

Hidden Treasures in SAS® Scheduler Land

Randolph S. Williams, SAS Institute Inc.

ABSTRACT

As a user of the Schedule Manager plug-in to SAS® Management Console, you might not realize all the features available for managing flows. This paper shows you how to unlock some hidden features and learn about new features designed to help with huge flows. This paper also provides an overview of the various supported schedulers and their niche. Come learn how to get more out of SAS® scheduling.

INTRODUCTION

The Schedule Manager plug-in to SAS® Management Console provides a range of features to help you create, manage, and schedule your flow definitions in a scheduler-neutral environment. A flow (also called a deployed flow or job flow) is a set of jobs, dependencies, and triggers that are suitable to submit to underlying scheduling servers. As new challenges arise, new features are added. However, not all the features are visible by default. In addition, SAS supports several types of schedulers. Determining which type of scheduler to use depends on multiple factors.

This paper discusses hidden features and how to unlock them, recent features that help manage huge flows, and tips on determining which scheduler type to use.

UNLOCKING HIDDEN FEATURES

This root node to the Schedule Manager plug-in has two undocumented actions: Import Flow and Validate Job Command Properties. These actions are considered experimental in nature. To turn on these actions, one needs to manually update the user profile.

Each user has a workspace file for the Schedule Manager in the user data area. This file contains user information about options and settings. The file name is app.schmgr. For example, on Windows it would be in the directory
`C:\Users\`

In the workspace file, you need to add the following lines:

```
UserSetting_ImportActivated=True  
UserSetting_ValidateCmdlineActivated=True
```

In Figure 1Error! Reference source not found., see the menu of actions for the root node.

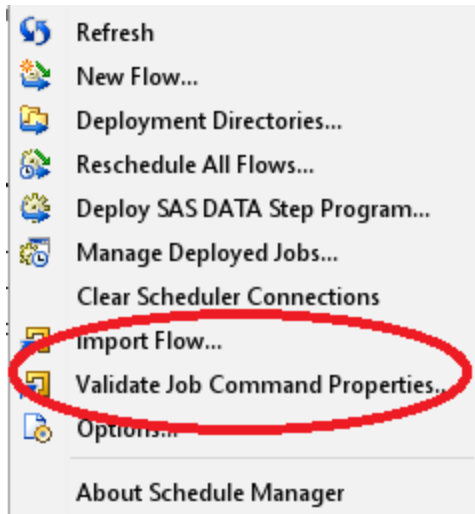


Figure 1. Actions for Root Node

IMPORT FLOW

As shown in Figure 2, the Import Flow action allows you to create a new flow or update an existing flow from one of three supported file formats. The three supported file formats are the same formats supported in the Report Flow action: XML for a selected flow, SAS data set, or CSV file. The data is assumed to be in UTF-8 encoding.

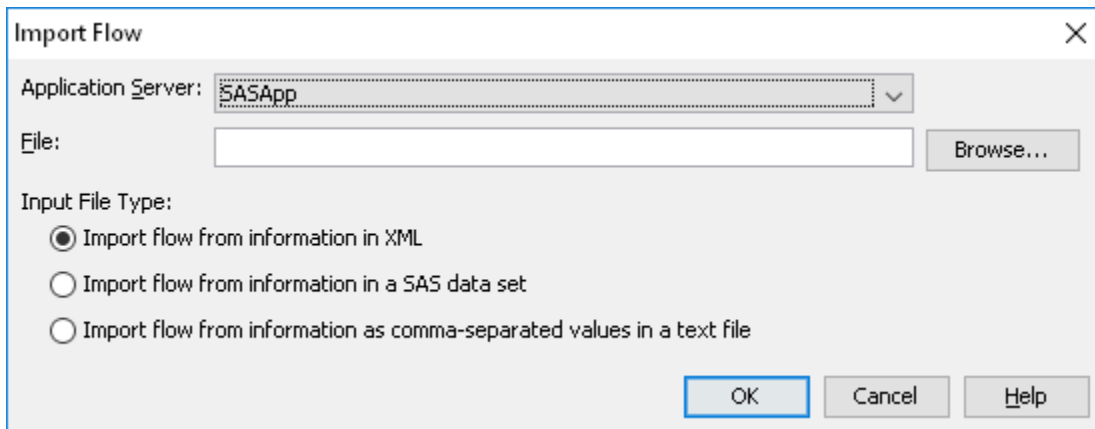


Figure 2. Import Flow

A new flow is determined when the ID for the flow is blank or omitted. If the ID is provided, **then it must match an existing flow's ID** and the flow is then updated based on the content of the file.

The easiest way to generate the input file is to first generate a report of a flow that is similar to the one you would like to create as your template. Then edit the resultant file as desired. The format of the SAS data set and the CSV file is as seen in the properties of the flow in the Dependencies tab. The columns are Predecessors, ID, Name, Type, Event, Location, Runas, and Description.

The flow XML contains one or more flow definitions. The flow definition contains major sections of the associated scheduler, triggers, jobs and subflow, and the dependencies among the jobs and subflows. The triggers and job dependencies are expressed as events. The top-level schema is as follows:

```

<?xml encoding="UTF-8"?>
<!ELEMENT xmlFlowDef (FlowDef)>
<!ATTLIST xmlFlowDef
  xmlns CDATA #FIXED ''
  Name NMTOKEN #REQUIRED
  Version CDATA #REQUIRED>

<!ELEMENT FlowDef (Scheduler,TriggerEvents,JobDef+,Dependencies+)>
<!ATTLIST FlowDef
  xmlns CDATA #FIXED ''
  Description CDATA
  Id NMTOKEN #REQUIRED
  Location CDATA
  Name NMTOKEN #REQUIRED>

<!ELEMENT Scheduler EMPTY>
<!ATTLIST Scheduler
  xmlns CDATA #FIXED ''
  ClassID CDATA #REQUIRED
  Id NMTOKEN #REQUIRED
  Name CDATA #REQUIRED>

<!ELEMENT TriggerEvents (Events)>
<!ATTLIST TriggerEvents
  xmlns CDATA #FIXED ''
  CombinationType NMTOKEN #REQUIRED>

<!ELEMENT JobDef (JobCmdLine)>
<!ATTLIST JobDef
  xmlns CDATA #FIXED ''
  Description CDATA
  Location CDATA
  Name NMTOKEN #REQUIRED
  Type NMTOKEN #REQUIRED>

<!ELEMENT Dependencies (Events)>
<!ATTLIST Dependencies
  xmlns CDATA #FIXED ''
  Description NMTOKEN
  Name CDATA #REQUIRED
  Reference NMTOKEN #REQUIRED
  Type NMTOKEN #REQUIRED>

<!ELEMENT JobCmdLine EMPTY>
<!ATTLIST JobCmdLine
  xmlns CDATA #FIXED ''
  Value CDATA #REQUIRED>

<!ELEMENT Events (Events|Event)+>
<!ATTLIST Events
  xmlns CDATA #FIXED ''
  CombinationType NMTOKEN #REQUIRED>

```

```
<!ELEMENT Event (done)>
<!ATTLIST Event
  xmlns CDATA #FIXED ''
  Description CDATA #REQUIRED
  Type NMTOKEN #REQUIRED>
```

VALIDATE JOB COMMAND PROPERTIES

Users can modify the command line of a deployed job, which creates a property. This property overrides the generation of the command line and in many cases causes issues when migrating the deployed job to other systems. Often the edit of the command line was not intended and only contains changes to whitespace characters.

The Validate Job Command Properties action allows you to remove command-line properties that match the generated command line for any deployed job in SAS metadata. Usually, the command line is generated based on the command in the associated batch server and the job arguments stored in metadata at the time the job is deployed. The user can edit the command line, which will create a command-line property with the text as provided. The command-line property for a job is always used over the generation of the command line. When the command-line property exists, it can prevent smooth migration of the metadata to a new system where aspects of the command-line change, such as the location of the batch server wrapper script. This action on the source system prior to migration or use of the partial promotion framework can help to ensure that the command line is generated as desired on the target system.

MANAGING HUGE FLOWS

The flow editor, used when the Edit Flow action is selected, can be a challenge to use with huge flows. It can be difficult to find the node you want in the flow. When there are thousands of nodes, the flow editor can take a long time to determine the best layout, which is the one that minimizes overlapping lines in the graph. As a result, the flow editor was enhanced to provide two new features. The first is the Find feature that was added in SAS® 9.4M3 to move the focus of the diagram to the desired node. The second feature is the ability to keep the layout of the flow diagram instead of relying on auto-layout to render the flow layout.

FIND NODE

The Edit Flow action provides a process flow diagram to visualize the flow dependencies. When there are lots of jobs and subflows in the diagram, it can be a challenge to find the node you are interested in viewing. In SAS 9.4M3, the Find action was added. Figure 3 shows the pop-up menu illustrating the Find command. Figure 4 shows the resulting dialog box that allows you to enter the name of the node you wish to find. The name must be the complete name. Once entered, the command searches the entire flow diagram for a job or subflow of the exact name, and it changes the focus of the diagram to that node.

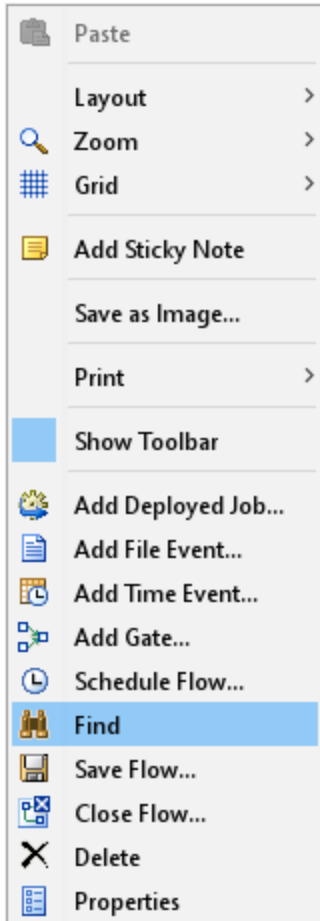


Figure 3. Find Node

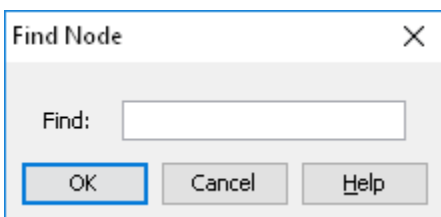


Figure 4. Find Node Name

KEEP LAYOUT

The Edit Flow action provides a process flow diagram to visualize the flow dependencies. By default, the layout of the process flow diagram is auto-layout. The underlying code automatically determines and saves the layout of the flow on open. When the flow is very large, it can take significant time to render the new layout. When the flow is huge, this process can take several minutes and even hours. Auto-layout is turned off when the flow is set to save the layout instead. This will prevent the auto-layout on save. This will also improve the drawing of the diagram on open as it just uses the saved layout rather than programmatically determining the optimal layout. To change the flow to save the layout instead of auto-layout, you right-click and select Properties. Next, go to the Attributes

tab and select the check box Save the layout of the flow in the flow editor, as shown in Figure 5.

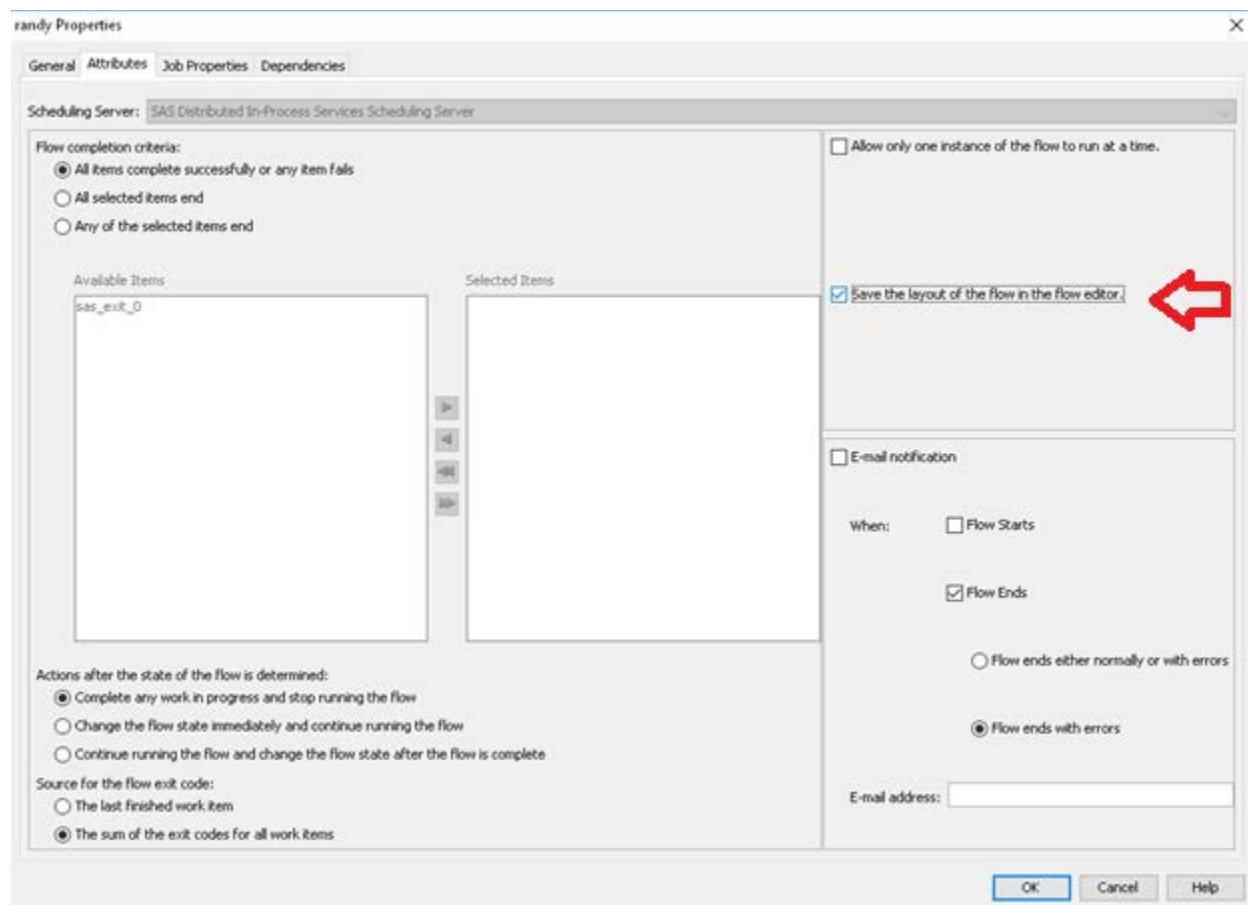


Figure 5. Keep Layout

SELECTING THE CORRECT SCHEDULER

The SAS scheduling integration supports multiple types of schedulers. Each scheduler offers **a unique feature or set of features that the other schedulers don't provide. This paper** briefly discusses each of the supported scheduler types and highlights some of the unique features that each provides.

PLATFORM PROCESS MANAGER

The Platform Process Manager is the first scheduler supported and provides enterprise-level capabilities with the most complete set of features as compared to other supported schedulers. This scheduler can impersonate a user for a job within the flow. The impersonated user can be different from the user that scheduled the flow. This scheduler supports variable substitution on the command line for jobs. This scheduler is grid-enabled and can submit jobs to the Platform LSF grid provider. This scheduler can be ordered separately, but it is included with SAS® Grid Manager for Platform, and can submit jobs directly to the associated grid of compute resources.

This scheduler is a third-party software that is OEM from IBM Platform Computing. It is designed for SAS use and distributed only by SAS. It does not depend on SAS software for managing its flow metadata or history. It provides two separate clients to help manage its flows. One is the Flow Manager for monitoring flow execution. The other is the Calendar

Editor for creating custom calendars to be used as either trigger events or job dependencies within a flow definition.

SAS JOB FLOW SCHEDULER

The SAS Job Flow Scheduler is the newest scheduler, and it was first introduced with SAS® 9.4M6. The initial release is limited to just Windows and Linux platforms. This scheduler provides an advanced level of capabilities with support for subflows, flow completion criteria, multiple triggers, complex job dependencies, parallel job execution, and more. This scheduler can be ordered separately, but it is included with SAS® Grid Manager, and can submit jobs directly to the associated grid of compute resources.

The SAS Job Flow Scheduler requires the SAS middle tier for managing its flow metadata and history. The triggering mechanism for a flow is done through operating system services such as crontab on UNIX and Windows Task Scheduler on Windows.

DISTRIBUTIVE IN-PROCESS SERVICES

The Distributive In-Process Services scheduler is designed for SAS application integration. It provides an intermediate level of capabilities. The unique aspect of this scheduler is that SAS applications can register certain types of jobs to be sent to the SAS applications for execution. This allows SAS applications to execute their own types of jobs. When it is a SAS web application, it can execute their jobs as a web user. Conversely, jobs that are sent to the operating system for execution will run as the user account that the Distributive In-Process Services process is running. This is typically the user that installed the software, and unlike the other schedulers, these jobs do not run as the user that scheduled the flow.

The Distributive In-Process Services scheduler requires the SAS middle tier for managing its flow metadata and history.

OPERATING SYSTEM SERVICES

The Operating System services is a basic scheduler with the smallest set of features. It generates a script native to the operating system, uses a triggering mechanism native to the operating system, and stores all its data in the file system. As a result, it has no runtime dependencies on any services such as the SAS middle tier and can be administered with basic operating system tools.

CONCLUSION

The SAS scheduling team is committed to improving the user experience with the Schedule Manager plug-in to SAS Management Console as well as SAS scheduling in general. New features are added to solve your business pains. Whether it is automation through advanced options, new features to solve issues, or guidance on how to use what is already available; we are here to serve you and help you bring order and simplification to your batch processing at scheduled times.

REFERENCES

Tran, Allen, Randy Williams, and Alan Wong. 2005. "SAS Scheduling: Getting the Most Out of Your Time and Resources." *Proceedings of the Thirtieth Annual SAS® Users Group International Conference*. Cary, NC: SAS Institute Inc. Available at <http://support.sas.com/resources/papers/SASScheduling.pdf>.

Williams, Randy. 2006. "Scheduling Options Within the SAS System." *Proceedings of the Thirty-First Annual SAS® Users Group International Conference*. Cary, NC: SAS Institute Inc.

RECOMMENDED READING

- *Scheduling in SAS® 9.4*

CONTACT INFORMATION

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

Randolph S. Williams
SAS Institute Inc.
randy.williams@sas.com

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.