

Using SAS® Job Monitor to Analyze SAS® Data Integration Studio Job Performance

Jeff Dyson, The Financial Risk Group

ABSTRACT

Obtaining run-time statistics and performance information from SAS® Data Integration Studio batch processes has historically been a challenge. And with growing source populations, tight batch windows, and the demand for SAS analytics and predictive model results, being aware of negative performance trends in the essential data management layer is as important as ever. With SAS® Job Monitor, the ETL (extract, transform, and load) administrator can take advantage of an integrated set of components that provides a more complete view of job load status, historical run-time trends, and other audit related information.

INTRODUCTION

SAS Job Monitor is designed to process logs from SAS Data Integration Studio or DataFlux Data Management Server. This paper is focused on processing and monitoring SAS Data Integration Studio logs in SAS Environment Manager with the SAS Job Monitor plug-in.

SAS Data Integration Studio is a client that uses metadata to create repeatable extract, transform, and load processes for data warehouses and marts. Jobs can be deployed and scheduled for batch execution. The SAS logs from the batch processes are analyzed by SAS Job Monitor.

SAS Environment Manager is a web-based administration application.

SAS Job Monitor is accessible as a SAS Environment Manager component.

SAS JOB MONITOR REQUIREMENTS

The following SAS Job Monitor components are required. The SAS Deployment Registry report can be used to verify that these SAS Job Monitor services are installed and available. If they exist, the entries will be found in the DeploymentRegistry section for mid-tier output:

- SAS Job Monitor Mid-Tier Services
- SAS Job Monitor Administration
- SAS Job Monitor Administration Help and Documentation
- SAS Job Monitor Data Server Configuration

If using SAS 9.4M5, each of these products will be at version 2.2_M1.

CONFIGURATON

To enable log monitoring, configuration steps are required at the environment level with SAS Management Console, at the SAS Job Monitor level, and during job deployment in SAS Data Integration Studio.

DATA STEP BATCH SERVER ARM LOGGING

To enable ARM logging in the Data Step Batch Server, start SAS Management Console with an administrative account such as sasadm@saspw. Locate the physical data step batch server that executes SAS Data Integration Studio batch jobs. This example assumes the SASApp – SAS DATA Step Batch Server is used for batch processing. To update the server, right click and select Properties as shown in Figure 1.

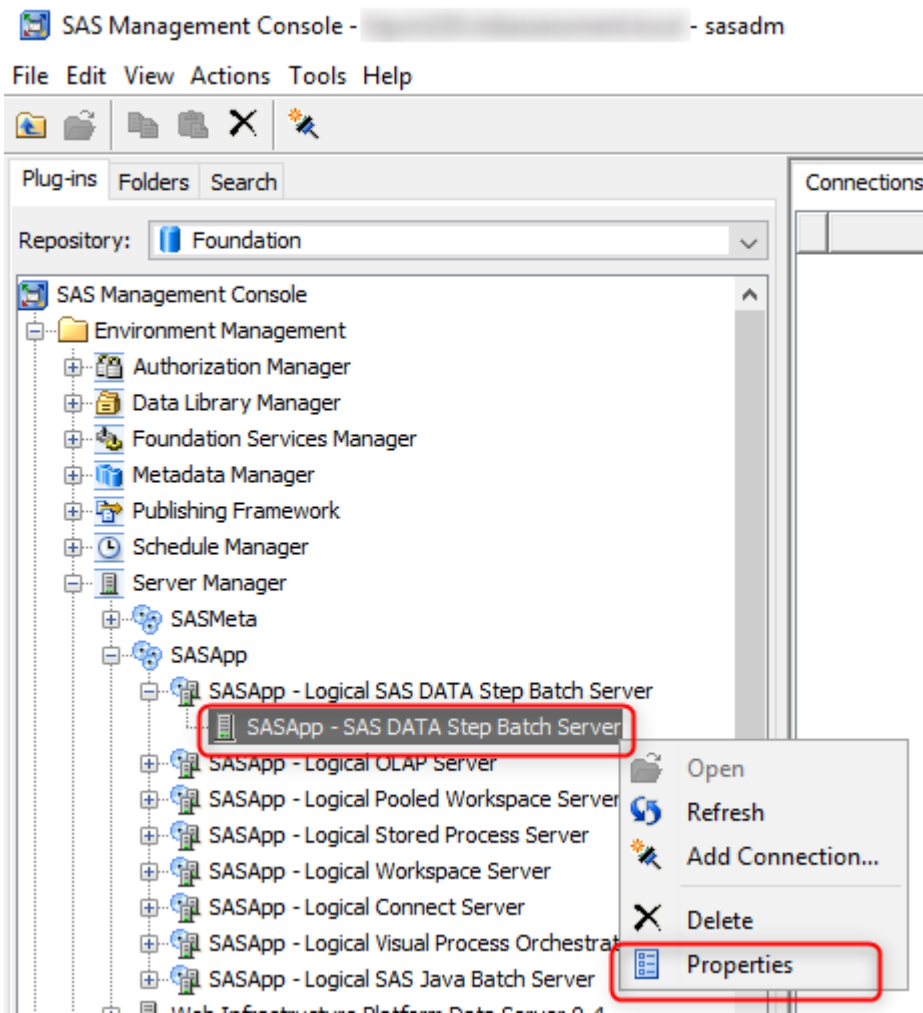


Figure 1. SAS Management Console - Batch Server Properties

There are multiple changes to make. First, select the Options tab and add the logconfigloc option to the batch command. Point the option to the logconfig.xml file located in the Lev1

BatchServer directory. The logconfig.xml file gives the batch server the ability to collect performance statistics.

In addition, the user must define the Logs Directory. One key to a successful logging configuration is to write the batch logs to a subdirectory of the defined logs directory. In this example that subdirectory is named *schedule*. Each of these settings are visible in Figure 2.

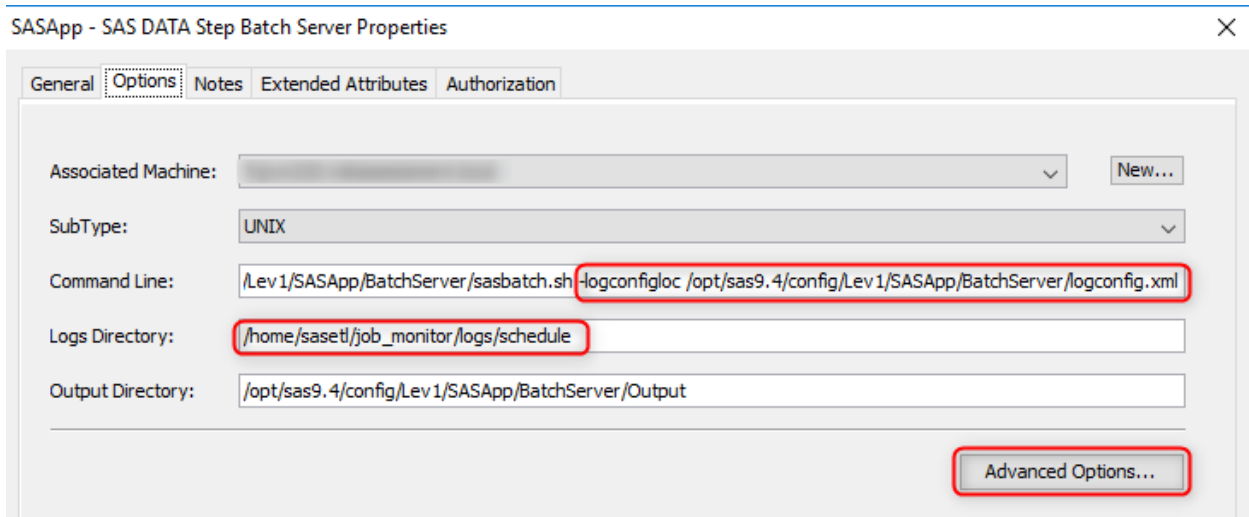


Figure 2. SAS Management Console - Data Step Batch Server Options

To make the final change select Advanced Options and replace the Rolling Log Options with a single underscore as seen in Figure 3. The Rolling Log Options will now be handled by the logconfig.xml file.

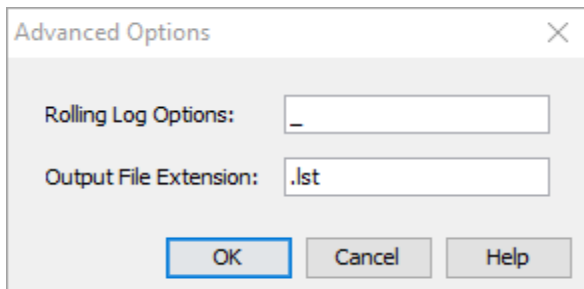


Figure 3. SAS Management Console - Data Step Batch Server Advanced Options

SAS ENVIRONMENT MANAGER AND SAS JOB MONITOR CONFIGURATION

SAS Environment Manager is a web application and is visible in Figure 4. It can be accessed through the SAS mid-tier by using the following url with an administrative account such as sasadm@saspw:

<http://<hostname>:7080>

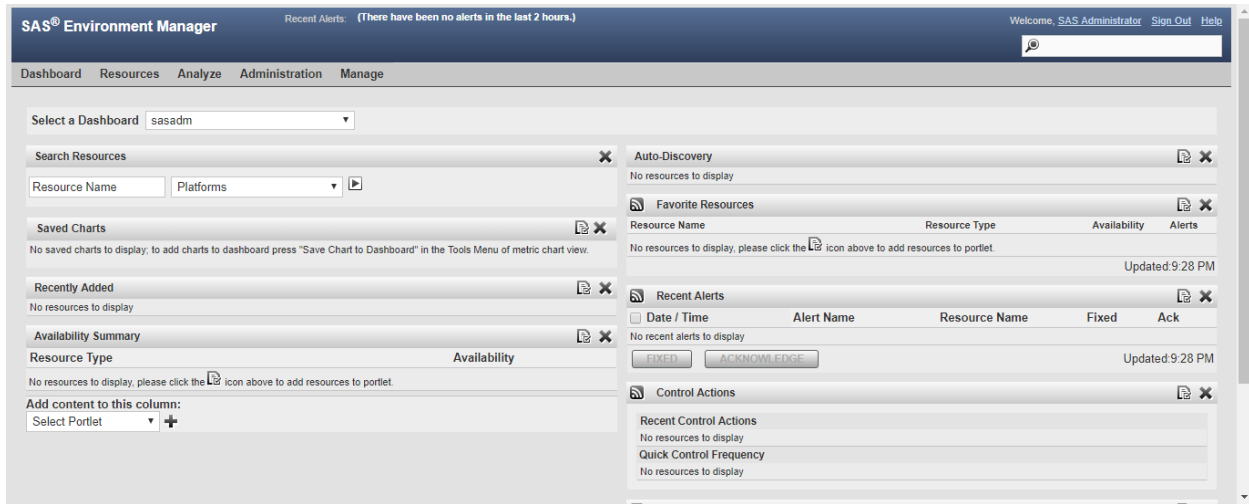


Figure 4. SAS Environment Manager

To configure SAS Job Monitor, list all available servers by using the Resources → Browse menus. The menus can be seen in Figure 5.

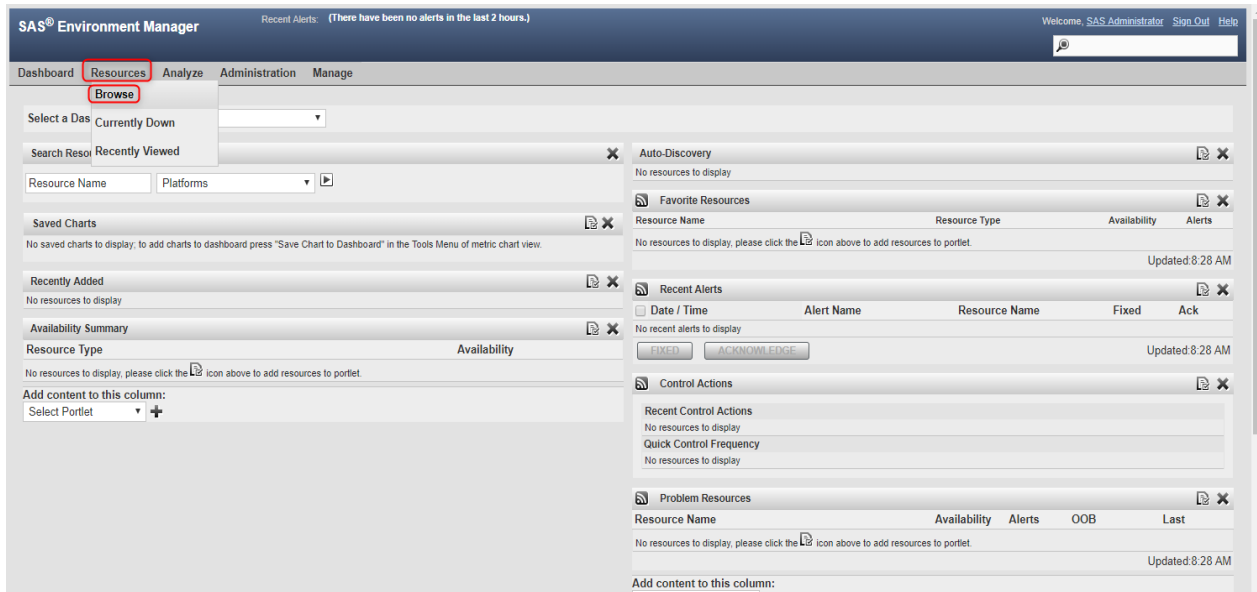


Figure 5. SAS Environment Manager - Browse Menu

Once all servers are displayed, find the SAS DI Studio server entry. This server can be identified by searching for the text *DI Studio* as illustrated in Figure 6. To view the details of the SAS DI Studio server, select the entry in the Server list.

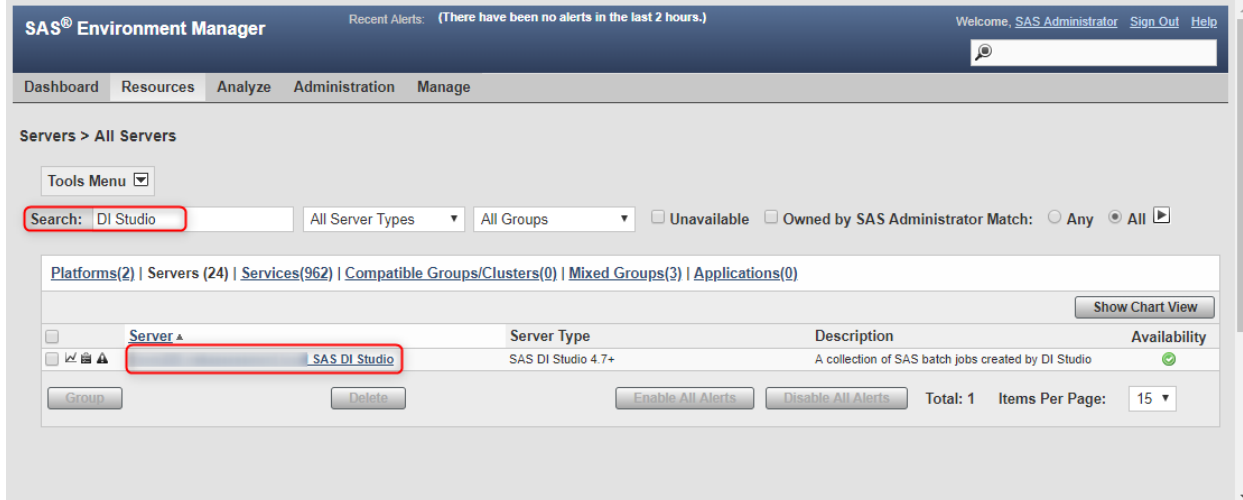


Figure 6. SAS Environment Manager - SAS DI Studio Server

If the SAS environment has multiple application servers, a SAS DI Studio entry will exist for each application server. Each SAS DI Studio server must be configured for SAS Job Monitor to process its logs.

The SAS DI Studio server resource is visible in Figure 7. To configure the resource, use the Tools Menu and select Configure Server.

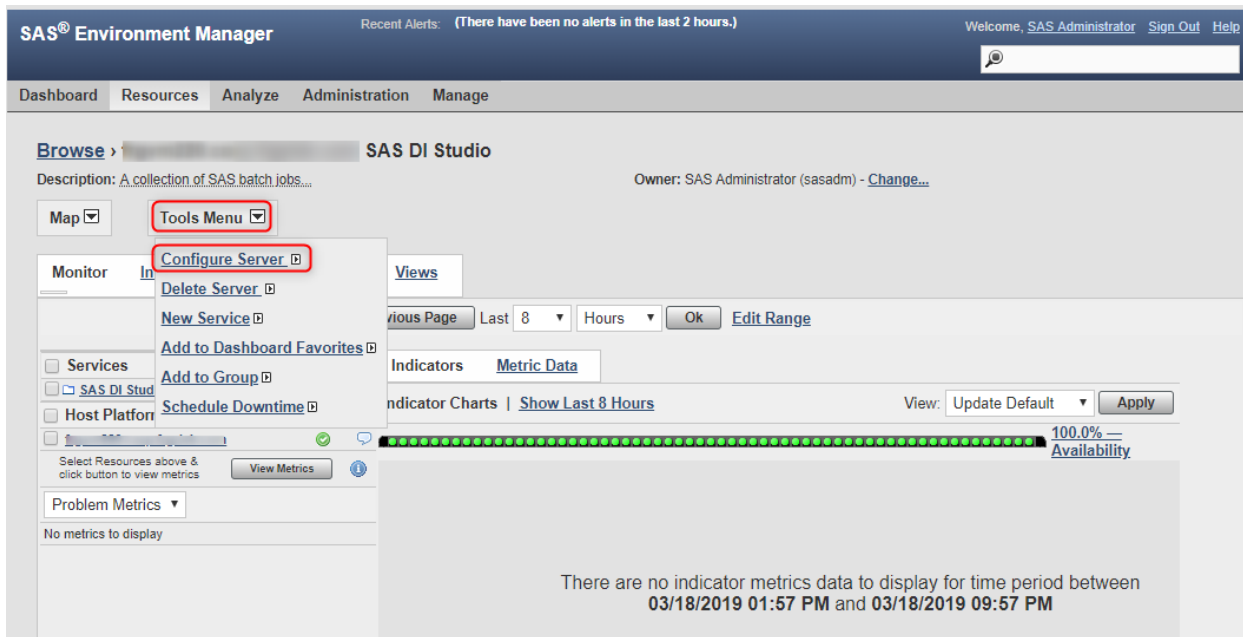


Figure 7. SAS Environment Manager - SAS DI Studio Resource

To enable the SAS Job Monitor plugin, select the checkboxes for properties *server.config_track.enable* and *Auto-Discover Batch Jobs?*. To point Job Monitor to the

batch logs, update the `server.config_track.files` property. In this example, the property is set to:

```
/home/sasetl/job_monitor/logs>true;.*\.log;
```

The property consists of three parameters that are delimited by semi-colons. The first value is the location of the SAS logs. Notice that the last directory listed is `logs`. The `schedule` sub-directory that contains the SAS logs is located within the `logs` directory and is not referenced in the property.

The second parameter is set to true and allows SAS Job Monitor to detect and process logs in sub-directories.

The last parameter is a regular expression used to determine which files to process. In this example all files ending in `.log` are processed. Users should not edit this value without having a solid understanding of regular expressions. Entering an invalid expression will prevent logs from appearing in SAS Job Monitor.

The user must also ensure that the `jobmonitoring.language` and `jobmonitoring.encoding` values correspond to the site's language and encoding settings. For the purposes of this paper, no changes were required.

These settings are shown in Figure 8. Select **OK** once configuration is complete.

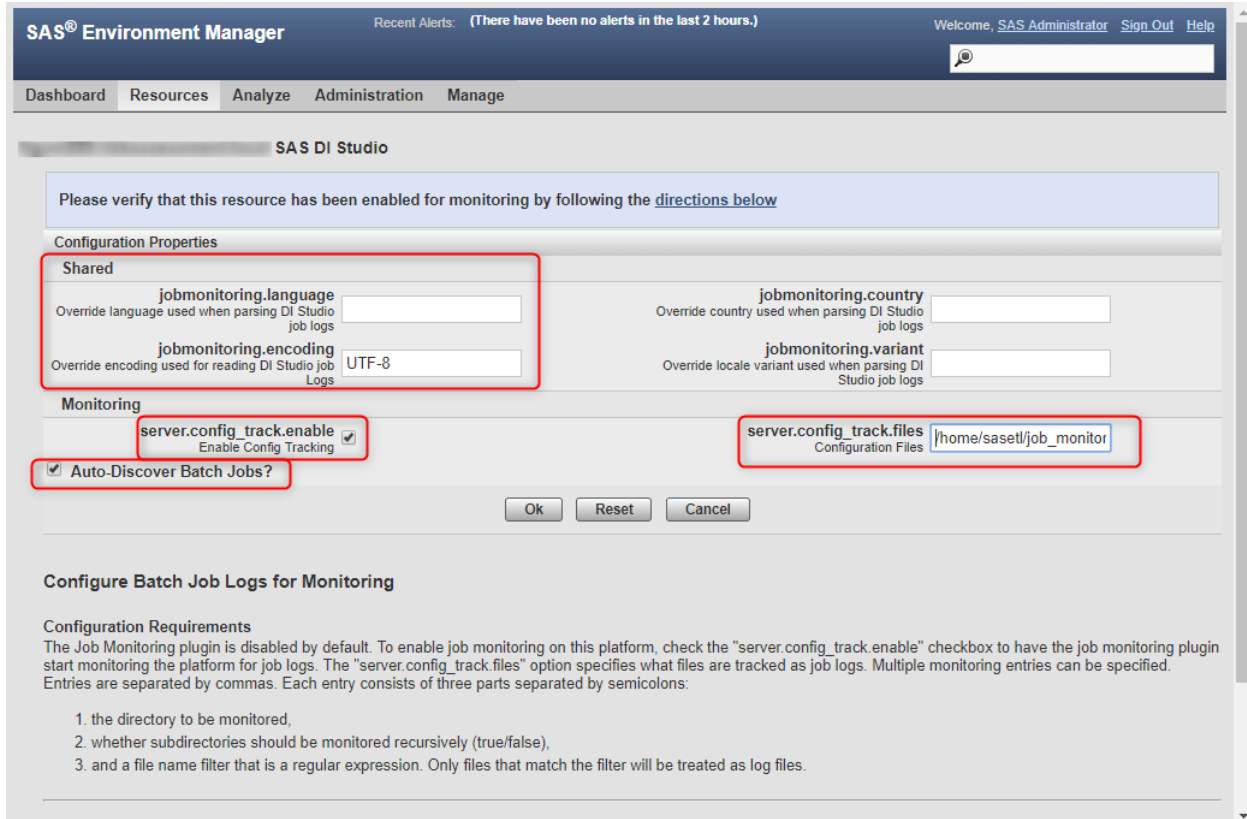


Figure 8. SAS Environment Manager - SAS Job Monitor Configuration

SAS DATA INTEGRATION STUDIO – COLLECTING STATISTICS

For SAS Job Monitor to process logs from SAS Data Integration Studio, jobs must be deployed with the Collect Runtime Statistics and Collect Table Statistics options enabled. These options can be turned on in each specific job or for all future jobs using a global option. To turn on the options in a single job, right click on the job canvas and select Collect Runtime Statistics and Collect Table Statistics as shown in Figure 9.

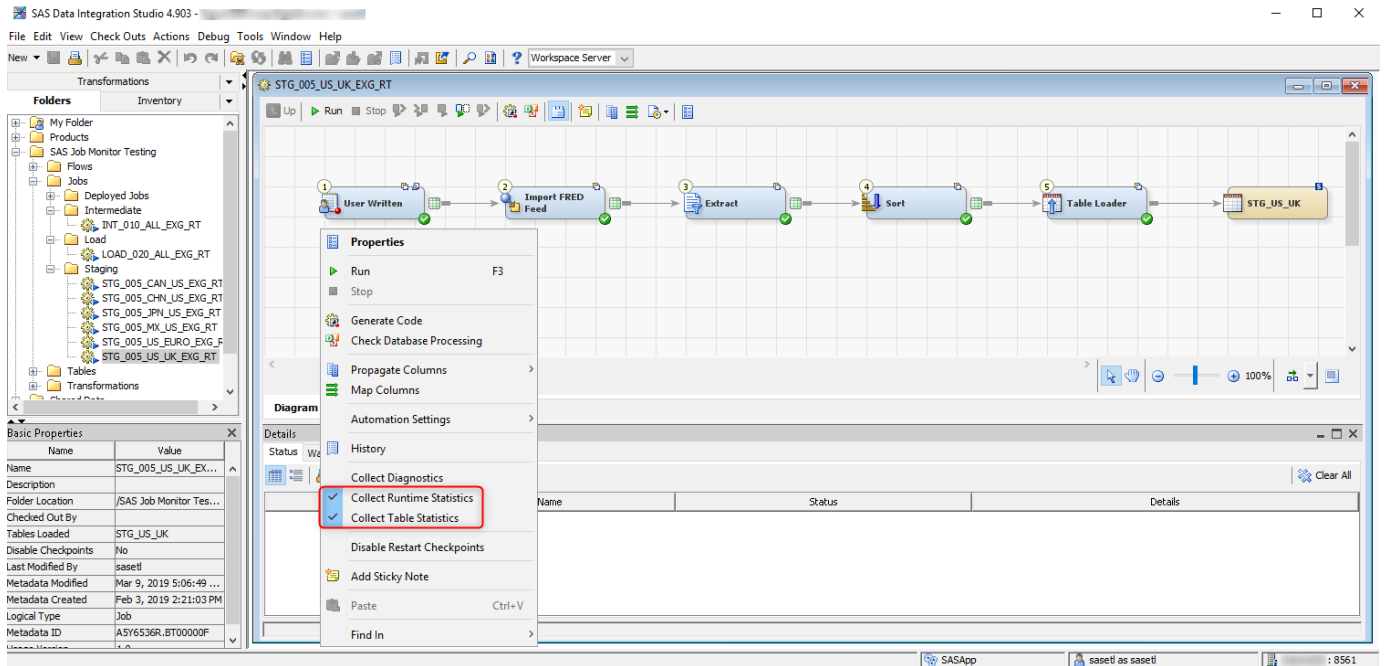


Figure 9. SAS Data Integration Studio - Job Level Statistics

To turn on these options for all new jobs, select the Tools → Options menus to open the Options window. Next, select the Job Editor tab and ensure Collect runtime statistics and Collect table statistics are selected as shown below in Figure 10. Again - these settings will not affect jobs that are already created.

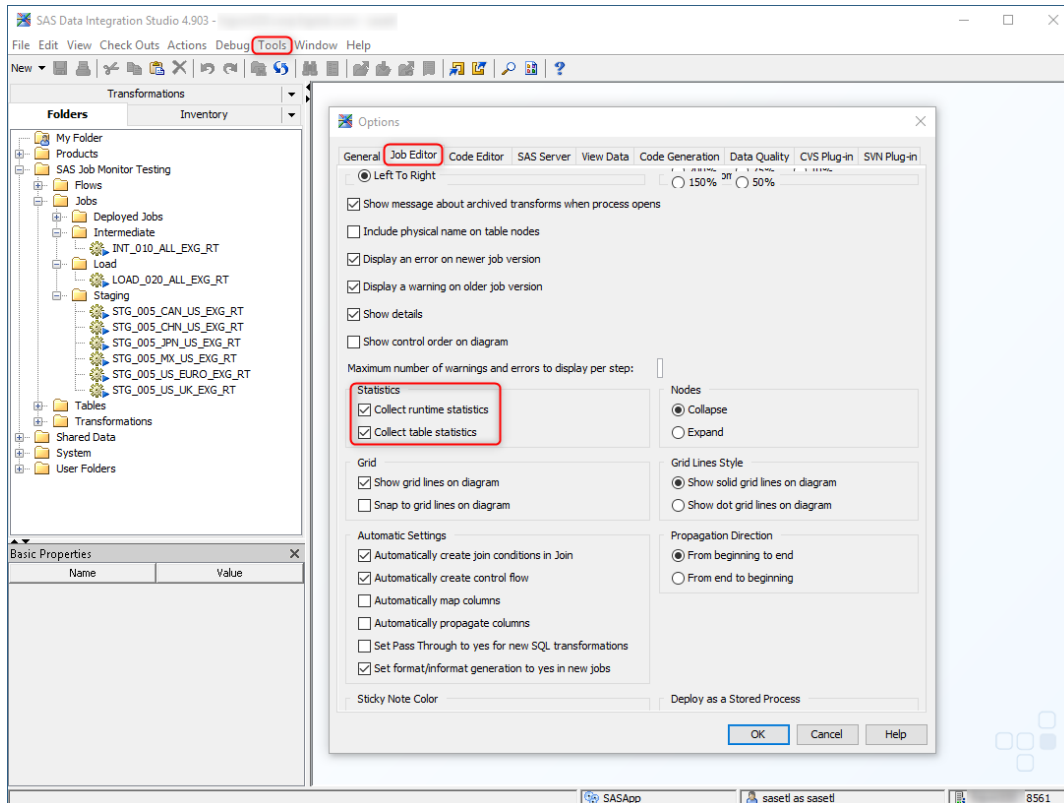


Figure 10. SAS Data Integration Studio - Global Level Statistics

In addition to turning on options that enable runtime statistics, the user must ensure those settings are not disabled during job deployment. The first step is to use the Tools -> Options menus to reopen the Options window. Next, on the General tab, make sure the *Override and disable performance statistics when a job is deployed* option is not selected. This option can be seen below in Figure 11.

Once the Collect Runtime Statistics and Collect Table Statistics options are turned on, remember to deploy or redeploy all SAS Data Integration Studio jobs.

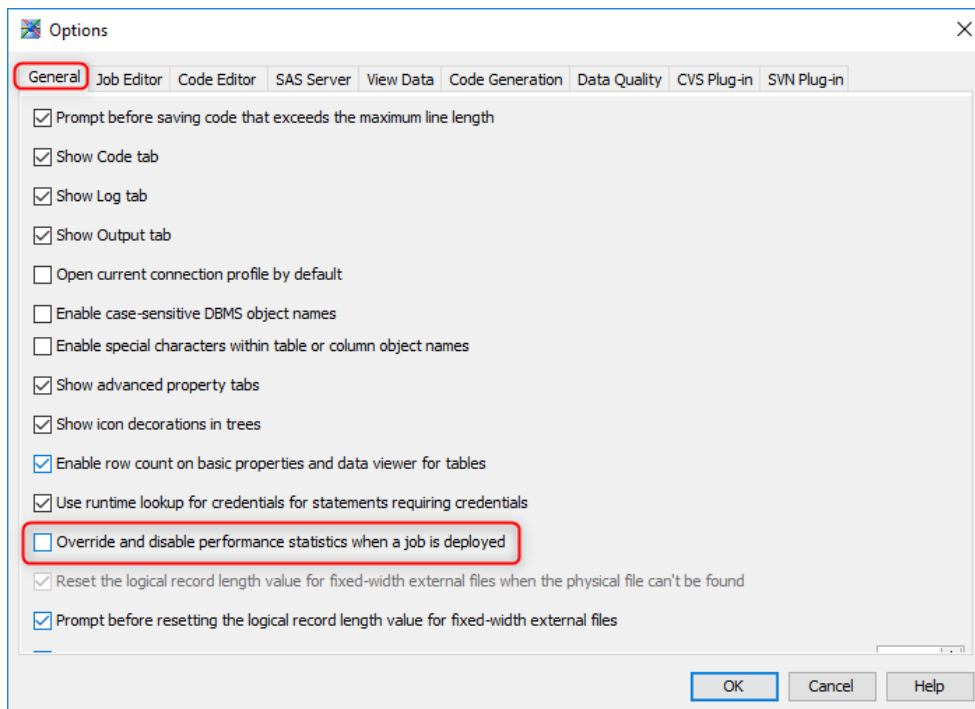


Figure 11. SAS Data Integration Studio - Override Option

APPLICATION RESPONSE MEASUREMENT

These options insert Application Response Measurement (ARM) code throughout the SAS Data Integration Studio job. The following is a sample of the ARM code:

```

/* Performance Statistics require ARM_PROC sub-system */
%macro etls_startPerformanceStats;
  %log4sas();
  %log4sas_logger(Perf.ARM, 'level=info');
  options armagent=log4sas armsubsys=(ARM_PROC);
  %global _armexec;
  %let _armexec = 1;
  %perfinit(applname="&applName");
  %global etls_recnt;
  %let etls_recnt=-1;
%mend;
%etls_startPerformanceStats;

/* Runtime statistics macros */
%etls_setPerfinit;
%perfstrt(txnname=%BQUOTE(_DISARM|&transformID|&syshostname|Extract),
metrNam6=_DISROWCNT, metrDef6=Count32) ;

```

```
%perfstop(metrVal6=%sysfunc(max(&etls_recnt,-1)));
```

When the SAS job is run, the ARM code adds performance information to the SAS log that is processed by SAS Job Monitor. Here is an example from a SAS log:

```
NOTE: PROCEDURE| _DISARM| STOP| _DISARM| 2019-02-15T08:00:12,069-05:00|  
_DISARM| BatchServer| _DISARM| SAS| _DISARM| | _DISARM| 11964416| _DISARM|  
11964416| _DISARM| 10| _DISARM| 10| _DISARM| 144| _DISARM| 432| _DISARM|  
0.020000| _DISARM| 0.017710| _DISARM| 1865854812.052231| _DISARM|  
1865854812.069941| _DISARM| 0.010000| _DISARM| | _ENDDISARM
```

SAS JOB MONITOR

Once all components are configured, SAS Job Monitor can be used to analyze SAS Data Integration Studio batch logs. Start SAS Environment Manager with url <http://<hostname>:7080> and use the Analyze menu in Figure 12 to access the Monitoring Center.

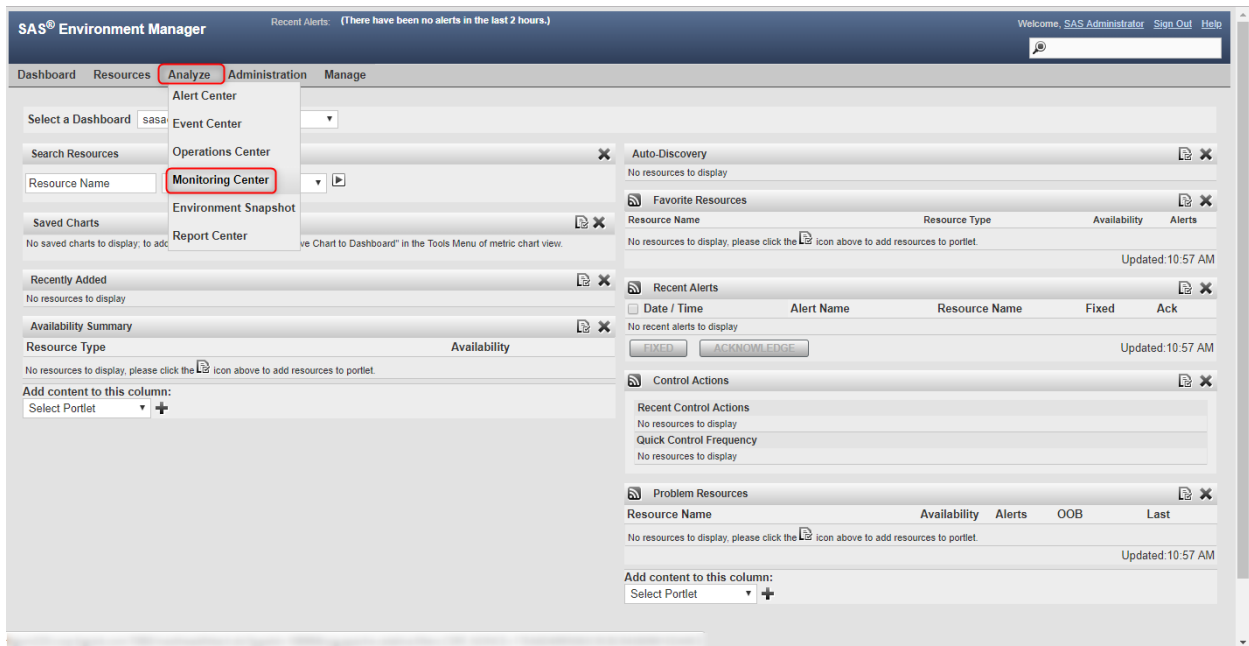


Figure 12. SAS Job Monitor - Menus

In SAS Job Monitor information such as job name and type, status, percent complete, starting and ending time as well as total run time is visible.

If SAS Job Monitor wants to highlight a potential problem, a warning symbol may be displayed to the right of the Run Time column. During testing, the following warning was displayed when hovering over the warning symbol:

Warning: The run time is significantly different from the 20 run mean time

SAS Job Monitor can be seen below in Figure 13 with the warning and trend symbols.

The screenshot shows the SAS Environment Manager Monitoring Center interface. At the top, it says 'SAS® Environment Manager' and 'Recent Alerts: (There have been no alerts in the last 2 hours.)'. Below the navigation bar, there are filters for 'Last 30 Days', 'Job Name', 'Job Type', 'Platform', 'Status', '% Complete', 'Start Time', 'End Time', 'Run Time', '20 Run Mean', and 'Trend'. The table below shows a list of jobs with their respective details and symbols.

Job Name	Job Type	Platform	Status	% Complete	Start Time	End Time	Run Time	20 Run Mean	Trend	User ID
STG_005_US_EURO_EXG_RT	DI Job		Completed	100	3/10/19 12:36:52 PM	3/10/19 12:50:54 PM	00:14:01.996	00:04:13.924		
STG_005_MX_US_EXG_RT	DI Job		Completed	100	3/10/19 12:18:50 PM	3/10/19 12:36:52 PM	00:18:01.951	00:03:49.953	Warning	Trend
STG_005_JPN_US_EXG_RT	DI Job		Completed	100	3/10/19 12:01:48 PM	3/10/19 12:18:50 PM	00:17:02.129	00:04:06.005		
STG_005_CHN_US_EXG_RT	DI Job		Completed	100	3/10/19 11:46:45 AM	3/10/19 12:01:47 PM	00:15:02.094	00:03:14.083	Warning	Trend
STG_005_CAN_US_EXG_RT	DI Job		Completed	100	3/10/19 11:27:43 AM	3/10/19 11:46:45 AM	00:19:02.091	00:03:42.171	Warning	Trend
LOAD_020_ALL_EXG_RT	DI Job		Completed	100	3/10/19 10:44:05 AM	3/10/19 11:02:05 AM	00:18:00.182	00:09:45.704		Trend
INT_010_ALL_EXG_RT	DI Job		Completed	100	3/10/19 10:34:05 AM	3/10/19 10:44:05 AM	00:10:00.074	00:02:04.336		Trend
STG_005_US_UK_EXG_RT	DI Job		Completed	100	3/10/19 10:19:03 AM	3/10/19 10:34:04 AM	00:15:01.928	00:03:18.976		Trend
STG_005_US_EURO_EXG_RT	DI Job		Completed	100	3/10/19 10:01:59 AM	3/10/19 10:19:02 AM	00:17:02.825	00:03:19.002	Warning	Trend
STG_005_MX_US_EXG_RT	DI Job		Completed	100	3/10/19 9:44:57 AM	3/10/19 10:01:59 AM	00:17:02.012	00:02:53.377	Warning	Trend
STG_005_JPN_US_EXG_RT	DI Job		Completed	100	3/10/19 9:27:55 AM	3/10/19 9:44:57 AM	00:17:01.98	00:03:10.578	Warning	Trend
STG_005_CHN_US_EXG_RT	DI Job		Completed	100	3/10/19 9:13:53 AM	3/10/19 9:27:55 AM	00:14:02.067	00:02:27.798	Warning	Trend

Figure 13. SAS Job Monitor – Monitoring Center

FILTERING LOGS

By default, SAS logs from the last eight days are available. Jobs that were run most recently appear at the top and can be filtered by name, job type, platform, and state. The Edit Range button can be selected to access date and time-based filters. Logs can be subset by *Last* days, hours, or minutes or the user can define a starting and ending datetime range.

It is often useful to show all logs that have encountered errors. To define this filter, locate the text box displaying text (*all states*), click in the drop-down menu, and select *Error*. To apply the filter, click on the checkmark that is to the right of the text boxes. Figure 14 shows logs that have encountered errors. Other options for this filter are *Running*, *Completed*, *Stopped*, and *Warning*. To clear the filter, select the two x's that are to the right of the checkmark.

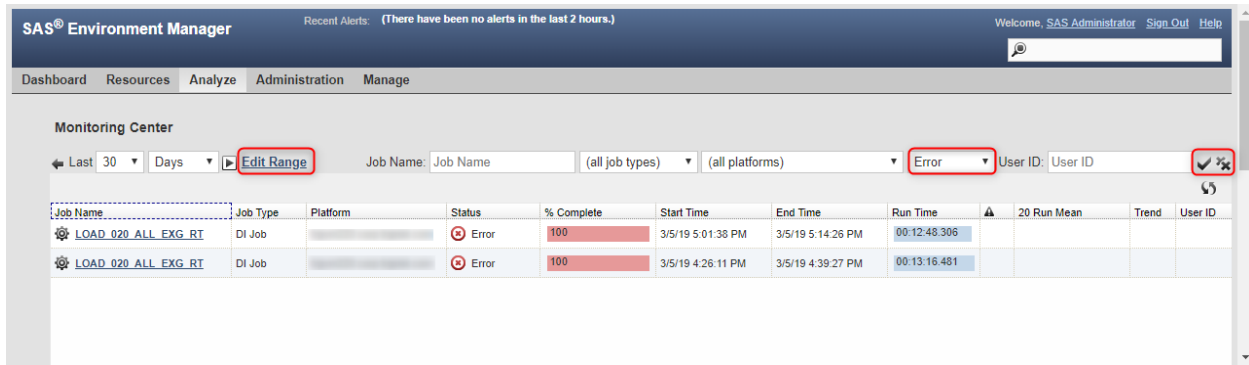


Figure 14. Monitoring Center – Completion Status Filter

Filtering with the *Job Name* text box can also be very helpful. This text box acts as a contains operator. If the ETL design includes predefined job naming conventions, the filter can show all logs for a given role in the ETL process. In this example, SAS Data Integration Studio jobs that stage data from the source have names starting with *STG*. To see all logs that stage data, enter *STG* in the Job Name text box and hit enter or select the checkbox. Figure 15 shows all staging logs.

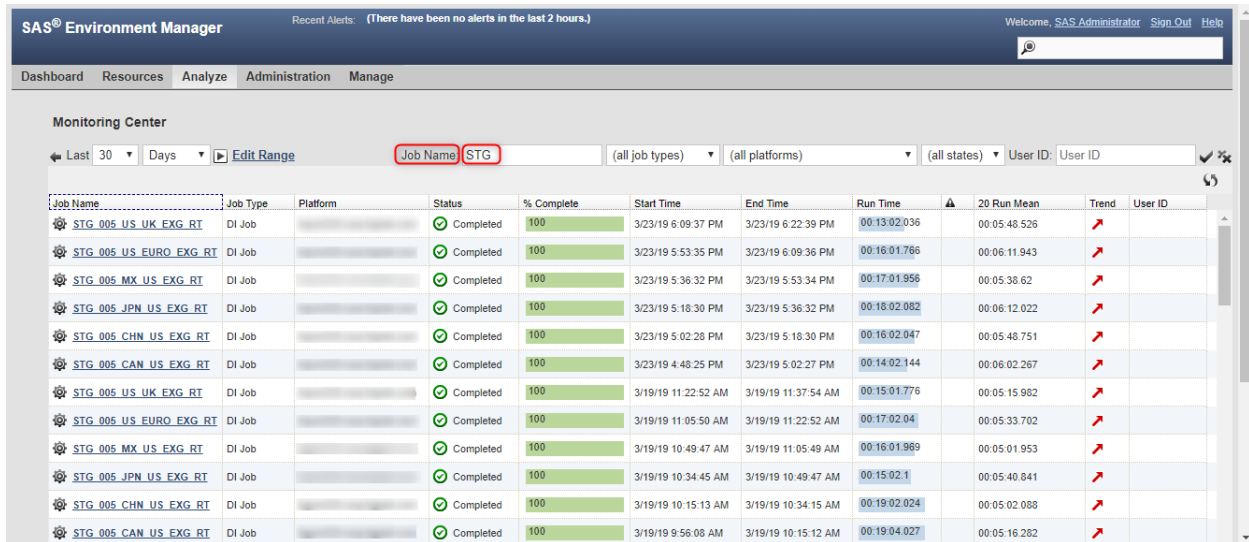


Figure 15. Monitoring Center – Job Name Filter

SAS Job Monitor will retain the filters even after the user logs out of SAS Environment Manager. To clear the filters, select the two x's that are to the right of the checkmark.

DETAILED INFORMATION

To see detailed information about a specific log, select it in the Job Name column as shown in Figure 16.

SAS® Environment Manager

Recent Alerts: (There have been no alerts in the last 2 hours.)

Welcome, SAS Administrator Sign Out Help

Dashboard Resources Analyze Administration Manage

Monitoring Center

Last 30 Days Edit Range Job Name: Job Name (all job types) (all platforms) (all states) User ID: User ID

Job Name	Job Type	Platform	Status	% Complete	Start Time	End Time	Run Time	20 Run Mean	Trend	User ID
LOAD_020_ALL_EXG_RT	DI Job		Completed	100	3/23/19 6:34:39 PM	3/23/19 6:45:40 PM	00:11:00.22	00:10:30.576	↗	
INT_010_ALL_EXG_RT	DI Job		Completed	100	3/23/19 6:22:39 PM	3/23/19 6:34:39 PM	00:12:00.08	00:03:36.722	↗	
STG_005_US_UK_EXG_RT	DI Job		Completed	100	3/23/19 6:09:37 PM	3/23/19 6:22:39 PM	00:13:02.036	00:05:48.526	↗	
STG_005_US_EURO_EXG_RT	DI Job		Completed	100	3/23/19 5:53:35 PM	3/23/19 6:09:36 PM	00:16:01.766	00:06:11.943	↗	
STG_005_MX_US_EXG_RT	DI Job		Completed	100	3/23/19 5:36:32 PM	3/23/19 5:53:34 PM	00:17:01.956	00:05:38.62	↗	
STG_005_JPN_US_EXG_RT	DI Job		Completed	100	3/23/19 5:18:30 PM	3/23/19 5:36:32 PM	00:18:02.082	00:06:12.022	↗	
STG_005_CHN_US_EXG_RT	DI Job		Completed	100	3/23/19 5:02:28 PM	3/23/19 5:18:30 PM	00:16:02.047	00:05:48.751	↗	
STG_005_CAN_US_EXG_RT	DI Job		Completed	100	3/23/19 4:48:25 PM	3/23/19 5:02:27 PM	00:14:02.144	00:06:02.267	↗	
LOAD_020_ALL_EXG_RT	DI Job		Completed	100	3/19/19 11:45:55 AM	3/19/19 11:57:55 AM	00:12:00.197	00:10:24.602	↗	
INT_010_ALL_EXG_RT	DI Job		Completed	100	3/19/19 11:37:54 AM	3/19/19 11:45:54 AM	00:08:00.082	00:03:21.231	↗	
STG_005_US_UK_EXG_RT	DI Job		Completed	100	3/19/19 11:22:52 AM	3/19/19 11:37:54 AM	00:15:01.776	00:05:15.982	↗	
STG_005_US_EURO_EXG_RT	DI Job		Completed	100	3/19/19 11:05:50 AM	3/19/19 11:22:52 AM	00:17:02.04	00:05:33.702	↗	

Figure 16. Monitoring Center – Log Selection

Once in the detailed log view, the user first sees the job name and its metadata path. The Monitoring Center also shows the date of the last job deployment and the job version. Details about this specific instance of running the job are shown next. This includes the job's completion status, it's starting and ending times, run time, and the 20 run mean.

The detailed log view offers five tabs that contain different sets of information. The *Steps* tab is visible by default and displays run time statistics broken down by the SAS Data Integration Studio transformations used in the job. Completion status, start and end time, run time, as well as rows processed and CPU time are available. This initial detail log view is shown in Figure 17.

SAS® Environment Manager

Recent Alerts: (There have been no alerts in the last 2 hours.)

Welcome, SAS Administrator Sign Out Help

Dashboard Resources Analyze Administration Manage

Monitoring Center

Jobs > LOAD_020_ALL_EXG_RT

Job Path: /SAS Job Monitor Testing/Jobs/Load/LOAD_020_ALL_EXG_RT
Deploy date: 3/9/19 5:15:08 PM Version: 4

Status: ✔ Completed

Start time: 3/23/19 6:34:39 PM Run time: 00:11:00.22
End time: 3/23/19 6:45:40 PM 20 run mean: 00:10:30.576

Trend: Distribution:

Steps Job History Job Trend Status Messages Log

Step	Instance	Status	Start Time	End Time	Run Time	20 Run Mean	Trend	Rows Processed	CPU Time
Sort		Completed	3/23/19 6:34:39 PM	3/23/19 6:35:39 PM	00:01:00.047	00:02:34.329	→	47580	20ms
Extract		Completed	3/23/19 6:35:39 PM	3/23/19 6:38:39 PM	00:03:00.036	00:03:42.689	→	47580	20ms
Join		Completed	3/23/19 6:38:39 PM	3/23/19 6:41:39 PM	00:03:00.06	00:03:06.624	→	47580	40ms
Table Loader		Completed	3/23/19 6:41:39 PM	3/23/19 6:45:40 PM	00:04:00.059	00:04:00.052	→	467945564	40ms

Figure 17. Monitoring Center - Detailed Log View - Steps tab

The next tab, titled *Job History*, contains all the logs for the job. A valuable feature of this view is the ability to see the version of deployed code for each log. If the user selects the

job status in the first column, SAS Job Monitor changes to the *Steps* view for that log instance. The Job History tab can be seen in Figure 18.

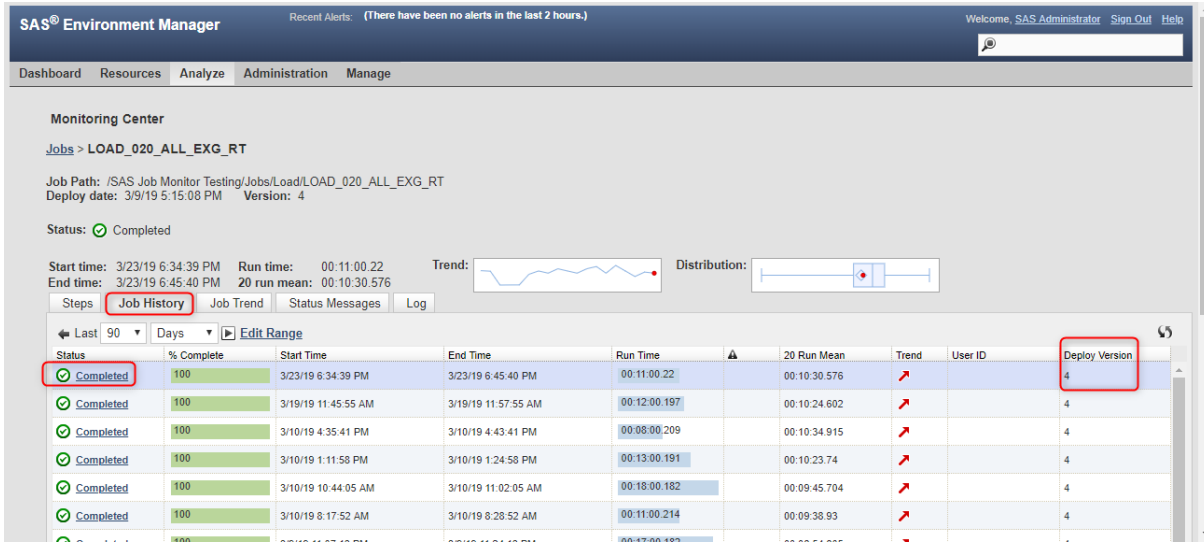


Figure 18. Monitoring Center - Detailed Log View - Job History tab

Job Trend is the next tab in the detailed view and provides a visual representation of elapsed run time by SAS Data Integration Studio job transformation. Graphing is limited to a maximum of eight logs. The Job Trend tab is shown in Figure 19.

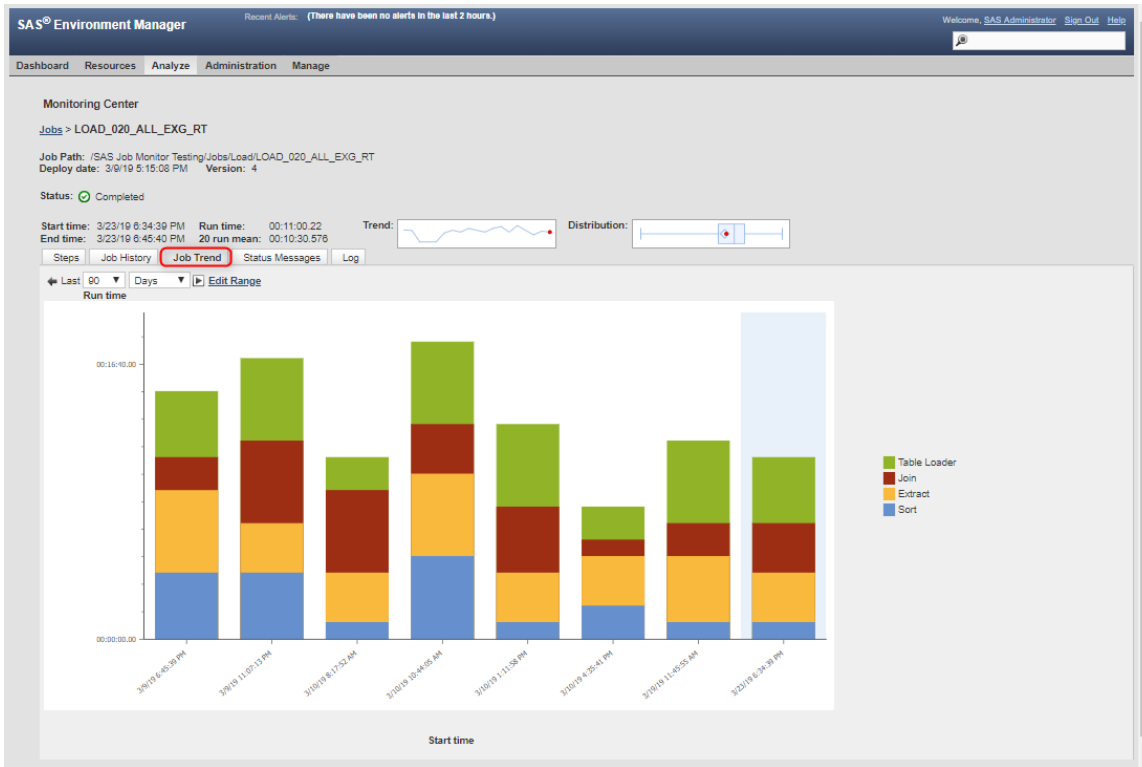


Figure 19. Monitoring Center - Detailed Log View - Job Trend tab

The next tab in the detailed view is *Status Messages*. This area displays any errors or warnings from the log. Figure 20 shows the *Status Messages* tab from a job that contained both errors and warnings.

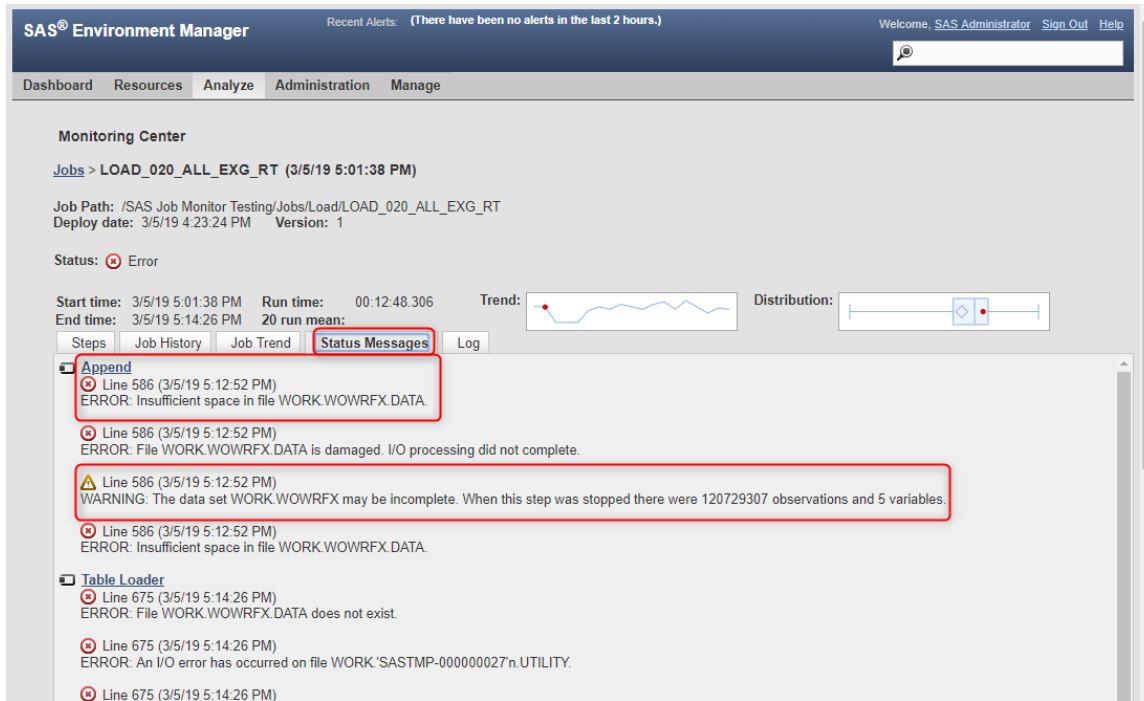


Figure 20. Monitoring Center - Detailed Log View - Status Messages tab

The last tab in the detailed view is the *Log* tab. This area simply shows the log that was created when the job was executed. By using the browser's find function, the user can search for specific steps or messages. The *Log* tab can be seen in Figure 21.

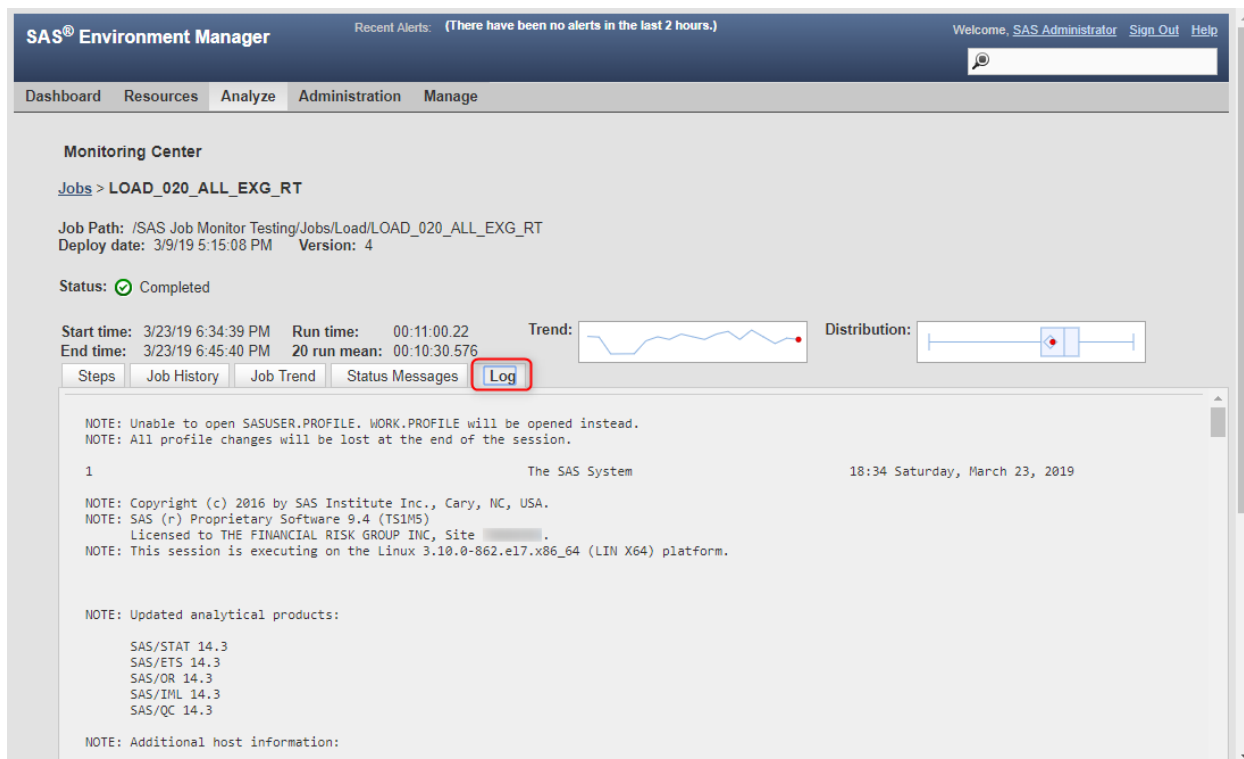


Figure 21. Monitoring Center - Detailed Log View - Log tab

ITEMS TO NOTE

- SAS Job Monitor stores processed logs in a PostgreSQL database that exists on the SAS mid-tier
- Once logs are processed by SAS Job Monitor, the physical logs no longer need to be stored in the logs directory on the compute server. Logs can be deleted from the filesystem if disk space is limited.
- SAS Job Monitor uses GIT, the open source revision control system, to detect changes in SAS logs

CONCLUSION

SAS Job Monitor is a great resource for any ETL tech lead or data warehousing administrator who supports SAS Data Integration Studio batch processing. The Monitoring Center offers basic information such as completion status, start and end times, and durations. But by surfacing supplemental details such as duration trends, job history, and break down by transformation, SAS Job Monitor proves to be an essential element in the ETL administration toolset.

ACKNOWLEDGMENTS

The author would like to thank Tom Hahl of SAS Technical Support for his assistance with this material.

RECOMMENDED READING

- SAS Job Monitor 2.2:
http://documentation.sas.com/?cdcId=evcdc&cdcVersion=2.5_M1&docsetId=evjobmtr&docsetTarget=titlepage.htm&locale=en
- SAS Data Integration Studio 4.904: User's Guide, Displaying Run-Time Statistics in SAS Job Monitor
<http://documentation.sas.com/?docsetId=etlug&docsetTarget=p151wl3sn1rbafn1jz1c7lqv55bn.htm&docsetVersion=4.904&locale=en>

CONTACT INFORMATION

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

Jeff Dyson
The Financial Risk Group
jeff.dyson@frgrisk.com
frgrisk.com