Paper PP20

Migration to SAS Grid for three hundred users From preparations to production

Cedric Marchand, Steve Wong, Sebastià Barcelo, Nicolas Rouillé and Angelo Tinazzi
Cytel Inc. Geneva, Switzerland

ABSTRACT

To remain competitive and achieve growth, CROs need to keep up with technological advancements and adopt to new ways of working, which are ever more resource intensive. In line with this, Cytel moved all its SAS® computing to an SAS Grid platform with the aim of utilizing computing power and support staff more efficiently.

As with previous migrations, we have prepared ourselves thorough investigations and documentation, by ensuring that our standard SAS macro catalogs were working properly, by adopting new work habits and by replacing SAS Display Manager with SAS studio and Enterprise Guide.

We also implemented quality control scripts and QA intervention and made difficult decisions such as the operating system of the Grid server.

From our experience as SAS users, our recommendation is to review the many preparation steps we took and the unexpected issues we faced when we moved to the production environment. We will also discuss the differences between SAS Studio, SAS Enterprise Guide, and a Batch run, and their implications.

Lastly, we will explore our future plans, including the integration of new features such as GitHub and SAS Studio snippet.

INTRODUCTION

The reason we migrated to a SAS Grid was in principle the increase of the number of users and the utilization of large volume of data generate lags and latencies for all users.

The main advantages we did identify together with our Cytel IT department were:

- These latest SAS client tools give the user advanced features to be more productive.
- GRID has better file\folder security management since users do not have direct access to SAS servers.
- SAS Grid Manager provides a modern, flexible infrastructure that turbo-charges SAS performance with distributed and parallel computing techniques. All under the automatic monitoring, resource management, and orchestration of a grid controller. When you leverage a SAS grid computing environment, you can automatically distribute SAS computing tasks among multiple computers on a network under the control of SAS Grid Manager User can access only those data where they have permission:
    - GRID has parallel processing and can execute multiple programs at same time thus saves end user time.
    - GRID can be auto scale whenever required like more nodes can be added in future as well. **Multi-user load balancing**
    - GRID has better SAS user management.

OLD VS NEW SAS INFRASTRUCTURE

Previous SAS Infrastructure	The new SAS Grid SAS Infrastructure
- Environment: Azure Cloud (US East Coast) - Architect: Single Server - OS: Windows Server - SAS Version: SAS 9.4 M6 - Material description process: Display Manager and Batch run - Memory: 432GB	- Environment: Azure Cloud (US East Coast) - Architect: Multi-tier servers - Metadata Server: one server (Memory: 32GB) - Compute Tier (Master Host): one server (Memory: 64 GB) - Compute Nodes: two servers (Memory: 128GBx2) - Mid-Tier: one server (Memory: 32GB) - OS: Windows Server - SAS Version: SAS 9.4 M7

MIGRATION CONSTRAINTS

The migration of the projects to GRID was progressive. Projects at risk stay under the actual environment => *the old and the new system run in parallel for few months*.

- Working processes need to remain the same as before the migration e.g., use batch to run programs and use standard Cytel macros.
- Windows has been selected as operating system for GRID instead of Linux to avoid following issues:
  - Update file path (such as different names and forward slash and back slash)
  - Test out scripting languages such as VBS vs. Linux Shell. Even AIS scripts will not work the same on Linux.
  - Although Excel import and export do work fine, some of the LIBNAME engines are not supported on all operating systems, such as EXCEL and XLSX engines.
  - Migrate SAS catalogues. Compile macros under the new system.
- SAS Display Manager is no longer available with GRID, Statistical programmers need to get used to either SAS Studio or Enterprise Guide.
  - List the different environment variables which are working only under SAS STUDIO or EG and no more with DM. e.g., "_SASPROGRAMFILES". Cytel macros need to be updated.

PRELIMINARY TEST PHASE

Before migrating to the production environment, a test environment has been set up. Some users initially ran standard Cytel macros on dummy studies to do a gap analysis and investigate major issues and breaches in the processes.

IDENTIFIED ISSUES

Batch Run

The batch to run a program needs to be updated. Utilization of SASGSUB to submit a request to the GRID.

```
call c:\sas\config\Lev1\Applications\SASGridManagerClientUtility\9.4\sasgsub.cmd
-gridsubmitpgm "%~d1%~p1%%~nA%%~xA" -gridsasopts "(-work
'G:\SASGRIDWork\$username' -sysparm '%~d1%~p1%%~nA%%~xA' -altlog
'%LogFolder%\%%~nA.log')" -GRIDWAIT
```

- GRIDWAIT: GRID will wait until the end of the run of the program before starting a new one. This is mandatory when we must run programs with dependencies.
- ALTLOG: Log of the execution of the program must be stored in a specific folder according to Cytel folder structure. ALTLOG allows you to create a copy of the log.
- GRIDSUBMITPGM: path and name of the program to execute.
- SYSPARAM: creation of a macro variable to store the path and name of the program. This information is needed for Cytel macros.
- WORK: path to the folder where the work library will be stored. This option has been removed since we were able to change the default location.

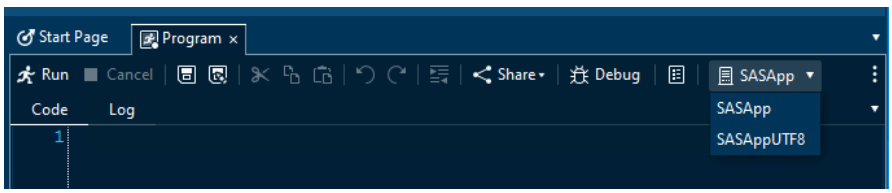
Wlatin / UTF-8 Encoding

The encoding by default is Wlatin because the data transmitted in English to the main regulators but sometimes, some extra characters (e.g., spanish, asian) required the UTF-8 encoding. We needed to find a way to configure this option in SASGSUB but also to have the possibility to switch the encoding in EG and SAS Studio.

For *batch run*, the SASGSUB was updated to configure the UTF-8 encoding by creating a specific configuration file given as an argument.

- `-config 'C:\Program Files\SASHome\SASFoundation\9.4\nls\u8\sasv9.cfg'`
- `-config 'C:\Program Files\SASHome\SASFoundation\9.4\nls\en\sasv9.cfg'`

For SAS Studio and EG, two distinct SASApp. The user has to select the one he wants to use depending on the encoding.



WORK library redirection for SASGSUB

In the initial configuration, the WORK library was pointed to the local drive of the user instead of a folder from the GRID server. Performance was severely affected by this issue. Our first resolution was to indicate as an argument the work library (`-work 'G:\SASGRIDWork\$username'`). Then IT was able to modify the configuration files to automatically point to the right folder.

UNEXPECTED ISSUES

Font issues with UTF-8 encoding

When we started to run code in UTF-8, we had the following issue in the log:

```
Output Added:
-------------
Name:        SGRender
Label:       The SGRender Procedure
Template:    figure1_
Path:        SGRender.SGRender
-------------
WARNING: The font Times New Roman is not available. Albany AMT will be used.
WARNING: The font Courier New is not available. Albany AMT will be used.
```

The solution proposed by SAS was to run the following code:

```
proc fontreg mode=all msglevel=verbose  usesashelp ;
   fontpath "C:\Windows\Fonts";
run;
```

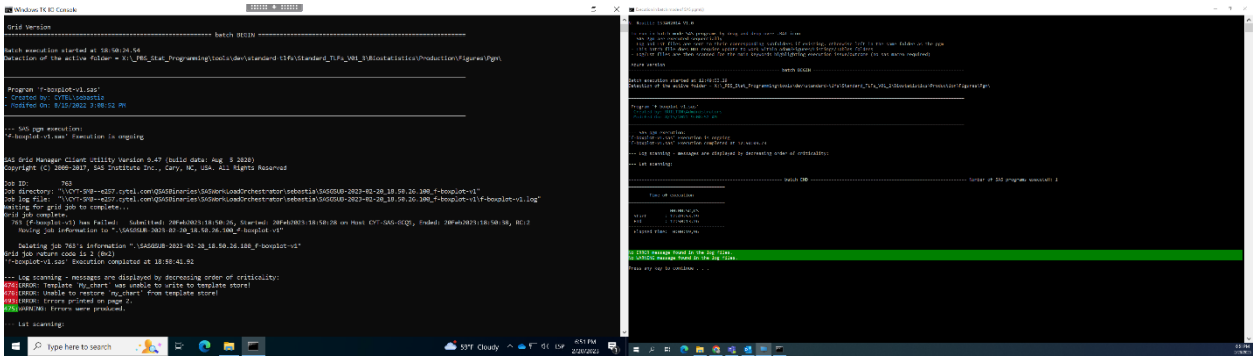
Running this code interactively, the issue disappeared from that specific session, but it reappeared again in the next session. Finally, the issue disappeared in all the sessions when IT ran the same piece of code as Admin.

ODS PATH /ODS RTF

```
ods    path    work.templat(update)    sasuser.templat(read)    sashelp.tmplmst(read)
cytels.styles(read);
ods graphics on / border=off;
ods rtf
file="&protdir\biostatistics\production\figures\output\%sysfunc(lowcase(&uniqueid.)
).rtf" style=print.chgRTF bodytitle nogtitle nogfootnote;
ods    path    work.templat(update)    sasuser.templat(read)    sashelp.tmplmst(read)
cytels.styles(read);
```

The issue we had was related to ODS RTF, it was deactivating the ODS PATH defined two rows before. The unique solution we found was to add the ODS PATH again after the call of ODS RTF.

This issue only appeared when we were running the code in batch and running programs containing PROC TEMPLATE. Obtaining the following ERRORs (left with PROC TEMPLATE, right with PROC SGPLOT):



PREPARING FOR THE MIGRATION

- Cytel Migration scripts
- To have consistency in the migration process we prepared three .vbs scripts:
  - _0_Grid_Migration.vbs
  - _0_undo-Grid_Migration.vbs
  - _Fix-Setup.vbs
- Test scripts/scenarios to validate Cytel macros on the new environment, several testers, collect of issues, resolution (detail macros)
- QA Requirements

STRESS TESTS

IT Stress Test

	IO Speed TEST		
GRID PROD	Hostname	Write MB/s	Read MB/s
	CYT-SAS-GCP1	622	2391
	CYT-SAS-NODEP1	570	4724
	CYT-SAS-NODEP2	528	4702
SAS 9.4 PROD	CYT-SAS-SRV01P	561	777
	CYT-SAS-SRV02P	358	539

SAS tool (51659 - Tool to test I/O throughput on Microsoft Windows platforms (sas.com)) to measure the time we need to write and read data on the server. This test verifies one thing: the slowness problem is not due to the server access time. It means that we can quickly send and receive information. Minimum IO of 100MB/s is recommended by SAS.

Initial User Stress Test

The purpose of the test is to run at the same time, same programs on the old and the new environment and to collect the full execution time (Durations are in format MM:SS).

PGM	Azure	GRID	PGM	Azure	GRID	PGM	Azure	GRID
t-teae (3 calls in the pgm)	00:35,9	00:29,9	t-demo (1 call, small pgm)	00:18,1	00:19,9	t-demo + t-teae	00:53,7	01:21,8
	00:34,0	00:25,2		00:18,7	00:20,0		00:58,3	00:39,5
	00:38,6	00:19,7		00:23,2	00:14,8		01:26,2	00:39,6
	00:39,4	00:24,9		00:19,4	00:14,8		01:00,8	00:56,6
	00:35,2	00:19,8		00:21,2	00:14,8		01:09,6	00:39,6
	00:46,3	00:24,8		00:21,2	00:14,7		01:02,2	00:44,6

PGM	Azure	GRID	PGM	Azure	GRID
t-teaeBig	00:40,1	00:29,9	t-demo + t-teae + t-teaeBig	01:38,5	01:19,4
	00:34,3	00:24,8		01:26,0	01:04,2
	00:38,0	00:24,9			

This test simulates the workload for one user. The result is that GRID time execution is faster than the previous environment and stable. We expect the same results when there are many more users.

STRESS TEST IN REAL CONDITIONS

All the previous tests tried to evaluate the performance of the new environment. As results are not converging, it has been decided that a full-scale test is needed.

USER TRAINING

Although Cytel processes have not changed, the new environment requires training of users. They will use a VDI instead of CITRIX and will have to use SAS Studio or Enterprise Guide which are much more user friendly than the Display Manager and offer new features which could facilitate the programming.

The auto-completion of the code and the debugging while writing the code are a real advantage. The use of shared snippets can also facilitate programming.

REFERENCES

Deep Dive with SAS® Studio into SAS® Grid Manager 9.4

CONTACT INFORMATION

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

Cedric Marchand
Cytel Inc.
Geneva, Switzerland
Cedric.marchand@cytel.com

www.cytel.com