

Paper 489-2013

Key Aspects to Implement a Perfect SAS® BI Platform

Gaurav Agrawal, Major Financial Company, Phoenix, AZ

1. ABSTRACT

A perfect SAS architecture is not defined just by successful installation of a SAS platform but also by ensuring good performance, easy maintenance, compliance to all security, Secured environment, scalability, good administration practices, proper monitoring, seamless Integration with Interfacing system etc. SAS provides lot of flexibility in order to integrate with other interfacing systems. However a perfect SAS Enterprise Implementation requires matured implementation of other interfacing platforms. SAS user experience starts from the first click on the SAS client and driven from SAS environment capabilities and its integration to interfacing systems. User experience can be further enhanced only if systems around SAS environment provide seamless integration. Hence interfacing systems also have a key role to play in order to get perfect SAS architecture.

2. INTRODUCTION

As discussed this paper will be more focused around the key aspect of making existing/new SAS environment “A perfect BI SAS Intelligence Platform” to meet all customer needs. To achieve it the interfaces of SAS Environment should also provide seamless integration. Examples of such interfaces are servers on which SAS is installed, Database/DW Security/processing, LDAP/AD (used for Authentication), Network, SAN, FTP, SSO environment etc. This paper will provide details of some of the key aspect around same, which can be very valuable to enhance User experience on system.

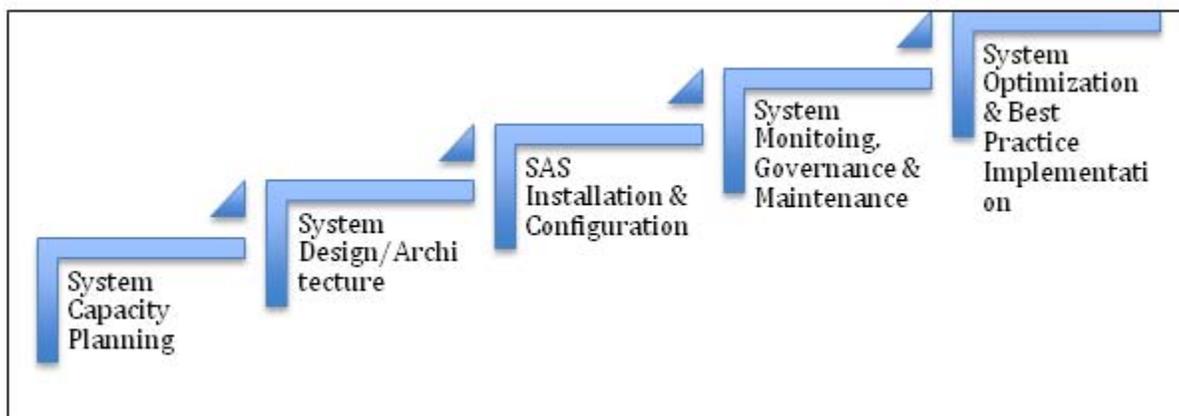


Figure 1: Key Phases of SAS Environment Deployment

Above depicted picture is a high level view of some of the phases to be performed for any new or existing environment. Many/All for the activities shown above can be categorized as repetitive considering every system can be designed only for certain period. Hence system has to be re-designed or completely reviewed after certain period of time. I've also seen some over-engineered SAS environments and such review/redesign activities help to simplify the system. This in turn enables company to bring more capabilities to environment by reducing complexity and maintenance effort.

Note: This paper is kind of booklet to give direction for different key aspects to be considering when designing SAS environment. Considering most enterprise does have their preferred hardware hence solution cannot be done for all possible combination in this paper.

3. SYSTEM CAPACITY PLANNING

This is very first phase of designing the system. In case you are upgrading an existing SAS system then it is very much possible to have/collect system performance statistics and usage patterns for doing planning. However if this exercise to be done for new SAS system then collecting/predicting required capacity of system is very challenging task. For example a program written by one SAS expert in comparison to absolutely new SAS user may show major performance difference and as well varied resource utilization.



I do recommend including SAS Institute's Enterprise Excellent group for the capacity estimation exercise because they do have a very well defined process for it. Hence outcome of that process provides a very good baseline for your platform's capacity. However key aspect to be considered that environment should be scalable. Capacity planning arises from the fact that nothing in your enterprise stays the same and after a certain period this exercise has to be repeated.

Considering SAS environment is used for analytics hence any variation in data size, reporting requirement, number of users etc. will impact the Capacity Planning exercise carried out for system. An architect has to consider all growth plan of organization during designing a system. It is also to be considered that we should design a system, which is cost effective hence optimal hardware should be considered for system. Considering SAS grid environment system can be scaled up any time from CPU and Memory perspective hence it should be considered while designing SAS system. Regarding system throughput (IO), SAS institute does have its own recommendation per core throughput, which should be considered and achieved. These recommendations are made considering the processing architecture of SAS jobs hence getting recommended throughput will achieve better performance for users. I do recommend that at the time of capacity planning we should consider "Business Capacity Management (Business Projection)", "Resource Capacity Management (Hardware Projection)" & "Service Capacity Management (System Maintenance Projection)".

4. SYSTEM DESIGN

This is very important phase for designing SAS system. In this phase we design that how SAS system will look like after implementation. Hence all things should be very well considered and accordingly designed. SAS Architects has to take input from various stakeholders in organization for completing this phase. Sizing exercise done before this will be input to design SAS system. There are certain standard recommendations for environment and in this section we'll discuss same. These recommendations help in building or enhancing the topology of SAS environment.

4.1 Metadata, Compute and Web-Tier on Separate Server

Administrators have a great deal of flexibility when defining the topology for a SAS deployment. However it is always recommended if we can define Metadata, Compute, Web and Client tier separately. As per the SAS Environment requirement resources can be defined to these separate boxes.

Below mentioned are the details of different standard tier of SAS environment.

- Metadata tier—manages information shared by all components in a deployment, controls authentication and authorization and enables communication among server processes
- Compute tier—also known as the SAS Application Server tier, provides SAS analytic and business processing to requests initiated by the client and middle tiers. In Grid environment it can be collection of Grid Node.
- Middle tier—also known as the Web tier, supports SAS Web applications and services. Starting from SAS 9.3 even all data mining products, which were deployed in SAS Analytical Platform in earlier versions, will be deployed on mid tier. However flexibility of keeping these analytical tools in separate Application Server is still there.
- Client tier—also known as the end-user or business-user tier, personal workstation hosting user interfaces to the SAS environment, including executable applications (such as SAS Enterprise Guide®) and the Web browsers in which SAS Web applications run

4.2 SAS Environment Segregation for Development, Test & Production Environment

SAS provides flexibility of implementing/installing/configure SAS environment as per customer needs. Recommended approach will be to maintain Development/ Test (UAT)/ Production environment on separate servers. In case an IT budget cannot afford OR not willing to have separate DEV/Test/Prod environment then DEV/TEST can be configured on same server(s) and Production can reside on other server(s). In order to achieve same there are different solution that can be considered

1. Configure SAS environment as Lev2 (TEST) & Lev3 (DEV) on different port on same server(s). Under this approach two Metadata servers will be running in one SAS implantation, which can work for Test and Development respectively
2. Second method can be that define separate Metadata folders for DEV and Test environment and segregate contents using it. Also separate Application Servers Context can be configured for Development and Test environment in same SAS implementation. Hence in this case there will be only one Metadata service instance but it will work for both. This is my personal preferred approach to avoid multiple Metadata server load on server.

In the scenario if funding is very less and even its challenge to get production environment separately then above mentioned method could be extended even for production environment. However it will bring risk to Production environment from load perspective.

You can see more details for it in paper <http://www2.sas.com/proceedings/forum2008/028-2008.pdf>

4.3 Segregation of Adhoc Analysis and True Production Jobs/Reports

This recommendation requires some additional funding from infrastructure however I feel it brings lot of benefits from load perspective. Considering the capabilities of SAS we know that SAS environment is used for lot of Adhoc analysis. Such kind of analysis, users like to do directly in production environment on real data and that is nature of approximately all SAS environment implemented in organization. However this brings a challenge that such kind of load can impact other true production jobs running in same environment. Hence it is recommended if we can segregate true production environment jobs from such Adhoc Analysis activities. When I say "true production activities" then my mean is that all queries/reports, which are well tested in Test environment and need to run/schedule in production environment. By separating environment we'll ensure that Adhoc analysis activities do not impact real production jobs/reports. However this approach poses a challenge that in this case data will be redundant if environments are completely separate. Hence in order to reduce SAN foot print suggestion can be to create a shared SAN for Adhoc and Production analysis work. Only drawback for this approach will be that a defined IO provide by storage system will be shared between both environments.

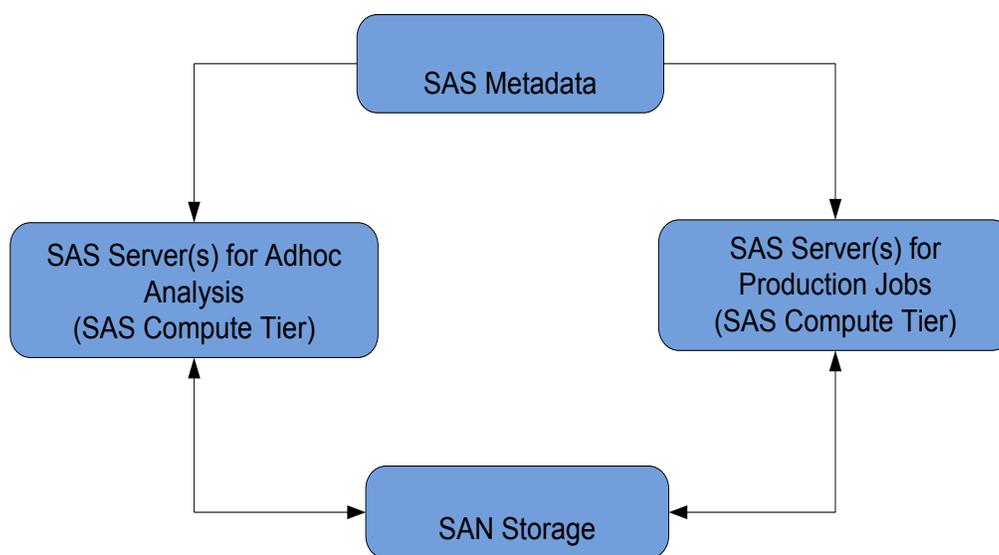


Figure 2: Shared SAN between Adhoc and Production

For example in SAS Grid environment we anyways do implement Shared SAN to make processing independent of node. Hence if we can consider segregating server(s) for Adhoc and Production activities then it will ensure that all production jobs are not impacted because of users activities. As we know that SAS Application Server context can trigger user session on defined server(s) however in order to trigger Batch Jobs on specific server there various architecture can be consider and need to decide on the basis of specific situation of your organization. Some methods are outlined below

1. All true production jobs can leverage SAS Grid Queue concept to be executed on specific node in case LSF scheduler is used for scheduling jobs
2. If enterprise scheduler is used for scheduling SAS jobs then GSUB utility can be used to trigger job, which can leverage Queue and in turn will trigger job on specific servers. Bottleneck is that before SAS version 9.3, GSUB utility do not give functionality to wait till job completes or do not return job status properly. Starting from SAS 9.3 this utility has been enhanced and can be leveraged with extra coding to capture Job return code
3. It will be good idea to leverage Enterprise Scheduler Load Balancing techniques for scheduling jobs. If that functionality is not available then a load balancer switch can be considered for load balancing the job. This is my personal preferred approach.

4.4 SAS Environment Security – SAS Server(s) Security

Default Authentication configured for SAS environment on Windows platform can easily leverage Company Active Directory for authentication. Hence Windows environment configuration from the perspective of authentication is very straight forward. However challenges starts in UNIX environment. If SAS environment is configured for default UNIX authentication then user has to maintain multiple credentials as User Desktop and SAS UNIX Authentication. In order to avoid this situation companies generally utilizes any of below mentioned three methods

1. Configure Microsoft Active Directory or LDAP Authentication for SAS Metadata. In additional to that configure Group login for spawning process on UNIX environment using Tokenized method. Challenge for this method is that UNIX security can be implemented only at group level. Another challenge becomes is the governance of data as all data will be written under group id. However by this architecture users need to not maintain multiple credentials. This method is specially used only in case only local UNIX authentication is used.
2. Configure LDAP authentication for all UNIX servers. In this scenario SAS Metadata will be configured for default UNIX authentication, which in-directly provides authentication against LDAP. The LDAP system may provide authentication as well as UNIX group maintenance service. In this scenario mostly users maintain two credentials as mentioned above. However if LDAP and AD are synched then it can be avoided, which also can be seen as another solution. Identity management solutions and other approaches can be considered to keep the credentials and groups in sync across directory servers.

PAM solution is well recommended solution. There are various tools available in market, which can be placed before Microsoft Active directory, and they provide seamless authentication & authorization for UNIX boxes. Such kind of implementation will help big way to users, as they are not required to maintain multiple credentials now.

4.5 SAS Environment Security – SAS Metadata Server Security

SAS Metadata is centralized piece for SAS environment. SAS Metadata maintains all authorization for users for all SAS client including Web. Below diagram represent most generic configuration of Application Server in SAS Metadata environment. As per company requirements multiple Server Contexts can be created. Multiple Application Server contexts make more sense when it is required that we do separate configuration parameters between different divisions and groups. Every Application Server context can have separate default Macros and Formats defined in environment. We can maintain generic macro with same name but different business logic using different application server context. However if architect feels such kind of need do not exist in environment then single application server context also can work.



Figure 3: SAS Logical Application Server Configuration in Metadata

Generally you'll find various SAS implementations, which are decoupled between SAS Metadata and UNIX security. Hence if a security change happens in SAS Metadata then that is not reflected on UNIX or vice versa.

In order to avoid such situation it's recommended that we design our security environment in such a manner that it becomes centralized and both Authorizations (UNIX, Metadata) are maintained from same place. In UNIX security section we already discussed that how we can utilize company's Active directory on UNIX. SAS also provided some

SAS Macro by which we can ensure that SAS Metadata also sync with Authentication/Authorization Provider for users and groups. These macros will take care of user addition, deletion, addition of group, removal of group, defining new group etc. if we utilized all these feature/concept together than we can get a nicely coupled architecture.

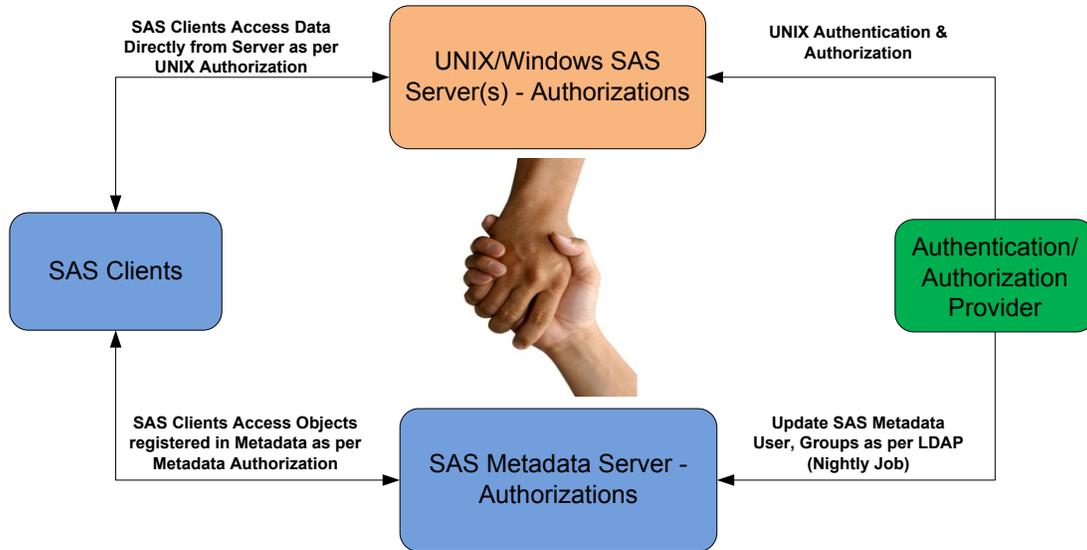


Figure 4: Centralized Authorization for SAS Environment

Please refer guide "SAS 9.3 Intelligence Platform: Security Administration Guide" to get more details around sync of Metadata and Authentication provider for authorization.

Below Mentioned are also some more general recommendations for SAS Metadata environment security

1. It is also recommended that as per business requirement SAS Metadata folders are created to maintain SAS Metadata object into it. In order to maintain security, group based Access Control Templates can be defined. Further using these ACTs as per user/group role different privileges can be provided on the folder shows in this picture. Such kind of segregation helps in maintaining SAS contents properly and even reduce the maintenance effort. Some administrators do recommend that direct permission can be defined using groups on folder. However my recommendation will be to utilized Access Control Template and avoid utilizing ACEs on folders for permissions.



2. It is recommended that every object in SAS metadata has to be secured. It has been commonly noticed that some of the metadata objects created over the period of time are not secured. Some of the examples are like AUTHDOMAIN, Users Identity, ACTs, and Metadata Groups etc.
3. Another recommendation is to avoid giving write privileges on SAS Application Server Context. Only SAS Enterprise Miner users are exception to it. OLAP Server developer and SAS Stored Process developer also do not need to have write privileges on SAS Application Server context. In case of Stored Process Server if you define target code location "SP Source Directory" for user then provide write privileges only on that Metadata object. By doing so, need to provide write privileges on application server goes away.
4. Another recommendation is that we should avoid modifying Default Access Control Template rather utilize extra permission on objects to secure environment.

4.6 SAS Environment Security – Interface Security

As we discussed earlier that interfacing system also play major role in term of enhancing SAS user experience hence we need to ensure that interfacing systems also configured in a matured manner to meet all requirement. This section will define the aspect of security.

Database/DW Security

An organization may have many databases/data-warehouse with which SAS environment interacts. There may be many users who would be interacting with more than one database. In such situation if database are not authenticated with centralize mechanism then users has to maintain multiple credential in order to interact with these database from SAS environment. Attempts should be made to consolidate the authentication provider of the databases.

User Login Object		
Authentication Domain	UserID	Password
DefaultAuth	Userid	
web	userid	
TeradataAuth	userid	*****
oraAuth	userid	*****
db2Auth	userid	*****

Figure 5: Account Tab Configuration for Individual User

However, this does not eliminate the need to maintain the credentials in SAS Metadata. Hence, somewhere need comes that how we can ensure that single sign on can be established. Identity Management solutions can be considered to help here by updating the user's password in SAS metadata when the password in authentication provider changes.

Some company goes for faceless/group id solution to interact with database/data-warehouse as depicted in below diagram. However this concept works well for production web Report or Jobs otherwise for Adhoc analysis such kind of arrangement is not good especially from DBA perspective. These faceless Ids even pose challenge to maintain security at user level in database. Below are some of the configuration diagrams that represents that how above-mentioned things will be achieved in SAS environment.

Group Login Object		
Authentication Domain	UserID	Password
TeradataAuth	Faceless_id	*****
oraAuth	Faceless_id	*****
db2Auth	Faceless_id	*****

Figure 6: Account Tab Configuration for Group User

4.7 SAS Environment Security – SAS Web Environment Security

SAS EBI implementation for web environment also can be configured in many ways for authentication however authorization is taken care by SAS Metadata. Default configuration of SAS Web also authenticates client via the SAS Metadata authentication mechanism defined. However this authentication can be changed to make it independent of SAS Metadata. Such kind of authentication conversion is called SAS Web Authentication. In order to achieve custom Web Authentication certain configuration file of SAS web installation will be updated as well Application Server property will be updated.

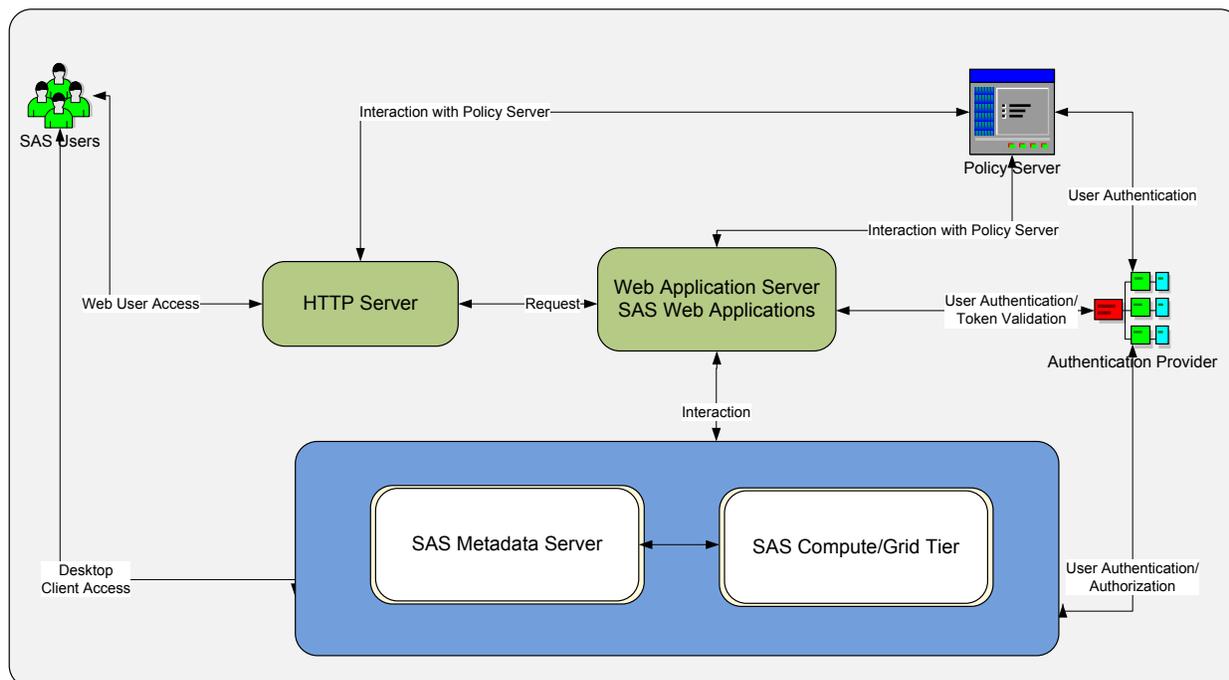


Figure 7: SAS Web Environment Interaction

It is recommended that SAS Web environment be configured to use SAS Web Authentication. In order to enhance customer experience Single Sign on can be implemented. Most of the enterprises do have standard SSO implementation for Web application. In order to achieve SSO there are various mechanism defined according to the application used. Such kind of implementation even helps in integrating other Web Application with SAS web Application and vice versa. Please find some of the benefits of Integration at <http://support.sas.com/resources/papers/proceedings11/042-2011.pdf>

Note: There are documentations available to distribute the load between different JVMs and various application Servers. In order to get latest document please interact with SAS Tech to implement solution as per recommendation.

4.8 SAS Environment Scalability – CPU, Memory, IO

As we know requirements do change very fast in current technology era. Requirements around data processing are also increasing day by day. Hence there is no doubt that a better SAS implementation will be the one, which can be scaled up to meet future requirements. It is always recommended that SAS environment should be configured in a manner that it can be scaled up to meet future business needs. Hence procuring hardware where extra resources can be added is always recommended.

SAS already propose Grid environment, which provides capability to horizontally scale environment. It is recommended to consider SAS grid architecture to meet all scalable requirements that may come. One may argue that if company does not see the need of having much more resources in coming years then why to consider configuring SAS environment using grid. This is a great point. However point to be noted that grid environment can

be configured with whatever hardware available hence there is no extra cost involved (Except LSF License if it's not already procured). However it will provide flexibility in future to add more resources. Also SAS Grid environment comes with free software called RTM, which helps in monitoring the resource utilized by SAS jobs.

One of the most critical factors in SAS environment remains SAN and throughput around it. I've seen so many environments and have noticed this always remains a biggest challenge. Whatever SAN you add if right governance policies are not applied then it will keep growing. It is required that business and tech leadership recognize this fact and aligned to reduce need of SAN addition. Even though SAN can be added to system but it will impact throughput received by users and over the period of time it will start impacting all users queries in system.

Hence system has to be designed in a manner that future need of adding more SAN can be accommodated and even throughput requirement also should be well documented. Considering solutions are more driven from the kind of storage you have hence I can not put forward just one solution here.

4.9 SAS Environment Backup & Restore

As per IT standard and for business continuity in all scenario a good backup and restore strategy should be implemented. SAS environment do have many components, which to be backed up and there may be different restore requirement as per the criticality of contents. Some of the contents, which should be backed up

- ✓ SAS Metadata (Mandatory)
- ✓ SAS Content Server (Mandatory)
- ✓ SAS Installation Folder (Mandatory)
- ✓ SAS Configuration Folder (Mandatory)
- ✓ SAS Datasets as per Business need (as per business and IT alignment)

Above mentioned backup can be broadly divided into two categories as Standard SAS Backup and SAS Environment File System Backup. It is to be considered that SAS Metadata and SAS Content Server backup have to be taken together otherwise things can be out of sync. Hence recommendation is that after pausing Metadata server, bring down SAS Content Application also and then go ahead with SAS Metadata and SAS Content Server backup. This will ensure right snapshot of backup.

In order to backup file system generally companies do have their own defined mechanism for backup like traditional Tape backup using LAN or LAN Free, Flash Copy, Separate Network for Backup etc. Some of these methods can be costly. This exercise has to be completed considering Company Standard, Size of data to backup, Cost, Backup and Restore Time etc.

5. SAS ENVIRONMENT INSTALLATION & CONFIGURATION

This section talks about the installation phase of SAS environment. Input for this section will be output of design document. Installation has to be carried out as per the architecture designed. In design document we should have defined that how Development, Test and Production environment will look like. Even we would have defined that how & where Metadata, Compute, Mid-tier will be installed. According to that SAS Institute contact person would have provided you required plan file for installation. If all required configuration are done on Windows/UNIX server with firewall rules then installation can be started. Steps more or less will remain same for new installation or upgrading existing installation. However in case of upgrading existing installation always remember to keep required backup to restore.



Figure 8: High Level Installation & Configuration Steps

6. SAS ENVIRONMENT MONITORING, GOVERNANCE & MAINTENANCE

After SAS environment launch, it is must that we have a strong SAS environment monitoring and governance plan. Missing of strong monitoring can lead to lot of outage/issues in systems. Hence strong plan around monitoring is must. Considering the nature of SAS, all SAS environment always do have strong need of huge SAN for data processing. Hence good governance plan around SAS data also need to be documented. Business involvement is must to define a strong governance plan. Governance plan will be enhanced further as system usage changes.

6.1 SAS Environment Monitoring

In order to maintain a stable SAS environment it is very much required that we do have right monitoring tools implemented/developed for SAS environment. Failure to do so will result in bad user experience even though there is a excellent SAS Architecture implementation.

SAS also provides its Audit & Performance package, which is a very good tool to have. However this is having one-day lag in order to understand the system behavior considering all collected data in day time will be processed in night. Hence we have to implement good real time monitoring tools with some additional tool to meet custom requirement of enterprise monitoring. With my personal experience I would highly recommend to monitor below area on real time basis & scheduling some scripts

- SAS Work Area monitoring
- SAS UTILLOC Monitoring
- SAS Process Monitoring
- Infrastructure Resource Utilization (CPU, Memory, IO, Run Queue, Wait Queue) and Alert around it
- SAN Alert
- Heap Area Monitoring
- Kill Orphan Process or Long Running process
- Clean of Work and UTIL area
- Timely SAS Metadata and Content Server Backup
- SAN Utilization and related Alerts

It will be advisable to collect all above-mentioned data also for long-term analysis. Administrator has to segregate that which tool will be collecting what kind of data.

6.2 SAS Environment Governance

Continuous engagement with business will help in limiting the usage of SAN on server. It is also recommended that we regularly archive or delete data, which is old and created by users for Adhoc analysis. Hence we should define governance policy around that after how much period we'll archive/delete the data. There is one very good paper on compression also, which is available at <http://support.sas.com/resources/papers/proceedings09/065-2009.pdf>

It is definitely recommended that data created by users should be separate from project data (created by service ids). I do see below kind of script can be helpful

- 1) File System Usage report on daily basis to IT and business
- 2) Files not used from past defined period with automatic mail to users or business
- 3) Archiving and delete old files as per business alignment
- 4) Script to move files to cheap archival storage (help in reducing SAN Usage)
- 5) Trending of File System Usage
- 6) If possible create data dictionary of user data and see if any redundant data is kept under user areas. See possibilities to create some centralized data for users to avoid such redundancy of data.

As mentioned earlier that for all SAS environment storage remains a bigger challenge. Considering that its must that we do have right policies define to maintain it. It is required that Users work in Project/Group mode and accordingly data is maintained.

6.3 SAS Environment Maintenance

This section will talk about that what standard maintenance activity to be considered for SAS environment. Some are mentioned below

- 1) Patching: SAS provide a tool by which it automatically analyze that what patches to be applied to SAS Servers and client side. UNIX patching all to be aligned.
- 2) Server Rebooting: It is recommended to reboot all UNIX Servers every quarter at least and Windows Servers every month
- 3) Service Restart: It is recommended to recycle SAS Web Service every week. However other SAS services can be recycled every month at least

- 4) License Renewal: Applying appropriate license files
- 5) Metadata Cleanup: Removing metadata content that is no longer in use and reducing the size of your Foundation or custom repositories

7. SYSTEM OPTIMIZATION & BEST PRACTICE IMPLEMENTATION

This phase will always active for all environments. After system implementation we have to promote/implement some of the best practices in the environment. For example all companies promote In-Database processing for better execution time and optimal utilization of resources.

7.1 Centralize Repository to Share Documents with Users

It is highly recommended that all SAS environment should maintain a document repository for users. This repository can be a medium to interact with users and to share best practice documents. This repository can have Best Practice to follow for coding, standards to follow, environment details, database details, Tools information, link to SAS knowledge base, SAS Trainings etc.

7.2 SAS Users Groups and Interaction

It is highly recommended to create users groups to keep dialogue on to share capabilities and addressing their concern/feedback. It is highly recommended to keep conducting cafes, training, knowledge sharing, discussion forum etc. This will ensure that technology and business are on same page, which in turn will ensure that business is getting value out of deployed SAS environment.

7.3 Promote SAS In-Database Processing

“SAS In-Database” allows processing to happen inside Database to utilize resources much more efficiently and effectively. SAS Access Engine does have limited in-built functionality to convert user query to get processed at database level. However users can explicitly also mention in code to pass query to database. SAS Analytics is also one of the extremely useful areas for business. Business/IT users, develops various models to address business problems and to do some prediction using these models. SAS In-Database functionality also provides capabilities to get such analytics done completely in database, which becomes extremely beneficial for industries.

Please refer below paper for details around in-database processing.

<http://support.sas.com/resources/papers/proceedings12/355-2012.pdf>

8. CONCLUSION

As discussed in this paper there are various ways a SAS environment can be deployed. It is required that we do careful designing of system before deploying SAS Intelligence Platform to environment. There are various building blocks (architecture components/solutions), which can be leveraged from enterprise architecture in order to deploy a strong SAS architecture.



Figure 9: Conclusion

9. ACKNOWLEDGEMENT

I would like to thank and appreciate efforts of Hitesh Sharma and Monika Singhal for reviewing and recommending changes in this paper to make it more useful for other Architects.

10. RECOMMENDED READINGS

- SAS® 9.3 Intelligence Platform: System Administration Guide

- SAS® 9.3 Intelligence Platform Security Administration Guide
- SAS(R) 9.3 Intelligence Platform: Web Application Administration Guide
- SAS® 9.3 Intelligence Platform Installation and Configuration Guide
- Grid Computing in SAS® 9.3
- <http://support.sas.com/rnd/scalability/grid/gridinstall.html>

11. CONTACT INFORMATION

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

Name: Gaurav Kumar Agrawal

E-mail: gaurav_agrawal@yahoo.com

Web: <http://www.linkedin.com/pub/gaurav-agrawal/8/227/360>

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.