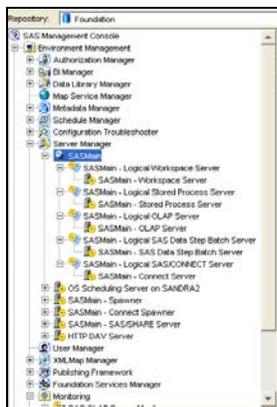


Figure 4: SAS Application Server

Generally speaking, if the application server is located on the same machine as the SAS Metadata Server then the default name of SASMain will be given, as in the above example. Otherwise, the application server will be called SASApp.

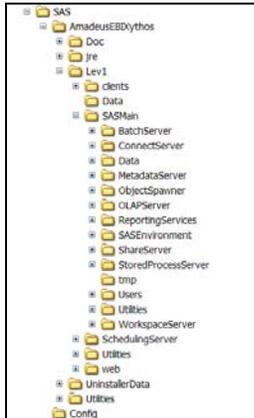


Figure 5: SAS Installation directory

The configuration directory which will exist on the machines where SAS is installed will look something like Figure 5.

Exactly which folders and sub-folders are present depends on what has been installed.

So what is the point of grouping together physical servers into logical servers and application servers? Well, just imagine that you are running data warehouse load jobs in SAS Data Integration Studio. You can get your jobs to all run on the same physical servers by simply specifying the SAS application server name rather than worry about the names of servers stored within it. It is also possible to assign a library or other such resources to a specific SAS application server, meaning that it will be available to only those jobs running on that specific SAS application server.

Logical servers are also useful when setting up user access rights, so if you allow a user access to a logical server it mean that they can all access all of the physical servers grouped within it.

The reason you may wish to set up multiple SAS Stored Process Servers or SAS Workspace Servers is so that you can spread the workload amongst them. There are two methods you can use for this; load balancing and pooling.

LOAD BALANCING

Load balancing is a method of sharing the workload amongst other SAS Stored Process Servers and SAS Workspace Servers defined in the logical servers. If you find that you are in need of more resource, then you can add a new physical SAS Stored Process Server or SAS Workspace Server to its logical server. An Object Spawner would also need to be set up for each of the SAS Workspace or SAS Stored Process Server added. Object Spawners for each server talk to each other and to the SAS Metadata Server to find the least loaded server. This is useful for applications such as SAS Data Integration Studio which submit long jobs.

As an example, if you are running multiple jobs in SAS Data Integration Studio, rather than have the jobs wait in a queue for resource on the SAS Workspace Server, the Object Spawners would check with the SAS Metadata Server to see which Workspace Server had the lighter load and it would send the client request in that direction.

POOLING

Pooling is available only for SAS Workspace Servers. Each time a client application makes a request to the SAS Workspace Server a new server connection is made. This can take time, especially if you are using a SAS Web Application such as SAS Web Report Studio, which makes lots of short connections to the SAS Workspace Server. Pooling is a way of avoiding this, as it means that you can set up some server connections which are reusable, thereby avoiding the need to create a new connection for each request.

WEB TIER

This tier is sometimes called the Middle Tier and it contains everything that you need to be able to run the SAS Web Applications, such as SAS Web Report Studio or the SAS Information Delivery Portal.

The Web Tier actually includes the following:

- Servlet container or J2EE Application Server
- Java 2 Software Development Kit
- WebDAV Server
- SAS Application Services
- SAS Foundation Services
- SAS Web Infrastructure Kit (WIK)

- SAS Web Applications (e.g. SAS Information Delivery Portal and SAS Web Report Studio)

All of this seems like a lot, but it's all configured by the installation process.

SERVLET CONTAINER AND J2EE APPLICATION SERVER

The Web Tier needs either a Servlet container or a J2EE Application Server to be able to function correctly.

So what does this all mean? Well, basically all SAS Web applications, such as SAS Web Report Studio and the SAS Information Delivery Portal are built using Java Servlet technology. To be able to run effectively, the Java Servlets (applications) and JavaServer Pages first need to be compiled using the Java 2 Software Development Kit. They then need the correct environment to execute in, which is where the Servlet container comes in to play. The job of a Servlet container is to provide the environment needed for Java Servlets (applications) to run. A common Servlet container used is Apache Tomcat. Apache Tomcat is free and open source and so may not meet security requirements, and so this is where a J2EE Application Server may be preferable. A J2EE Application Server is a Servlet container, but also includes a lot more. Examples of J2EE Application Servers commonly used with the SAS Enterprise Intelligence Platform are BEA WebLogic Server and IBM WebSphere Application Server.

WEBDAV SERVER

WebDAV stands for Web-based Distributed Authoring and Versioning, and it is a way of storing and organising your files on your web server, similar to the way in which you can organise your files on a Windows machine using Windows Explorer. It means that you can store all of your files on a web server, but you can manage them from any desktop with a web browser. A WebDAV server is required if you would like to enjoy all of the SAS Information Delivery Portal capabilities. At the time of writing this paper Xythos WebFile Server is most commonly used with SAS.

SAS APPLICATION SERVICES

This is a set of services, including the SAS Query and Reporting Services, which provide business-oriented query and reporting services to requesting client applications.

SAS FOUNDATION SERVICES

This allows custom built Java applications to be incorporated into the SAS EIP, via Java-based Application Programming Interfaces (APIs). It also needs to be installed if you would like to access the 'BI Manager' plug-in in the SAS Management Console.

SAS WEB INFRASTRUCTURE KIT (WIK)

The SAS WIK provides the infrastructure necessary to make full use of the SAS Information Delivery Portal. It allows for:

- User logon and logoff
- Page navigation
- Searching
- Integration with SAS via stored processes
- Interacting with basic content types

It also includes:

- SAS Stored Process Web Application - to allow stored processes to be run from the requesting web clients
- SAS Services Application - which provides a set of reusable services to the SAS Web Applications

CLIENT TIER

The client tier is where all of the client applications will be installed. There are many client applications available to the SAS EIP, and they are installed on the users PCs. These applications are the main interface to the SAS EIP. They all allow the user to log on the same Metadata Server, so that regardless of the application being used, the same data and metadata will always be used. These client applications are either java based, windows based or accessed via a web browser.

Types of Client Applications:

- Java
 - SAS Management Console
 - SAS Data Integration Studio
 - SAS OLAP Cube Studio
 - SAS Information Map Studio
- Windows
 - SAS Add-In for Microsoft Office
 - SAS Enterprise Guide

- Web (surfaced via a Web Browser)
 - SAS Web Report Studio
 - SAS Information Delivery Portal

SAS METADATA SERVER

As I mentioned earlier, the metadata is the glue that joins the tiers together. Metadata describing the entire SAS environment is stored in SAS Metadata Repositories, and is accessed and managed by the SAS Metadata Server. The SAS Metadata Server will access these repositories when it receives a request from a client application.

Some examples of what the metadata server does:

- When you open up any of the SAS client applications and enter the log on credentials, the SAS Metadata Server will check in the repository whether you are a registered user and whether you are allowed access to it or not.
- When you are running a stored process, the SAS Metadata Server will receive a request from the object spawner to start a stored process server, so it will check to see which one is free and then let the object spawner know.
- If you are in SAS Web Report Studio to view a report, the SAS Metadata Server will check to see if you are allowed to view that particular report.

The SAS Metadata Repositories are simply a collection of SAS datasets and contain Metadata Objects. Metadata Objects are descriptions of different elements of the SAS Enterprise Intelligence Platform, rather than the actual element itself. Examples of different Metadata Objects and what they describe can be found in Table 3:

Table 3: Metadata Objects

| Metadata Object | What it describes |
|------------------|--|
| Data tables | This object describes what is contained within the table, such as column names, formats and the physical location of the table. |
| Libraries | This object will describe everything necessary to generate a base SAS libname statement, e.g. the name of the library, the physical location it points to and any options. |
| Jobs | This object describes all of the transforms, source and target data which make up the jobs. It is basically everything necessary for SAS code to be generated and successfully executed. |
| Cubes | This object describes how the cubes will look, what the dimensions and hierarchies are. |
| Information Maps | Metadata about data source used to build the map, how the different sources relate to each other, what the end user is going to see, details of any filters applied. |
| Reports | This object describes what the reports will contain and how they will be structured. |
| Servers | Each server metadata object describes the host machine name, the port(s) on which the server listens for requests and details of how to start the server. |
| Users | This object describes the user logon details what they have access to. |

SAS METADATA REPOSITORIES

There are three different types of repository; Foundation, Custom and Project.

- Foundation
 - This is the parent repository and contains metadata to be accessed by everyone, e.g. server and library definitions.
- Custom
 - This repository is dependent on either the Foundation repository, or another Custom repository. It will typically contain metadata specific to a certain department or project team, e.g. jobs to load the sales department data warehouse.
- Project
 - This repository is used in a change managed environment in SAS DI Studio and can be dependent on either the Foundation or Custom repositories. It protects against multiple users potentially making changes to the same files.

As the metadata is stored only once, in one place, this means that no matter which application you use to ask the question, the answer will always be the same.

SAS CLIENT APPLICATIONS

There are many different client applications available with the SAS Enterprise Intelligence Platform. This section will give descriptions of the following most commonly used SAS client applications:

- SAS Management Console
- SAS Data Integration Studio
- SAS OLAP Cube Studio
- SAS Information Map Studio
- SAS Add-In for Microsoft Office
- SAS Enterprise Guide
- SAS Web Report Studio
- SAS Information Delivery Portal

There are many more “solution specific” client applications, such as SAS CRM, SAS Enterprise Miner, SAS Forecast Studio etc, but these will not be discussed in this paper.

SAS MANAGEMENT CONSOLE

The SAS Management Console is the SAS Enterprise Intelligence Platform administration application used to manage metadata, and perform many administrative tasks. It allows you to manage the following:

- Server definitions
- Library definitions
- User definitions
- Resource access controls
- Metadata repositories
- SAS licenses
- Job schedules
- XML maps

These are all managed through Plug-Ins. For example, the Server Manager plug-in allows you to manage the server definitions. The User Manager plug-in allows you to add or remove users and groups. Extra plug-ins can be installed as and when required.

The SAS Management Console can also be used to promote (move) metadata between environments.

SAS DATA INTEGRATION STUDIO

SAS Data Integration Studio (formerly SAS ETL Studio) is the application used to extract data from source systems, transform this data into useable business data and load it into a data warehouse or data marts. During the ETL process you can also clean up and validate your data using DfPower Studio, a data quality tool which can be integrated into the process. SAS Data Integration Studio can also be used to create OLAP cubes.

In SAS Data Integration Studio, different process flows (or jobs) are built up using the Process Designer window. These processes are dragged and dropped into a process flow specific options will build up PROC SQL code behind the scenes. Custom processes can also be created, and incorporated into the process flow. Any Source and Target tables used will first need to be defined using the Source Designer and Target Designer.

SAS Data Integration Studio allows a change managed environment. Any source and target tables and jobs can be ‘checked out’ into a Project Repository when they are being updated. Only the person who has checked them out can work on them, until they are checked back in again.

SAS OLAP CUBE STUDIO

There are two ways of creating OLAP cubes in the SAS9 platform; with SAS DI Studio, or with SAS OLAP Cube Studio. This application contains a user friendly point and click Cube Designer to build up the cube from scratch.

Cubes can be built from either a detail table or a star schema structure. Care must be taken if you have a snowflake schema data warehouse, as you will not be able to build a cube if you have a dimension hanging off another dimension. For those of you who are SAS programmers, what the cube designer is actually doing is building up a PROC OLAP step behind the scenes. This PROC OLAP is sent to the SAS Workspace Server to be executed.

SAS OLAP Cube Studio cannot be used to view OLAP cubes. If you would like to look at the cube you have just built then you would need to open it through one of the following applications:

- SAS Enterprise Guide
- SAS Web Report Studio

- SAS Information Delivery Portal
- SAS Add-In for Microsoft Office

To be able to access OLAP cubes, the SAS OLAP Server needs to be installed.

SAS ADD-IN FOR MICROSOFT OFFICE

If you are a Microsoft addict then you will be pleased to know that it is now possible to work with SAS data from within the different Microsoft Office applications. Once this has been installed on the client machine an extra menu called 'SAS' will appear at the top. This will give you the ability to view and analyse SAS data sets, execute stored processes and do much more. For example, you could run a Stored Process from within Microsoft PowerPoint, which will automatically update the relevant charts and figures on the slides.

SAS ENTERPRISE GUIDE

SAS Enterprise Guide is an application design to allow you to use all of the capabilities of the SAS programming language, without having to really know the code. It is an easy to use point and click, drag and drop environment. For example, instead of writing 'procedures' to build up a SAS program you will add 'tasks' to an Enterprise Guide project to build up a process flow. By selecting different options in the tasks you are essentially building up the SAS code behind the scenes. The one downside to this application if you are a SAS programmer is that there are no tasks which replace the data step, however it is possible to incorporate custom made data step code into the projects.

Enterprise Guide, like the base language, has strong analysis and reporting capabilities. You will also be able to create SAS Stored Processes, which can then be run either in Enterprise Guide, in the Microsoft Office applications through use of the SAS Add-In, or in SAS Web Report Studio and the SAS Information Delivery Portal.

SAS INFORMATION MAP STUDIO

This is an application to allow you to build Information Maps, and is needed to be able to create web reports, through SAS Web Report Studio. An Information Map is essentially a way of simplifying the data for the end business user. After all, the end user is not going to care whether the columns that they want to use in their report come from one table or five. The Information Map hides the origin of the data and all of complex joins between tables, and will supply the end user with a list of columns which can be selected in their reports. It is basically the same as a SAS data set view.

When the Information Map is being built, PROC SQL code is being generated behind the scenes. If you are using the Map to provide a view across a range of tables, be aware that the all of the tables will be joined regardless of whether they contain the columns you want to look at or not. This can impact on performance time, so it is often better to create a series of 'mini-maps' to join together a few tables at a time, rather than to create one big map, which joins everything at once. This will all depend on the reports needed, so you will need to be very clear about the requirements.

SAS WEB REPORT STUDIO

SAS Web Report Studio is an end user web-based reporting application, and is accessed via a web browser. It is a point and click, drag and drop application which allows users to create, view and share reports. SAS Web Report Studio works from an Information Map.

SAS INFORMATION DELIVERY PORTAL (IDP)

The SAS Information Delivery Portal is a single point of entry to business intelligence content. It is a customisable webpage which, once logged into, allows you to view all of the different Business Intelligence reports that you have access too. The welcome page from where you are able to log in is called the Public Kiosk, and this can also be customised.

All of the BI content in the Portal is organised into pages and portlets. The pages are different tabs within the website and portlets are the different windows or boxes that make up each page.

The sorts of things that can be viewed in a portlet are:

- Web links
- SAS Information Maps
- SAS Web Reports
- SAS Stored Processes

It is common practice to use a WebDAV server such as Xythos WebFile Server to access all portal features.

CONCLUSION

The SAS Enterprise Intelligence Platform is an end-to-end framework for delivering BI across an organisation. Through its multi-tiered, metadata-sharing architecture, it meets the needs of different types of users, and opens up

the power of SAS to a wider audience.

REFERENCES

Pratter, Frederick. *Access to SAS® Data Using the Integrated Object Model (IOM) in Version 9.1*. Eastern Oregon University, La Grande OR.

Bostic, Pat, Thacher, Clarke, Crow, Dave and Roth, John. *Top Deployment Considerations for the SAS® 9.1 Intelligence Architecture*. Cary, NC: SAS Institute Inc.

SAS OnlineDoc® 9.1.3

The SAS® 9.1.3 Intelligence Architecture: Planning and Administration Guide along with other reference documentation can be found at <http://support.sas.com/documentation/configuration>

CONTACT INFORMATION

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

Sandra Sayers
Amadeus Software Ltd
Orchard Farm
Leafield, Oxfordshire, OX29 9PG, UK
Work Phone: +44 (0)1993 878287
Fax: +44 (0)1993 878042
E-mail: sandra.sayers@amadeus.co.uk
Web: www.amadeus.co.uk

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

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