The Top Ten SAS® Studio Tips for SAS® Grid Manager Administrators

Edoardo Riva, SAS Institute Inc.

ABSTRACT

SAS® environments are evolving in multiple directions. Modern web interfaces such as SAS® Studio are replacing the traditional SAS® Display Manager system. At the same time, distributed analytic computing, centrally managed by SAS® Grid Manager, is becoming the standard topology for many enterprises. SAS administrators are faced with the task of providing business users properly configured, tuned, and monitored applications. The tips included in this paper provide SAS administrators with best practices to centrally manage SAS Studio options and repositories, proper grid tuning, effective monitoring of user sessions, high-availability considerations and more.

INTRODUCTION

In the paper I presented last year at the SAS Global Forum 2016, I focused on SAS programmers and end users, and explained how to leverage SAS Studio to take advantage of workload balancing and parallel computing provided by SAS Grid Manager.

This year's paper is targeted at SAS Administrators, in order to help them with the daily tasks associated with configuring, monitoring, and maintaining SAS Grid Manager environments.

TIP # 1 UNDERSTAND SAS STUDIO ARCHITECTURES IN A GRID DEPLOYMENT

The simple question "How does SAS Studio integrate with a SAS Grid Manager environment?", actually has my favorite answer: it depends. That's because SAS Studio comes in multiple editions, and each has its own way to leverage a grid environment.

SAS STUDIO SINGLE-USER

SAS Studio Single-User is delivered with Base SAS® and runs on Windows operating environments. For SAS Studio Single-User, all the parts of SAS Studio are installed on the same machine (the end-user desktop or laptop), and only one user identity has access. If that instance of SAS has SAS Grid Manager licensed and LSF is installed, SAS Studio allows grid-enabled SIGNON statements to spawn remote processing on your grid.

Figure 1 shows the architecture and the interaction between processes when an end user runs a grid job from their local instance of SAS Studio Single User.

The key point to remember here is that, even if the user is working through a web interface, the user is using the locally installed SAS® Foundation to submit grid code, so all of the grid prerequisites apply to this client machine. The license on this client should include at least Base SAS, SAS CONNECT, and SAS Grid Manager. A local installation of the LSF client software is also required.
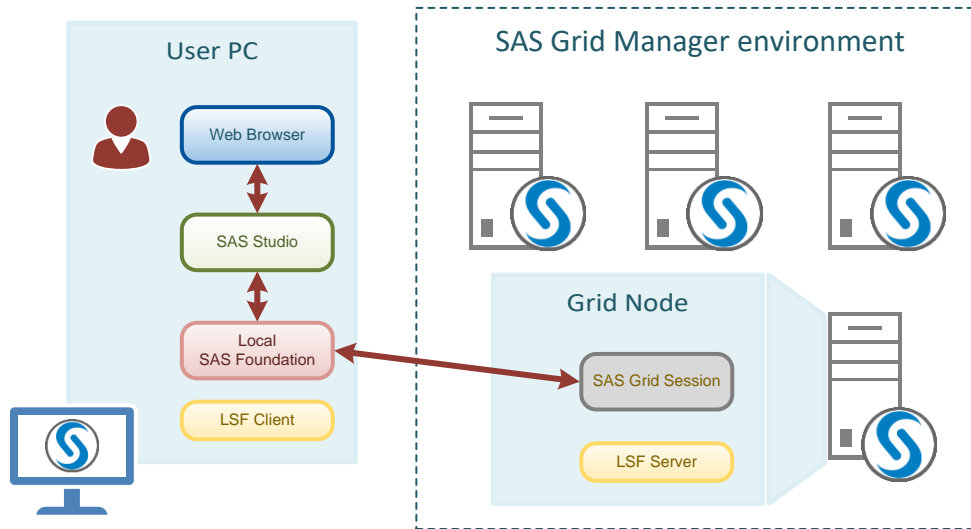
Figure 1. SAS Studio Single User Starting a Grid Session

SAS STUDIO BASIC

SAS Studio Basic is delivered with Base SAS and runs on Windows and UNIX operating environments. This edition includes the SAS® Web Application Server and the SAS object spawner. Any user who has an operating system account on the Windows or UNIX machine can log on through a web browser over the network. Though this environment does not include a full SAS enterprise platform, it is possible to use it as a client to submit code to an existing SAS Grid Manager environment. Figure 2 presents two users accessing SAS Studio Basic from their browsers in order to start SAS Grid Manager sessions on a different environment. In this example we are going multi-user, but still in a basic scenario. From an architecture point of view, the grid client requirements have to be fulfilled on the server where SAS Studio is running, because end-users' machines only have a web browser.

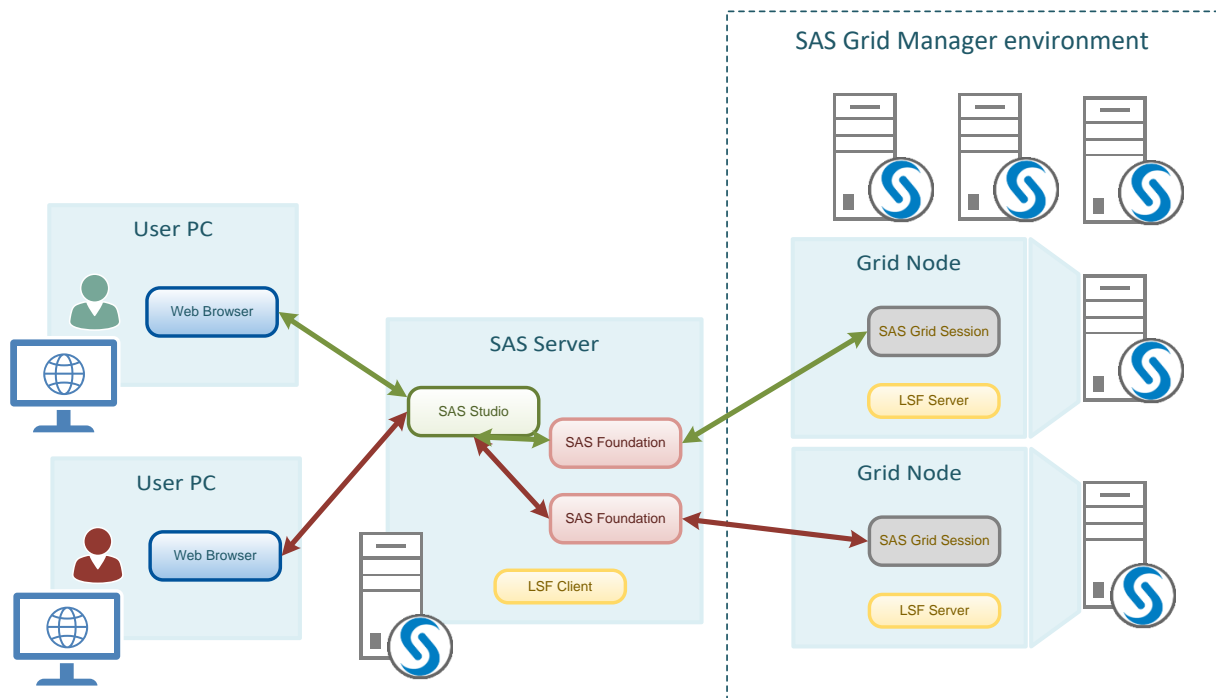Given the peculiar connection between two otherwise different environments, this is not a common scenario.

Figure 2. SAS Studio Basic Users Connecting to a Grid Environment.

SAS STUDIO MID-TIER (ENTERPRISE EDITION)

SAS Studio Mid-Tier (Enterprise Edition) is available if your site licenses SAS® Integration Technologies. This edition works with SAS® Metadata Server, SAS Web Application Server, and SAS® Web Infrastructure Platform Database. As such, it is enough to have SAS Grid Manager licensed in the same site, and it becomes possible to have a direct integration between the two. The easiest way to describe this case is to point out that SAS Studio Enterprise Edition uses workspace servers just like SAS Enterprise Guide. This means that at least three different use cases are available:

1. If the workspace server is licensed for SAS Grid Manager, it is possible to manually code grid-enabled SIGNONs (this is the same method we just described for SAS Studio Single-User and SAS Studio Basic)

2. If you configure grid load-balanced workspace servers, SAS Studio's workspace server will be grid load-balanced.

3. If you are using grid-launched workspace servers, SAS Studio's workspace server will be grid-launched.

This last use case is the recommended choice. With this configuration, we are leveraging the newest grid features, and the architecture becomes really simplified, as reflected in Figure 3. Please note that all the pictures are simplified. For the sake of clarity, some components have not been represented, such as the full SAS Web Infrastructure Platform on the middle tier or the SAS Metadata Server required by every grid deployment.
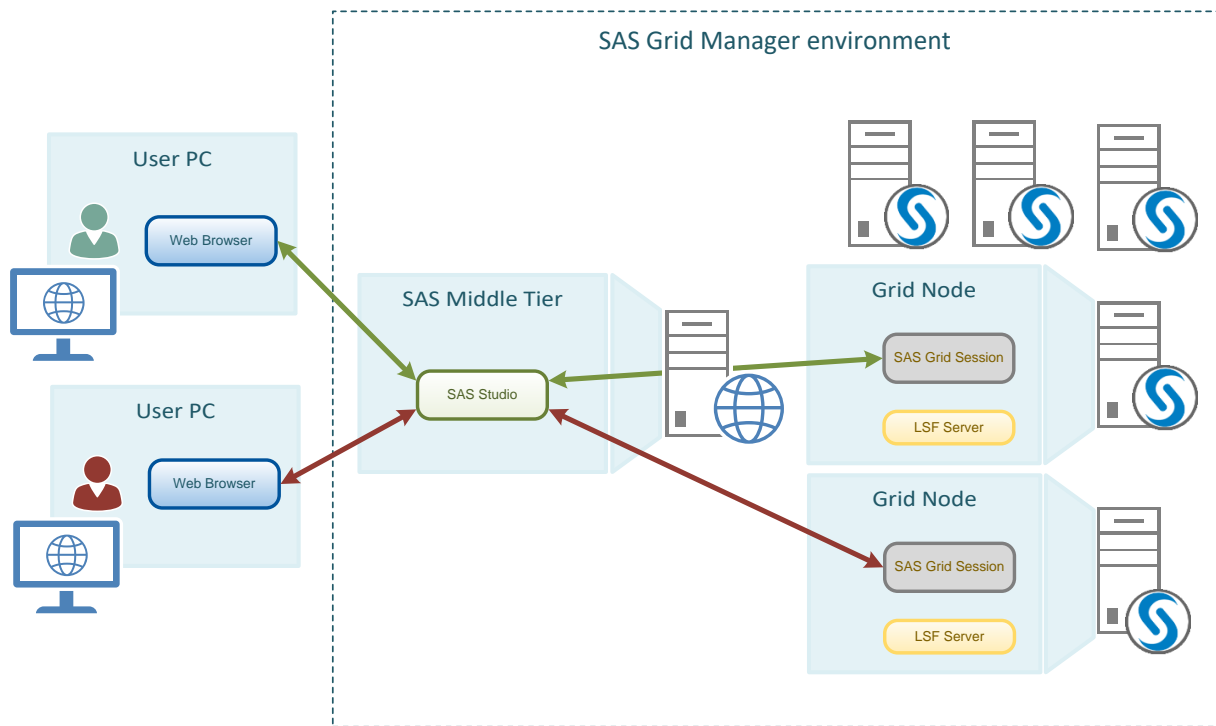
Figure 3. SAS Studio Enterprise Grid Architecture

For the rest of this paper, we will focus on SAS Studio Enterprise Edition, which is the only one that can actually be centrally managed by a SAS Administrator.

TIP # 2 CONFIGURE THE PROPER ALGORITHM FOR MULTI-USER LOAD BALANCING

SAS workspace servers are capable of performing load balancing across multiple machines. These servers can be configured to use one of the default algorithms to provide load balancing. However, with SAS Grid Manager installed, you can configure workspace servers to use SAS Grid Manager to both provide load balancing and to launch the servers. Because the grid-launched workspace servers are started by SAS Grid Manager, jobs on these servers appear as grid jobs. SAS Grid Manager submits the job to the grid provider, which determines the best server to run the job based on the policy of the queue. The grid provider uses factors such as CPU utilization, network and disk input and output, disk space, available memory, and queue limits in determining which machine to use.

This is a one-time configuration that is performed using SAS Management Console, as shown in Figure 4. Any SAS product or solution that uses workspace servers, including SAS Studio Enterprise Edition, will benefit from using SAS Grid Manager to provide load balancing.

Remember that using the grid to provide load balancing also increases overhead, so each session might take a few seconds longer to start.
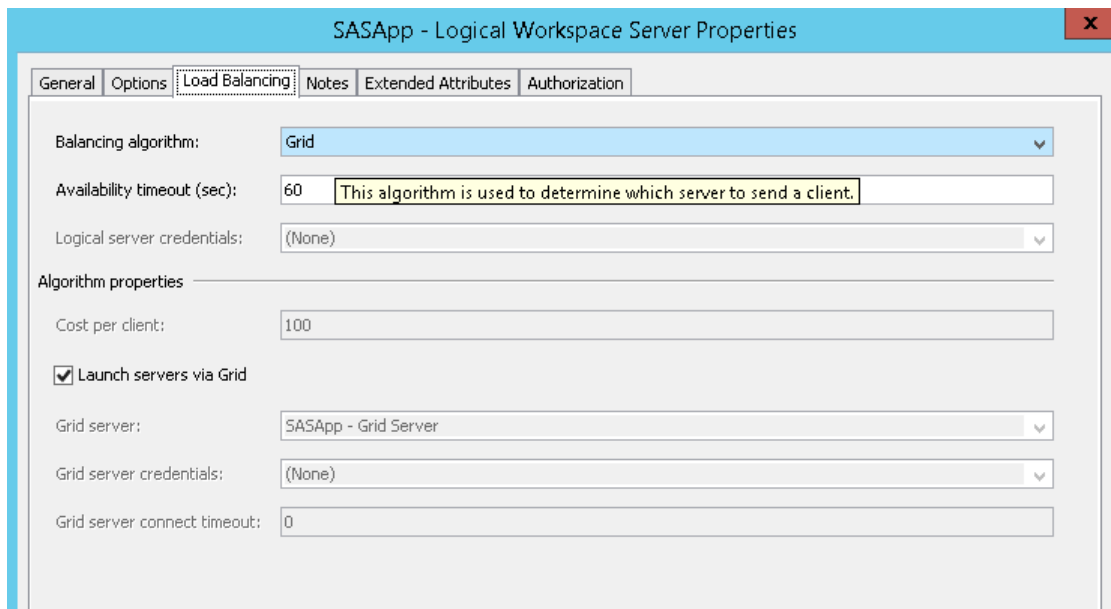
Figure 4 - Server Load Balancing Options Dialog Box in SAS Management Console

TIP # 3 CENTRALLY MANAGE USER PREFERENCES

SAS Studio has a Preferences window that enables end users to customize several options and change the behavior of different features of the software. By default, these preferences are stored under the end-user home directory on the server where the workspace server session is running (*%AppData%/SAS/SASStudio/preferences* in Windows or *~/.sasstudio/preferences* in UNIX). Does this sentence ring any alarm bells? With SAS Studio Enterprise Edition running in a grid environment, there is no such thing as "the server where the workspace server session is running!" One time it runs on one grid node, a few minutes later it runs on another one. For this reason, a preference that a user just set to a custom value might revert to its default value on the next sign-in. This issue can become worse because SAS Studio follows the same approach to store code snippets, tasks, autosave files, the WEBWORK library, and more.

Until SAS Studio 3.4, the only solution to this uncertainty was to have end users' home directories shared across all the grid nodes. SAS Studio 3.5 removes this requirement by providing administrators with a new configuration option: **webdms.studioDataParentDirectory**. This option specifies the location of SAS Studio preferences, snippets, tasks, and more. The default value is blank, which means that the behavior is the same as in previous releases. As an administrator, you can specify the option value to point to any shared location, in order to provide seamless access to all this common data from workspace server sessions running on any grid host.

THIS OPTION SOUNDS INTERESTING, HOW CAN I CHANGE ITS VALUE?

SAS Studio Administrator's Guide provides information about this topic. The property is specified in the config.properties file in the configuration directory for SAS Studio, under the SAS Web Application Server where it is deployed. Remember that when you deploy using multiple SAS Web Application Servers (which is common with many SAS solutions, and mandatory in modern grid environments), SAS Studio is deployed in SASServer2_1, not in SASServer1_1. It is also worth remembering that, in the case of clustered middle tiers, this change should be applied to every deployed instance of SAS Studio web application.

After modifying any parameter in SAS Studio config.properties, you must restart the SAS Web Application Server that is hosting SAS Studio (for example SASSerrver2_1) in order for it to pick up the change.

TIP # 4 GIVE EACH USER THEIR OWN PREFERENCES

If the previous tip seems interesting, you do not want all end users to share a single common directory, because they would override each other's settings!

It is possible to use the option introduced with the previous tip, and still have a per-user director. The key is to append `<userid>` at the end of the path. This literal string includes the minor and major brackets. SAS Studio substitutes this token with the actual user ID of the user currently logged on. For example, this configuration

```
webdms.studioDataParentDirectory=/sas94/sharedwork/studioDataParentDirectory/
<userid>
```

produces the following directory structure:

```
[sasinst@sasserver02 ~]$ ls -la
/sas94/sharedwork/studioDataParentDirectory/*/.sasstudio/

/sas94/sharedwork/studioDataParentDirectory/sasdemo/.sasstudio/:
total 28
drwxr-xr-x. 7 sasdemo sas 4096 Aug 29 17:05 .
drwxr-xr-x. 3 sasdemo sas 4096 Aug 29 17:05 ..
drwxr-xr-x. 2 sasdemo sas 4096 Aug 29 17:05 autosave
drwxr-xr-x. 2 sasdemo sas 4096 Aug 29 17:05 preferences
drwxr-xr-x. 2 sasdemo sas 4096 Aug 29 17:05 snippets
drwxr-xr-x. 2 sasdemo sas 4096 Aug 29 17:05 tasks
drwxr-xr-x. 2 sasdemo sas 4096 Aug 30 16:36 webwork

/sas94/sharedwork/studioDataParentDirectory/sasdemo2/.sasstudio/:
total 28
drwxr-xr-x. 7 sasdemo2 sas 4096 Aug 26 12:22 .
drwxr-xr-x. 3 sasdemo2 sas 4096 Aug 26 12:22 ..
drwxr-xr-x. 2 sasdemo2 sas 4096 Aug 26 12:22 autosave
drwxr-xr-x. 2 sasdemo2 sas 4096 Aug 29 16:31 preferences
drwxr-xr-x. 2 sasdemo2 sas 4096 Aug 26 12:23 snippets
drwxr-xr-x. 2 sasdemo2 sas 4096 Aug 26 12:22 tasks
drwxr-xr-x. 3 sasdemo2 sas 4096 Sep  6 15:13 webwork
```

TIP # 5 KEEP TRACK OF YOUR SETTINGS

After modifying the value of the webdms.studioDataParentDirectory option, how can you know if the change has been correctly applied? This is the same as asking "how can I know where this option is currently pointing to?" Here is a quick tip - this property influences the location of the WEBWORK directory. Simply logon to SAS Studio, open the Libraries pane, right-click on WEBWORK, and select Properties. Figure 5 shows how the highlighted path reveals the current value of the option.
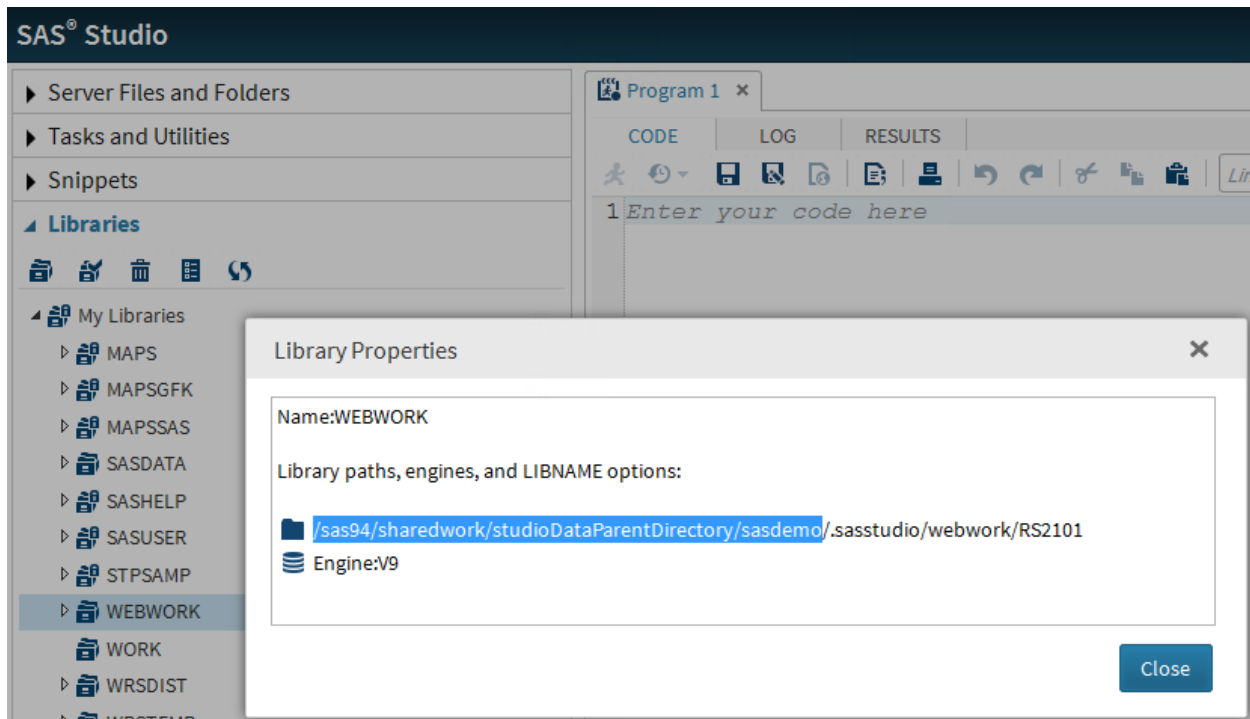
Figure 5 – Using SAS Studio to Find the Value of webdms.studioDataParentDirectory

TIP # 6 MANAGE GLOBAL SETTINGS

There are additional settings that an administrator can leverage to properly configure a multi-user environment. As you might imagine, these options deserve special considerations when SAS Studio is deployed in grid environments.

SAS Studio research and development and product management often collect customer feedback and suggestions, especially during events such as the SAS Global Forum. They received several requests for SAS Studio to provide administrators with the ability to globally set various options. The goal is to eliminate the need to have all users define them in their user preferences or elsewhere in the application. To support these requests, SAS Studio 3.5 introduced a new configuration option, **webdms.globalSettings**. This setting specifies the location of a directory containing XML files that is used to define these global options.

The procedure to configure this option is the same as that used for the **webdms.studioDataParentDirectory** property. They are both specified in the config.properties file in the configuration directory for SAS Studio, as shown in Figure 6.

```
44
45    webdms.allowRepositories=true
46    webdms.globalSettings=!SASRoot/GlobalStudioSettings
47    webdms.allowFolderShortcuts=true
48    webdms.allowBatchSubmit=true
49    webdms.maxNumActiveBatchSubmissions=5
50    webdms.maxNumActiveBatchSubmissionsSystem=5
51    webdms.batchSubmissionResultsRetentionPeriod=24
52    webdms.maxSessionTimeoutInHours=240
53    webdms.studioDataParentDirectory=
54    webdms.workspaceServer.largeTableRows=50000
55    webdms.workspaceServer.cacheTableRows=true
56    webdms.workspaceServer.nativeEngines=
57    webdms.workspaceServer.allowGetRecordCount=true
58    webdms.allowEGPOpen=false
59    webdms.maxSessionsByUser=-1
```

Figure 6 – SAS Studio config.properties file, Highlighting the globalSettings Property

By default, this option points to the directory path **!SASROOT/GlobalStudioSettings**. SASROOT translates to the directory where SAS Foundation binaries are installed, such as */opt/sas/sashome/SASFoundation/9.4* on UNIX or *C:/Program Files/SASHome/SASFoundation/9.4/* on Windows. It is possible to change the **webdms.globalSettings** property to point to any chosen directory. SAS Studio 3.6 documentation provides an additional key detail - in a multi-machine environment, the GlobalStudioSettings directory must be on the machine that hosts the workspace servers used by SAS Studio. We know that, in grid environments, this means that this location should be on shared storage accessible by every node.

TIP # 7 CONFIGURE GLOBAL FOLDER SHORTCUTS AND REPOSITORIES

By nature, grid environments are shared across multiple users. As such, it is better to provide them with central locations to share files, tasks and code snippets.

CONFIGURING GLOBAL FOLDER SHORTCUTS
In SAS Studio, end users can create folder shortcuts from the **Files and Folders** section in the navigation pane. An administrator might want to create global shortcuts for all the users, so that each user does not have to create these shortcuts manually. This is achieved by creating a file called shortcuts.xml in the location that is specified by **webdms.globalSettings**, as detailed in *SAS Studio 3.6: Administrator's Guide.*

The following example creates the shortcut shown in Figure 7.

```
<?xml version="1.0" encoding="UTF-8"?>
<Shortcuts>
<Shortcut type="disk" name="A shared location" dir="/sas94/shared/"/>
</Shortcuts>
```
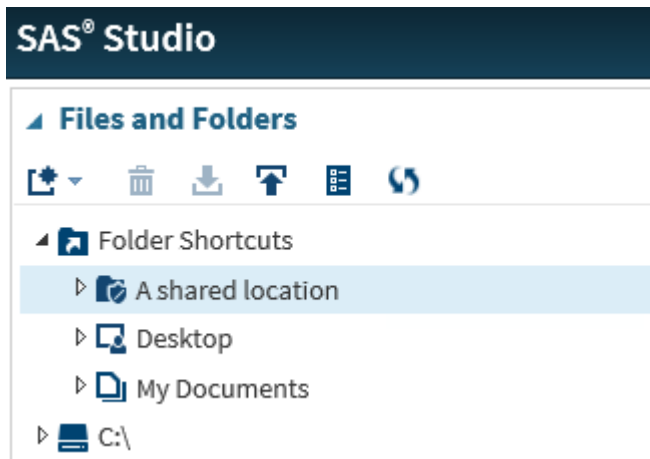
Figure 7 – A Global Folder Shortcut Displayed Under SAS Studio Files and Folders

CREATING GLOBAL REPOSITORIES

SAS Studio repositories are an easy way to share tasks and snippets between users. An administrator might want to configure one or multiple centralized repositories and make them available to everyone. SAS Studio users could add these repositories through their Preferences window, but it's easier for an administrator to create global repositories that are automatically available from the **Tasks and Utilities** and **Snippets** sections. As before, this is achieved by creating a file called repositories.xml in the location specified by **webdms.globalSettings**, as detailed in *SAS Studio: Developer's Guide to Repositories*.

The following example creates the link to a repository shown in Figure 8.

```
<?xml version="1.0" encoding="UTF-8"?>
<Repositories>
<Repository uri="http://my-http-server/Repository/sample"
   name="Example Repository" />
</Repositories>
```
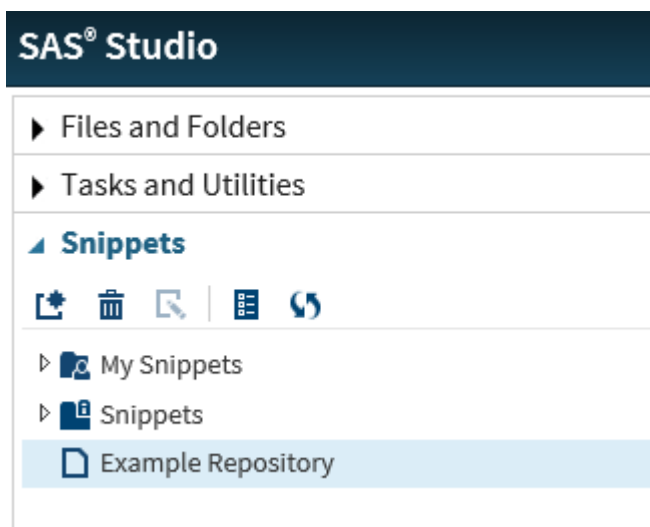

Figure 8 – A Global Repository Displayed Under SAS Studio Snippets

While this option provides end-users with repositories that are defined by an administrator, the users can also add user-defined repositories that point to locations of their own choice. In regulated environments this might not be a permitted option. To prevent end-users from adding any custom repositories to the

ones already set by the administrator, set the **webdms.allowRepositories** property to **false** in the *config.properties* file.

TIP # 8 TUNE SAS GRID MANAGER AND SAS STUDIO

SAS Studio interacts with back-end processes and uses grid resources in a similar way to SAS® Enterprise Guide®. As such, the required grid configuration is similar between the two applications. This configuration is described in great detail in the SAS Global Forum 2014 paper "Effective Usage of SAS® Enterprise Guide® in a SAS® 9.4 Grid Manager Environment", and the same considerations apply. In detail, grid administrators should perform these tasks:

- Tune LSF batch parameters such as MBD_SLEEP_TIME, SDB_SLEEP_TIME, MBD_REFRESH_TIME, JOB_SCHEDULING_INTERVAL.

- Create a dedicated SASStudio queue in LSF.

- Limit the number of jobs that each user can submit to the SASStudio queue.

- Create a dedicated SAS Studio grid options set in metadata, to associate SAS Studio with its queue.

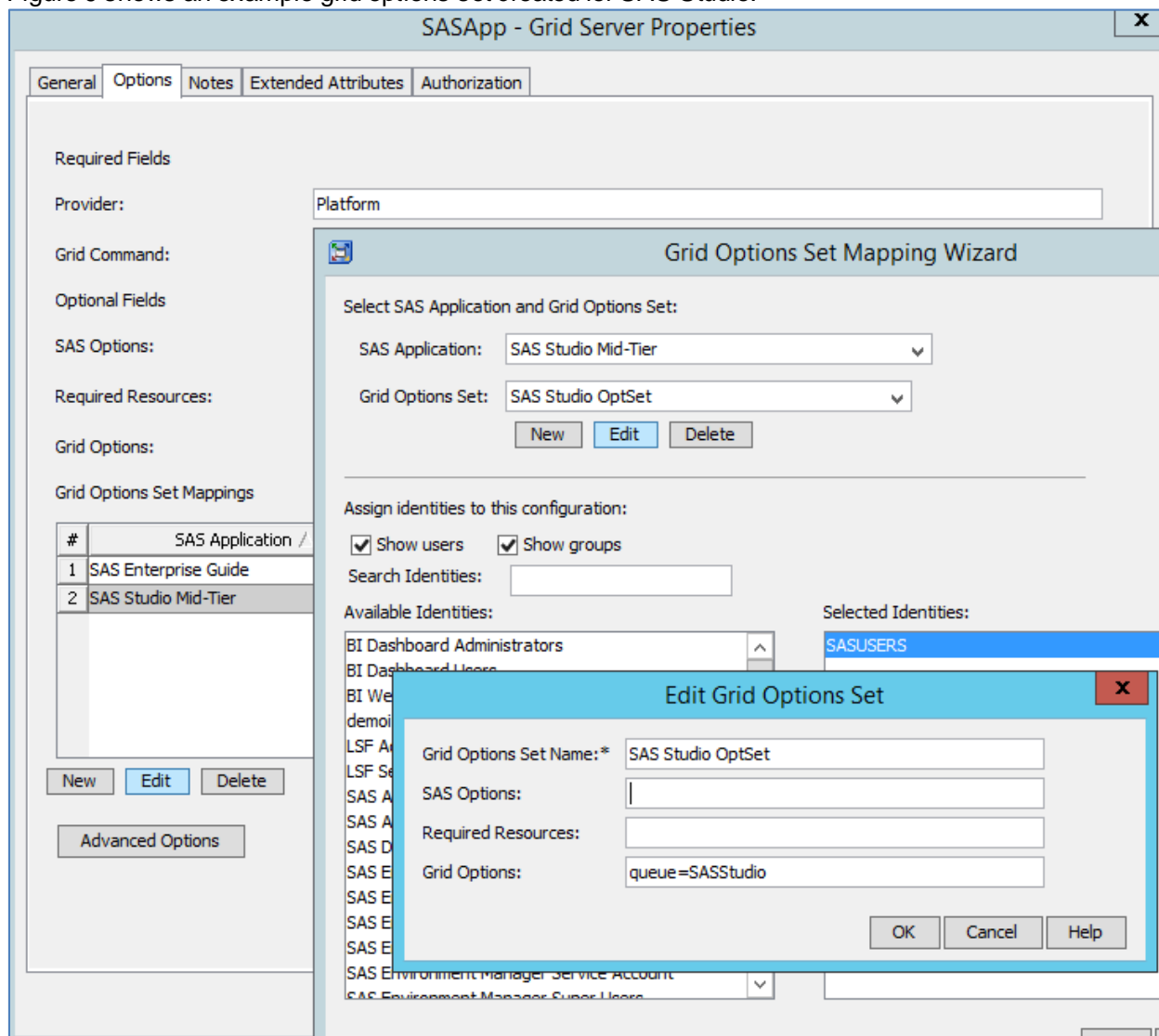Figure 9 shows an example grid options set created for SAS Studio.

Figure 9 – SAS Studio Grid Options Set in Metadata

Administrators should be aware of the different SAS Studio usage patterns and the configuration parameters that influence the number of workspace servers started on the back-end hosts for each end-user session. This information is important in determining the maximum number of job slots to assign to each user of the SASStudio queue.

By default, as soon as an end user logs in SAS Studio, two workspace server sessions are created. One session is used to submit user code, while the other performs internal tasks such as listing files and libraries to populate the web interface. These two workspace servers remain open, even if nothing is running, until the user logs out from the web application. However, they are used only for the default execution mode.

If a user switches to SAS Studio Visual Programmer perspective and selects to run the process flow in parallel mode, up to three additional SAS session are started, for a total of five. Once the process flow is finished, if there is no further activity for 30 seconds, the three additional SAS processes terminate in order to release resources.

Finally, starting with SAS Studio 3.5, end users can submit code in background. Each background submission starts a new, dedicated workspace server that remains active until code execution ends, and then automatically terminates.

All these additional sessions can overload a grid environment, if it is not properly managed.

An administrator can use the **webdms.maxParallelWorkspaces** configuration property to specify the maximum number of workspaces that can be used when SAS Studio is running in parallel mode. The default value is 3. The maximum value is 8.

Background submissions can be controlled with two parameters.
webdms.maxNumActiveBatchSubmissions specifies the maximum number of active background jobs for the current SAS Studio user. The default value for SAS Studio Enterprise Edition is 3.
webdms.maxNumActiveBatchSubmissionsSystem specifies the maximum number of active background jobs that can be submitted for a given instance of SAS Studio, across all users. The default value for SAS Studio Enterprise Edition is 24. Note that this parameter is "per instance", which means that, in case of middle tier clustering, each SAS Studio instance could start up to 24 sessions running in background.
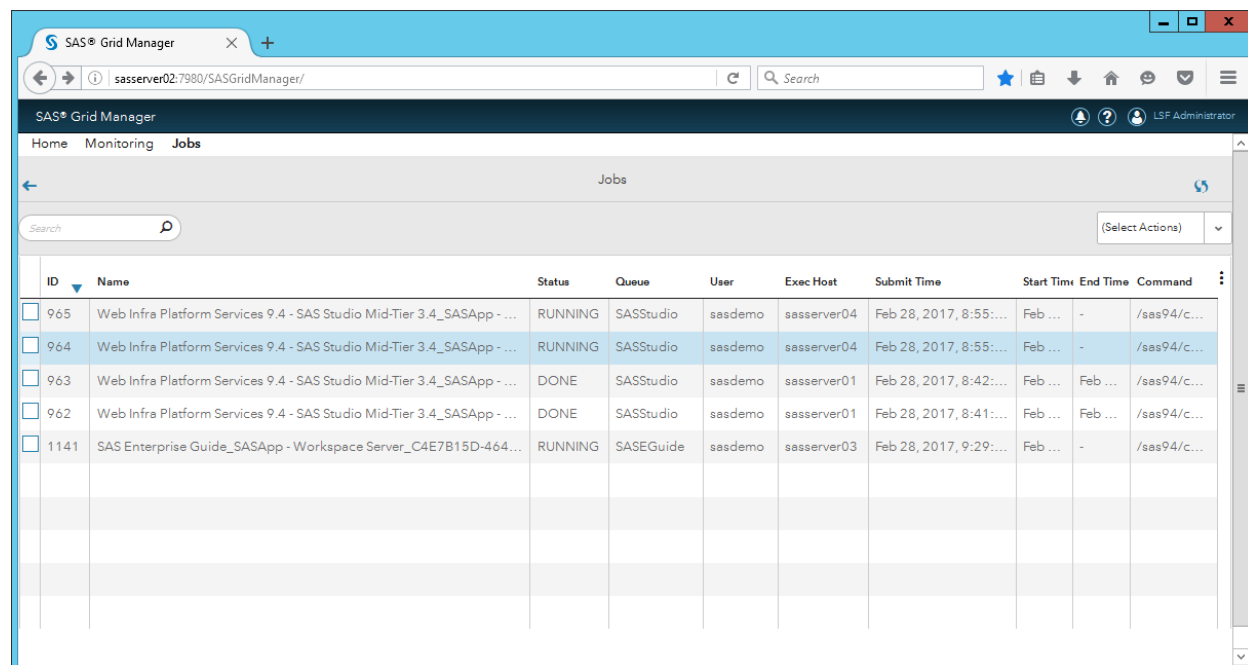
TIP # 9 MONITOR AND MANAGE USER SESSIONS

SAS Environment Manager provides a management module that enables administrators and end-users to monitor a SAS grid cluster. As a grid administrator you can see and manage the whole grid, including jobs submitted by any user, while end users can take advantage of this module to monitor their own sessions. When using SAS Grid Manager for Hadoop, a similar monitoring capability is provided by Hadoop, using the YARN ResourceManager web user interface.

Figure 10 shows the Jobs Monitoring window. SAS Studio jobs can be identified looking for names that start with "Web Infra Platform Services 9.4 - SAS Studio Mid-Tier". It possible to perform management actions on a specific job, such as killing it, by selecting the desired job and choosing the action from the drop-down field in the top right side of the page, as shown in Figure 11.

Additional job details are displayed by simply clicking on a job line. Figure 12 shows the Job History page for a SAS Studio job.

Figure 10 – Monitoring Job Execution Using the SAS Environment Manager Grid Module
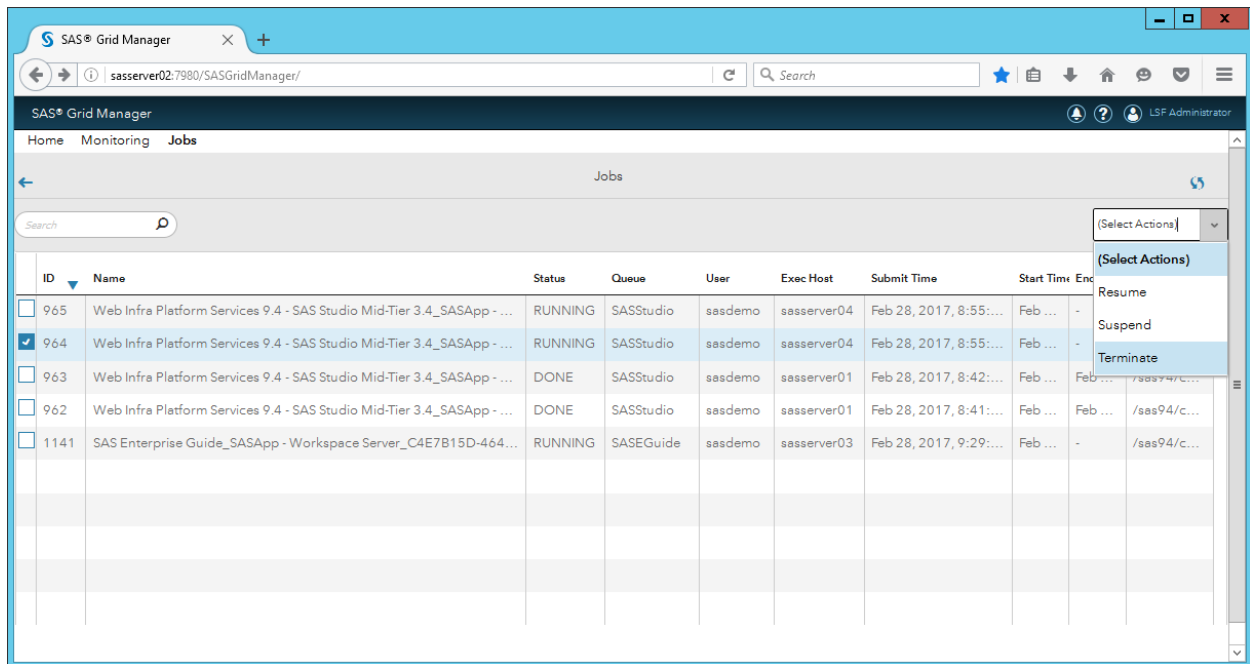
Figure 11 – Terminating a Job Using the SAS Environment Manager Grid Module
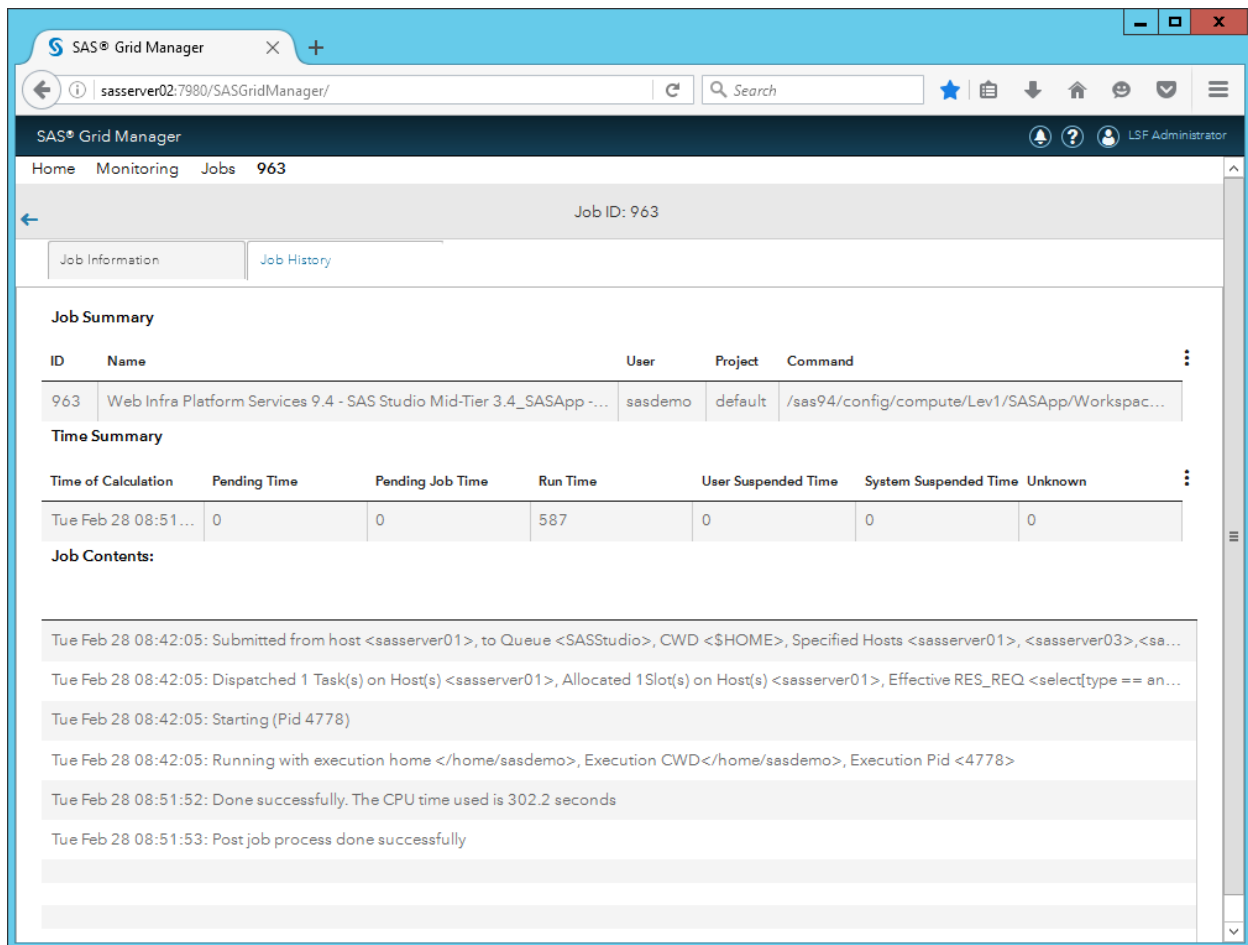
Figure 12 – The Job History Page of a SAS Studio Job

TIP #10 CONFIGURE SAS STUDIO FOR HIGH AVAILABILITY

SAS Grid Manager administrators are familiar with using grid capabilities to ensure high availability of back-end services, such as the object spawner, the SAS Metadata Server or the SAS Web Infrastructure Platform Data Server. Monitoring these services and configuring proper recovery policies is important to ensure the availability of the SAS processing engine. However, SAS Studio is a web application that is running inside an instance of the SAS Web Application Server (SASServer2 by default). SAS Studio also depends on the services provided by the SAS Web Infrastructure Platform web applications that are running in SASServer1. As such, the proper mechanism to ensure high availability is to create a clustered middle tier deployment, preferably spreading it across multiple hosts. The correct procedure to perform this deployment is described in the "High-Availability Features in the Middle Tier" section of *SAS® 9.4 Intelligence Platform: Middle-Tier Administration Guide*.

CONCLUSION

In this paper, we have presented useful best practices to administer SAS Studio in SAS Grid Manager environments.

These tips will help SAS administrators enable SAS Studio users to benefit from distributed and parallel computing techniques provided by the automatic monitoring, resource management, and orchestration of SAS Grid Manager.

REFERENCES

14

- Riva, Edoardo. 2016. "Deep Dive with SAS® Studio into SAS® Grid Manager 9.4". *Proceedings of the SAS Global 2016 Conference.* Cary, NC: SAS Institute Inc.

  Available http://support.sas.com/resources/papers/proceedings16/SAS6422-2016.pdf

- Riva, Edoardo. "SAS Studio tips for SAS Grid Manager Administrators: Where are my preferences?"

  Available http://blogs.sas.com/content/sgf/2016/09/21/sas-studio-tips-for-sas-grid-manager-administrators-where-are-my-preferences/. Last modified September 21, 2016. Accessed February 28, 2017.

- Riva, Edoardo. "More SAS Studio Tips for SAS Grid Manager Administrators: Global Settings"

  Available at http://blogs.sas.com/content/sgf/2016/12/27/more-sas-studio-tips-for-sas-grid-manager-administrators-global-settings/. Last modified December 27, 2016. Accessed February 28, 2017.

- Riva, Edoardo. 2014. "Effective Usage of SAS® Enterprise Guide® in a SAS® 9.4 Grid Manager Environment." *Proceedings of the SAS Global 2014 Conference.* Cary, NC: SAS Institute Inc.

  Available http://support.sas.com/resources/papers/proceedings14/SAS375-2014.pdf

- SAS Institute Inc. 2016. *SAS 9.4 Intelligence Platform: Middle-Tier Administration Guide.* Cary NC: SAS Institute Inc.

  Available http://support.sas.com/documentation/onlinedoc/intellplatform/tabs/admin94.html

RECOMMENDED READING

- SAS Institute Inc. 2016. *Grid Computing in SAS® 9.4*. Cary NC: SAS Institute Inc.
  Available http://support.sas.com/documentation/onlinedoc/gridmgr/index.html.

- SAS Institute Inc. 2016. *SAS Studio: User's Guide*. Cary NC: SAS Institute Inc.
  Available at http://support.sas.com/documentation/onlinedoc/sasstudio/index.html.

- SAS Institute Inc. 2016. *SAS Studio: Administrator's Guide*. Cary NC: SAS Institute Inc.
  Available at http://support.sas.com/documentation/onlinedoc/sasstudio/index.html.

- SAS Institute Inc. 2016. *SAS Studio: Administrator's Guide*. Cary NC: SAS Institute Inc.
  Available at http://support.sas.com/documentation/onlinedoc/sasstudio/index.html.

- SAS Institute Inc. 2016. *SAS Studio: Developer's Guide to Repositories*. Cary NC: SAS Institute Inc.
  Available at http://support.sas.com/documentation/onlinedoc/sasstudio/index.html.

- *SAS Grid Computing Overview Video.* YouTube.
  Available https://www.youtube.com/watch?v=BIK8JzDsSQg. Accessed February 28, 2017.

- *Working in SAS Studio.* YouTube
  Available https://www.youtube.com/watch?v=usnucvpnGLM. Accessed February 28, 2017.

CONTACT INFORMATION

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

Edoardo Riva
100 SAS Campus Drive
Cary, NC 27513
SAS Institute Inc.

+1 919 531 7293
edoardo.riva@sas.com
http://www.sas.com