

SESUG Paper 152-2018

A Shorter Path to Moving Data Profiles from SAS® Data Management Studio to SAS® Data Management Server

Wayne Hileman, South Carolina Department of Education

ABSTRACT

Typically, one uses SAS® Data Management Studio to develop and run a data profile on a much smaller subset of data than the final target data set. Once the profile is ready for prime time or you need to profile a large data set, the data profile is run on the SAS® Data Management Server through a batch job using SAS® Data Integration Studio.

This presentation covers the steps to move a data profile from the user's machine to the SAS® Data Management Server and run through a DataFlux® Batch Job, in SAS® Data Integration Studio. Also covered are recommendations to follow that are helpful to know in advance, rather than hindsight, in getting my first data profile to run successfully. Before scaring you away, the intended audience is anyone who has access to SAS® Data Management Studio, SAS® Data Management Server, SAS® Data Integration Studio, and a simple data set that you would like to profile. Experience with either application is minimal.

INTRODUCTION

The primary purpose of this paper is to help the reader find a shorter path to moving data profiles from SAS® Data Management Studio to SAS® Data Management Server. As stated earlier, Data Management Studio is typically used to develop and test data profiles on small data sets. Once the profile works on the small dataset, it is typically deployed to the SAS® Data Management Server where the profile can be run using SAS® Data Integration Studio.

Deploying data profiles to Data Management Server frees up resources and Data Management Studio on the client machine. Time not wasted running data profiles using Data Management Studio is time and resources that can be allocated to more effective use of Data Management Studio (e.g., developing new data profiles, explorations, batch jobs; managing business rules, and monitoring data quality reports across the organization).

When I am learning to do something for the first time, often, it helps me to have a simple example from which I can follow along with step by step detailed instructions, to help get me across the finish line, which is nothing short of successful completion of the task at hand. The bottom line is I want to see an example that leads me from point A to point B, step by step, with point B being across the finish line---not short of it! While SAS® Support provides voluminous documentation on all three of these SAS® products, the documentation was spread across three manuals and multiple sections, while still lacking enough of the specific details to help me complete the task at hand. At the other extreme, my search on this topic at LexJansen.com was unsuccessful. Coming to my rescue was SAS® Technical Support. Between the two of us, we found all the right needles and cut the haystack down to size. I hope this paper helps rescue the reader at SESUG and at LexJansen.

In my example, point A is creating a Server Repository database. Point B is successfully completing a DataFlux® Batch Job using SAS® Data Integration Studio. The batch job runs a data profile located on the SAS® Data Management Server. This data profile used to run on the client machine using SAS® Data Management Studio. Along the way, I cover some of the details of tasks left out of the SAS® documentation---primarily through screenshots. Finally, I list a few lessons learned.

STEPS TO A SHORTER PATH CHECKLIST

Before diving into the steps for a shorter path, it is important to keep in mind that the steps listed below should be followed in order as each subsequent step depends on the successful completion of earlier steps. Here are the required steps:

1. Create “Server Repository” database.
2. Create DataFlux® 32-bit ODBC connections on client machine.
3. Create data profile using SAS® Data Management Studio.
4. Create DataFlux® 64-bit ODBC connections on SAS® Data Management Server.
5. Create “Server Repository” Definition using SAS® Data Management Studio.
6. Update the server.rcf file on SAS® Data Management Server.
7. Start SAS® Data Management Server on the SAS® compute tier.
8. Add Data Management Server in SAS® Data Management Studio.
9. Copy data profile to the “Server Repository”.
10. Configure DataFlux® Batch Job using SAS® Data Integration Studio.

STEP 1. CREATE “SERVER REPOSITORY” DATABASE

Create a new database to store the DataFlux® “Server Repository” repository (see Step 5). Unlike DataFlux® repositories created in SAS® Data Management Studio, SAS® Data Management Server cannot use the typical.rps DataFlux® repositories. In our example, we will be using SQL Server database as the database repository.

STEP 2. CREATE DATAFLUX® 32-BIT ODBC CONNECTION ON CLIENT MACHINE

SAS® Data Management Studio is a 32-bit application. Therefore, SAS® Data Management Studio can only work with 32-bit ODBC data connections. So, you will want to create 32-bit ODBC data connections on your client machine, to develop your data profiles using SAS® Data Management Studio. Note that jobs that run on SAS® Data Management Studio can use most any 32-bit ODBC connection. However, I recommend using the DataFlux® 32-bit SQL Server Wire Protocol to make the move to SAS® Data Management Server easier (see Figure 1). Further recommend prefixing the name for all DataFlux® SQL Server Wire Protocol ODBC connections with df_ (see Figure 2). This makes it easier to differentiate between DataFlux® ODBC connections on the client machine and standard 32-bit ODBC connections pointing to the same database (e.g., SQL Server). This same strategy works well on the SAS® Data Management Server when you have to create DataFlux® 64-bit SQL Server Wire Protocol ODBC connections.

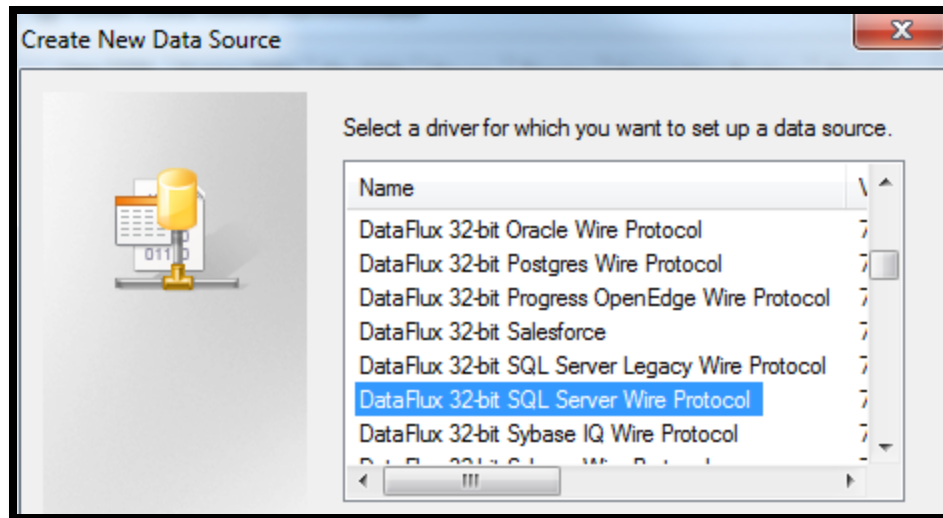


Figure 1. DataFlux® 32-bit SQL Server Wire Protocol ODBC on client machine running SAS® Data Management Studio.

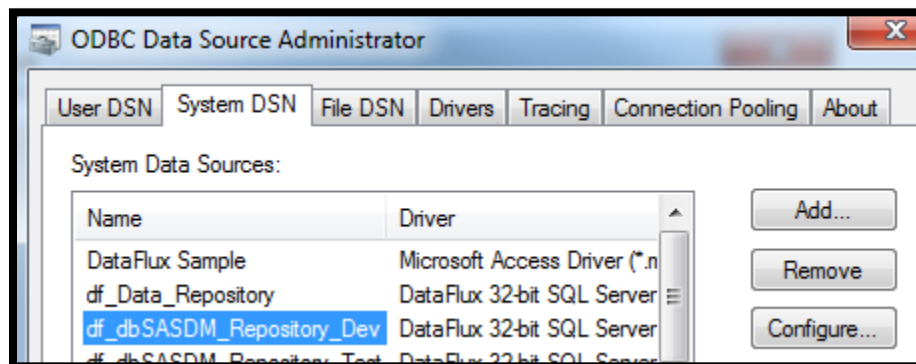


Figure 2. Prefix DataFlux® SQL Server ODBC connection names with df_

STEP 3. CREATE DATA PROFILE USING SAS® DATA MANAGEMENT STUDIO

Next, create a data profile using SAS® Data Management Studio. For our example, we will use the data connection created in Step 2. Figure 3 shows the data profile.

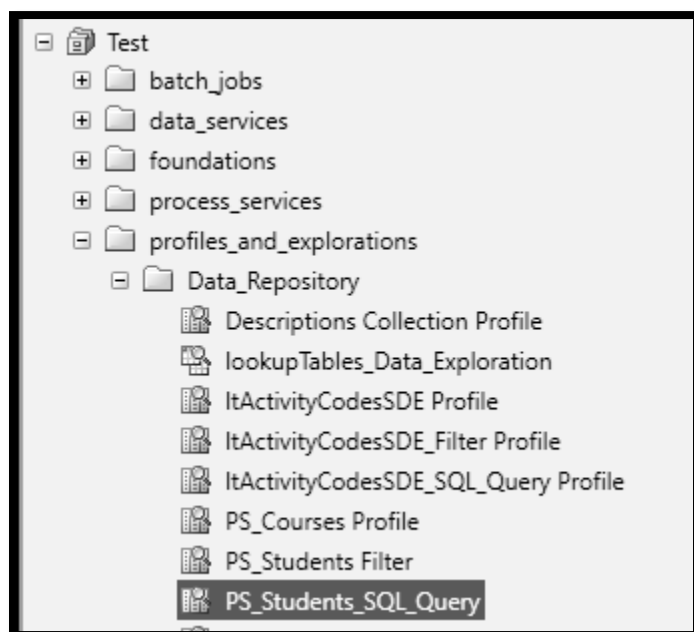


Figure 3. Data profile on Test Repository shown in SAS® Data Management Studio.

STEP 4. CREATE DATAFLUX® 64-BIT ODBC CONNECTIONS ON SAS® DATA MANAGEMENT SERVER

Next, create a DataFlux® 64-bit SQL Server Wire Protocol ODBC connection on the SAS® Data Management Server pointing to the Server Repository database created in Step 1 (see Figures 4 and 5).

SAS® Data Management Server requires 64-bit ODBC data connections---unlike SAS® Data Management Studio. So, you will need to create a 64-bit ODBC data connection on the SAS® Data Management Server that points to the “Server Repository” database created in Step 1. Next, create a DataFlux® 64-bit SQL Server Wire Protocol ODBC connection for each DataFlux® 32-bit SQL Server Wire Protocol ODBC connection created on your client machine for use with SAS® Data Management Studio. Tip: use the same name.

Again, recommend prefixing your DataFlux® SQL Server Wire Protocol ODBC connection names with df_, to make it easier to differentiate between DataFlux® ODBC connections on the client machine and standard ODBC connections pointing to the same database (see Figure 4).

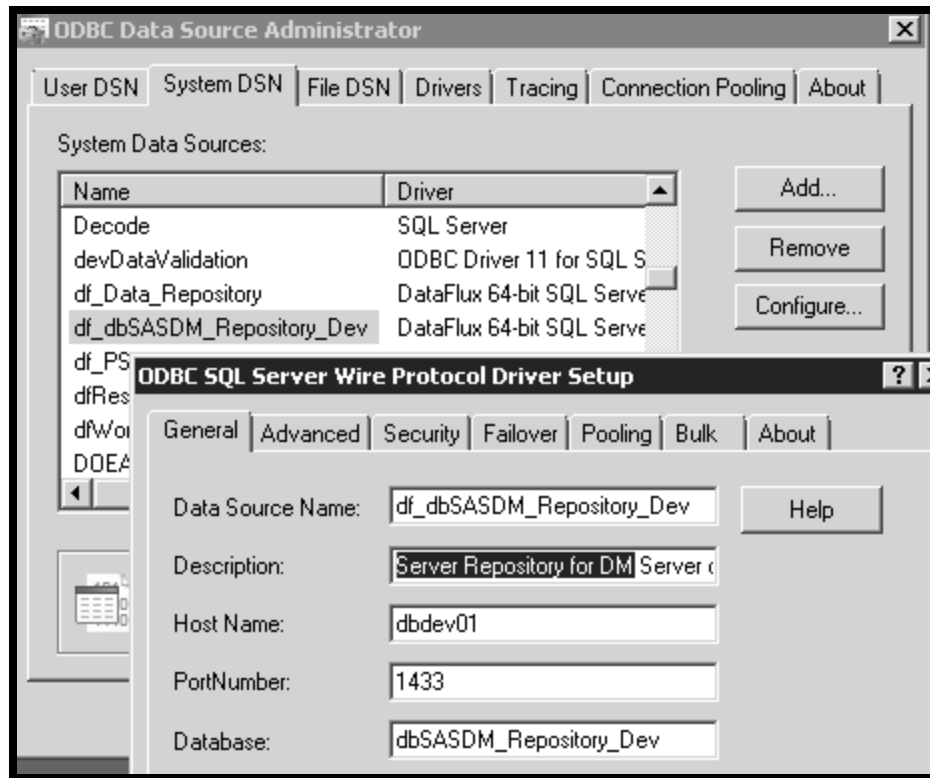


Figure 4. DataFlux® 64-bit SQL Server Wire Protocol ODBC Connection to Server Repository

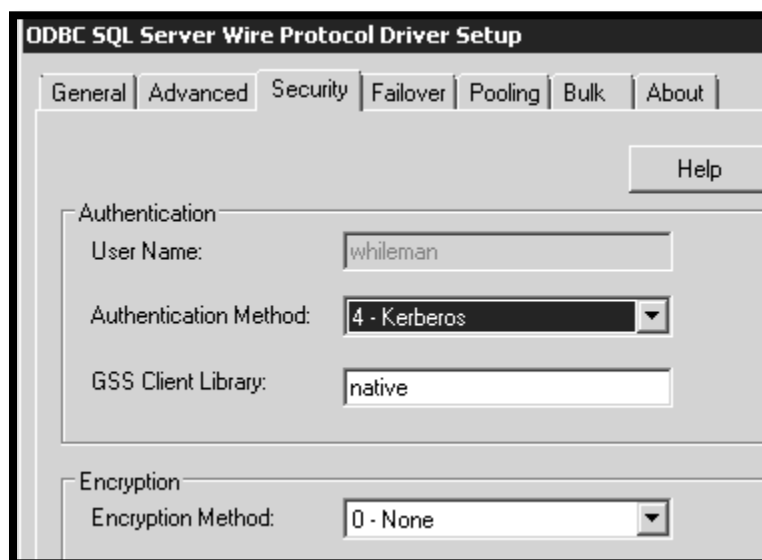


Figure 5. Security settings for DataFlux® SQL Server Wire Protocol ODBC

STEP 5. CREATE “SERVER REPOSITORY” DEFINITION IN SAS® DATA MANAGEMENT STUDIO

In SAS® Data Management Studio, from the Administration riser bar, create a new Repository Definition named “Server Repository”. It must be named “Server Repository” (see Figure 6).

New Repository Definition

Name:

Data storage
Specify a database for the storage of items requiring a database.

☐ Database file
Location:

☒ Database connection
Name:
Table prefix (optional):

File storage (optional)
Specify a folder for the storage of file-based items.

Folder:

☒ Connect to repository at startup
☒ Private

Figure 6

STEP 6. UPDATE THE SERVER.RCF FILE ON SAS® DATA MANAGEMENT SERVER

First, it is recommended that you have only one file stored in the \repositories folder with a 'rcf' extension (see Figure 7). Next, check the server.rcf file and ensure that the Name property is set to "Server Repository" and that the value of the UnifiedRepositoryConnect element is set to the name of your ODBC data connection created in Step 4 (see shaded text in Figure 8).

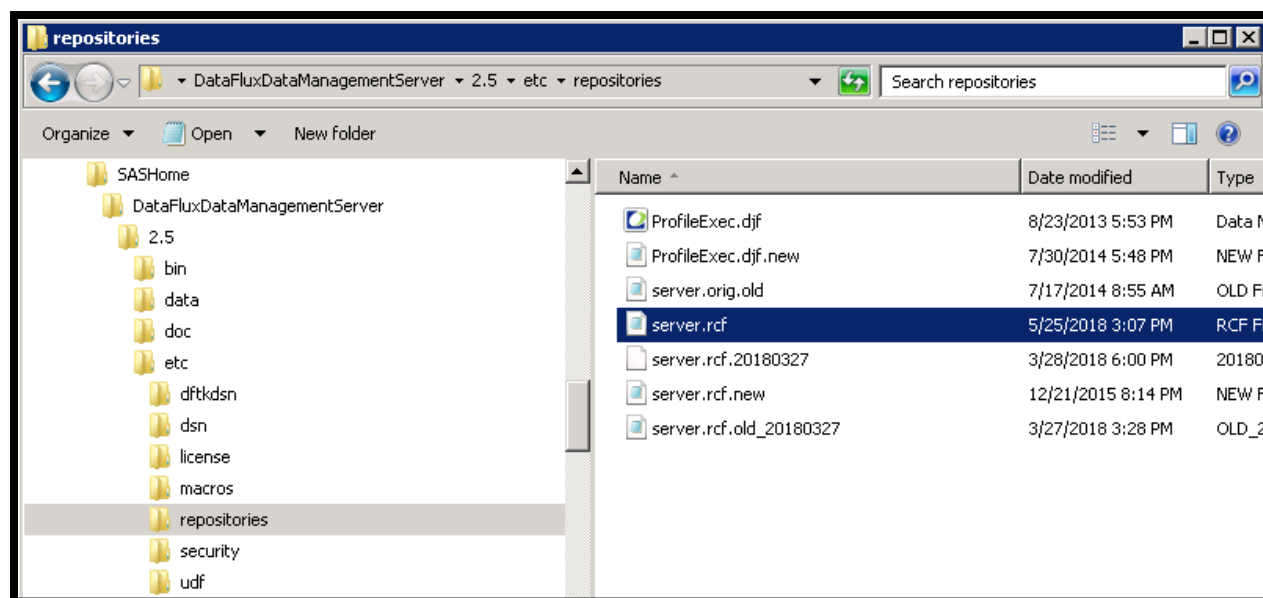


Figure 7. etc/repositories folder

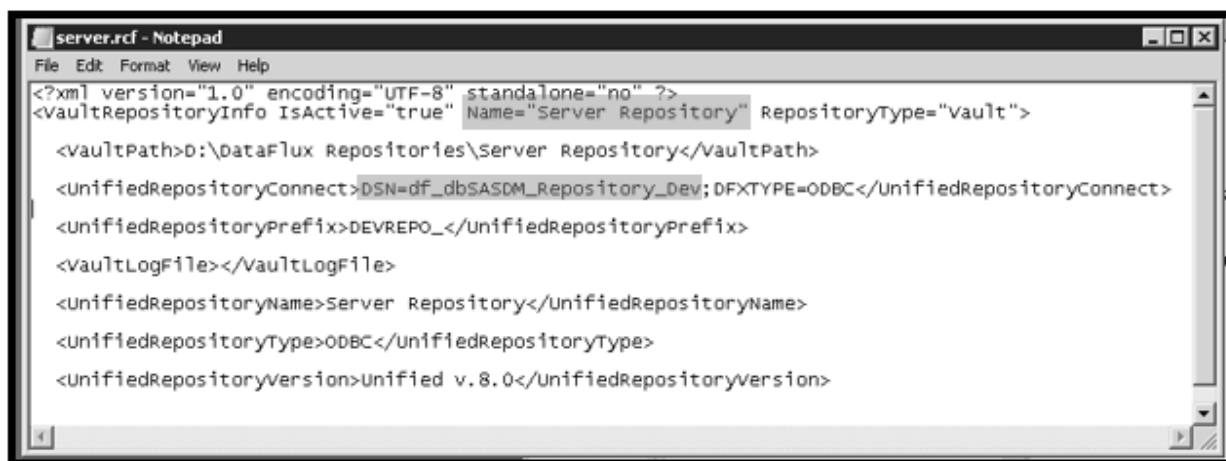


Figure 8. Server.rcf

STEP 7. START THE SAS® DATA MANAGEMENT SERVER

Connect to the SAS® Data Management Server and start or restart (if already started) the DataFlux® Data Management Server service from the Services window in Windows (Figure 9). Note that you will need to configure the properties of the service with a sufficient logon account.

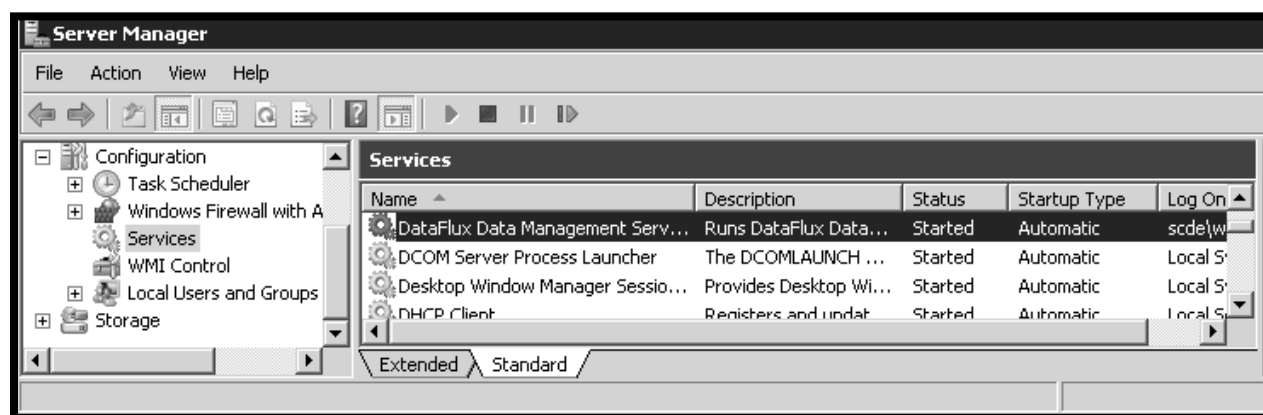


Figure 9. DataFlux® Data Management Server service - Started (on SAS® compute-tier)

STEP 8. ADD SAS® DATA MANAGEMENT SERVER TO SAS® DATA MANAGEMENT STUDIO

Next, from the Data Management Servers riser, on the client machine, in SAS® Data Management Studio, add a new Data Management Server. Enter Name and Server. Then click Test Connection. You should be prompted by Log On (see Figure 10). If not prompted for User ID and Password, confirm you entered the proper Server (e.g., sasdev03) and that the Data Management Server is started (see Step 7). If connection test is successful, click OK. You should see the new Data Management Server listed in SAS® Data Management Studio (see Figure 11).

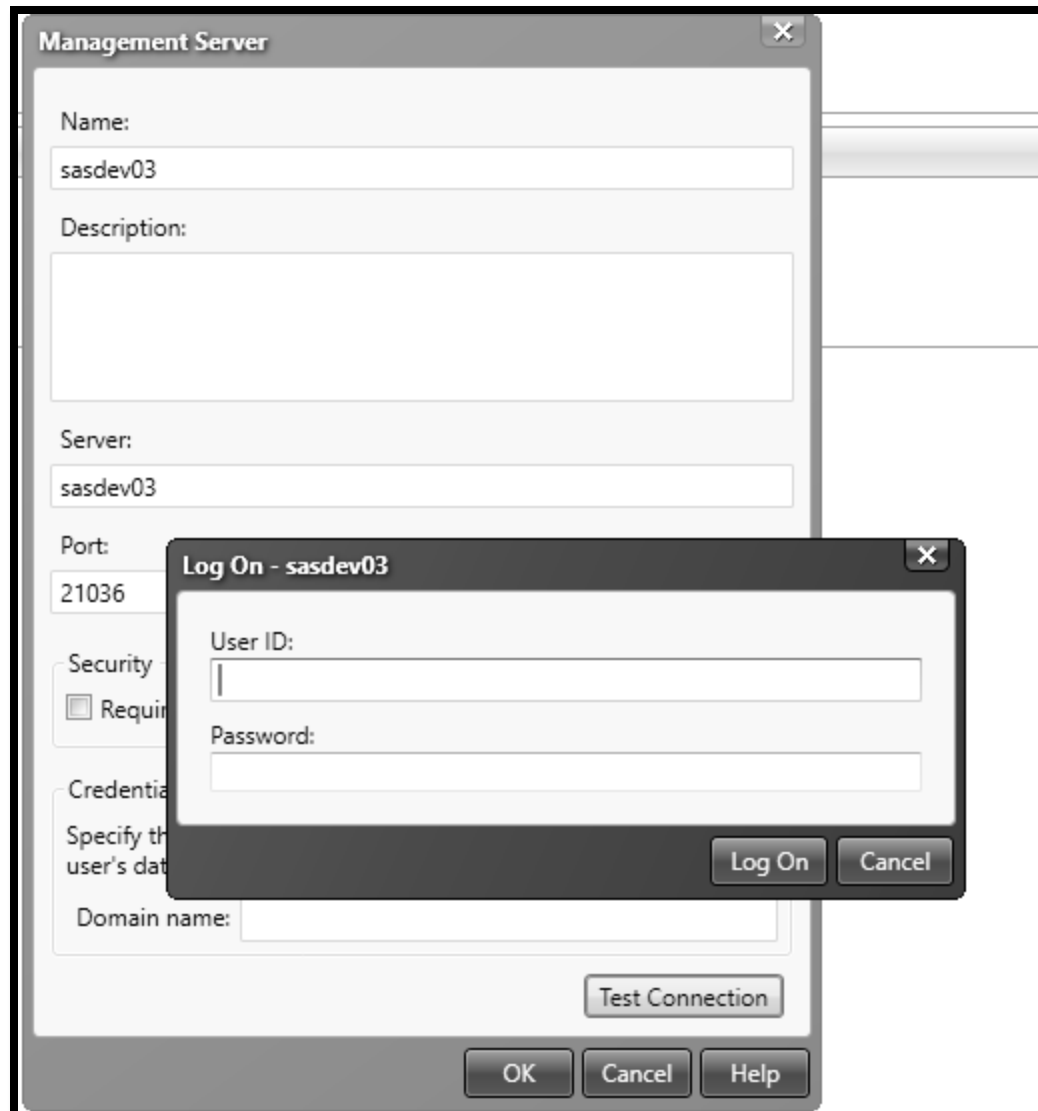


Figure 10. SAS® Data Management Studio – SAS® Data Management Server Test Connection Logon Prompt

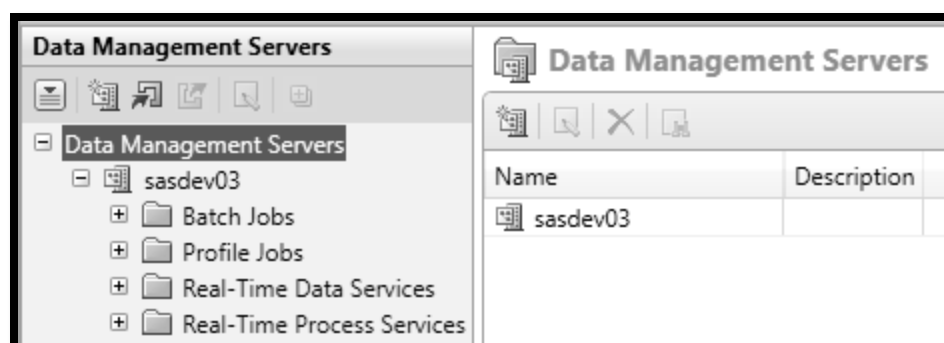


Figure 11. SAS® Data Management Studio – Data Management Servers folder

STEP 9. COPY DATA PROFILE TO THE “SERVER REPOSITORY”

Now, you are ready to copy the data profile created in Step 3 to the Data Management Server. So, in SAS® Data Management Studio, select the Folders riser bar, copy the data profile, and paste the data profile to the Server Repository (see Figure 12). Now, you are ready to test run the data profile, on the SAS® Data Management Server, using SAS® Data Management Studio, on the client machine. So, right click the profile and select run... (see Figure 13).

Check the status under the Run History tab (see Figure 14). If it read “Completed Successfully”, congratulations you have run your first data profile on the SAS® Data Management Server successfully. Feel free to stop here. You now have a data profile running on the SAS® Data Management Server that used to consume the resources on your local machine. However, to really take advantage of the SAS® Data Management Server, why not set up the data profile to run in SAS® Data Integration Studio? Doing so will free up your client machine to handle other SAS® Data Management Studio tasks (e.g., creating new profiles, batch jobs, business rules).

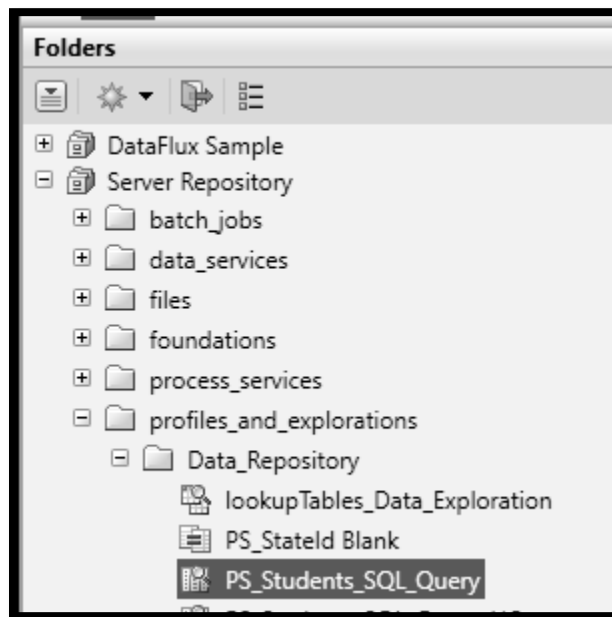


Figure 12. SAS® Data Management Studio – Folders riser bar – Profile selected

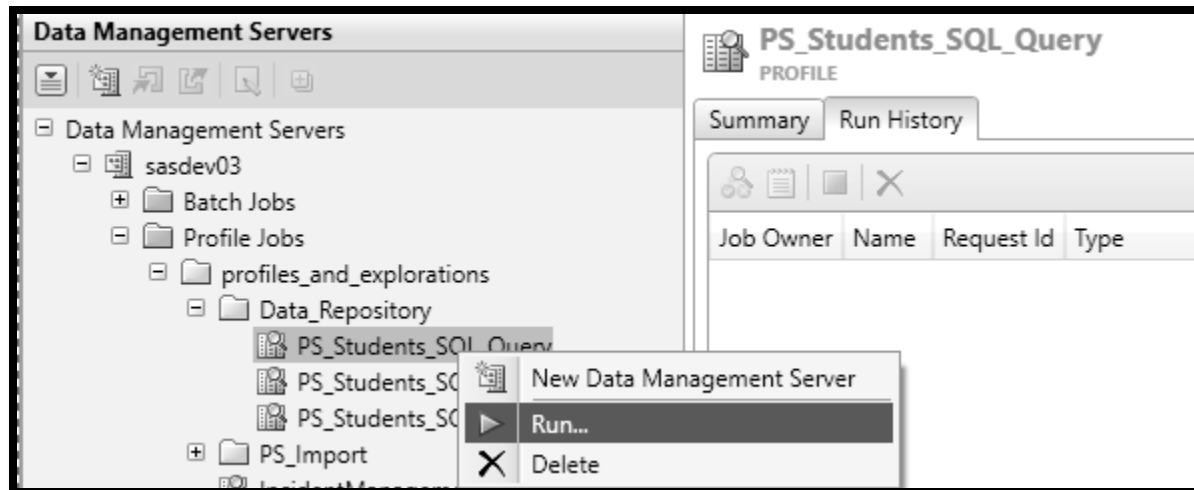


Figure 13. SAS® Data Management Studio - Data Management Servers – Run profile menu option

The screenshot shows the 'Run History' tab for the 'PS_Students_SQL_Query' profile. The table below displays the run history data.

Job Owner	Name	Request Id	Type	Start Time	End Time	Run Time (minutes)	Status
whileman	PS_Students_SQL_Query_12	12	Profile	8/13/2018 11:59 AM	8/13/2018 12:00 PM	00:31.00	Completed Successfully

Figure 14. SAS® Data Management Studio – Data Management Servers - Run History Tab

STEP 10. CONFIGURE DATAFLUX® BATCH JOB IN SAS® DATA INTEGRATION STUDIO

If you are still reading, great! You must be contemplating running your data profile on the SAS® Data Management Server using SAS® Data Integration Studio, rather than from SAS® Data Management Studio (see Step 9). So, create a new job using SAS® Data Integration Studio. Next, from the Transformations tab, drag a DataFlux® Batch Job to the job canvas, from the Data Quality section (see Figure 15). Next, right click the DataFlux® Batch Job transformation, on the job canvas, and select Properties, and then click the Job tab (see Figure 16). Next, select the listed server from the Server dropdown, Repository from the Job Type dropdown, and the job you copied to the SAS® Data Management Server in Step 9 from the Job dropdown (see Figure 17). Click OK and Save the job. Finally, test run the job created in SAS® Data Integration Studio until you get a completed successfully status (see Figure 18). Alternatively, using SAS® Data Management Studio, you can check the status (as well as view the log) of a DataFlux® Batch Job run from SAS® Data Integration Studio. To do so, in SAS® Data Management Studio, from the Data Management Servers riser bar, right click the profile, and select Open Log (see Figure 19).



Figure 15. SAS® Data Quality Transformations available in SAS® Data Integration Studio.

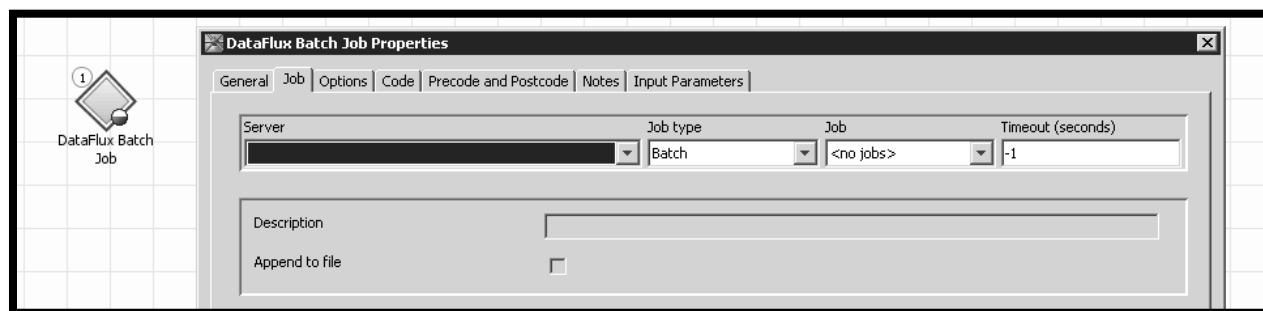


Figure 16. DataFlux® Batch Job Properties – Job Tab – ready to configure new job.

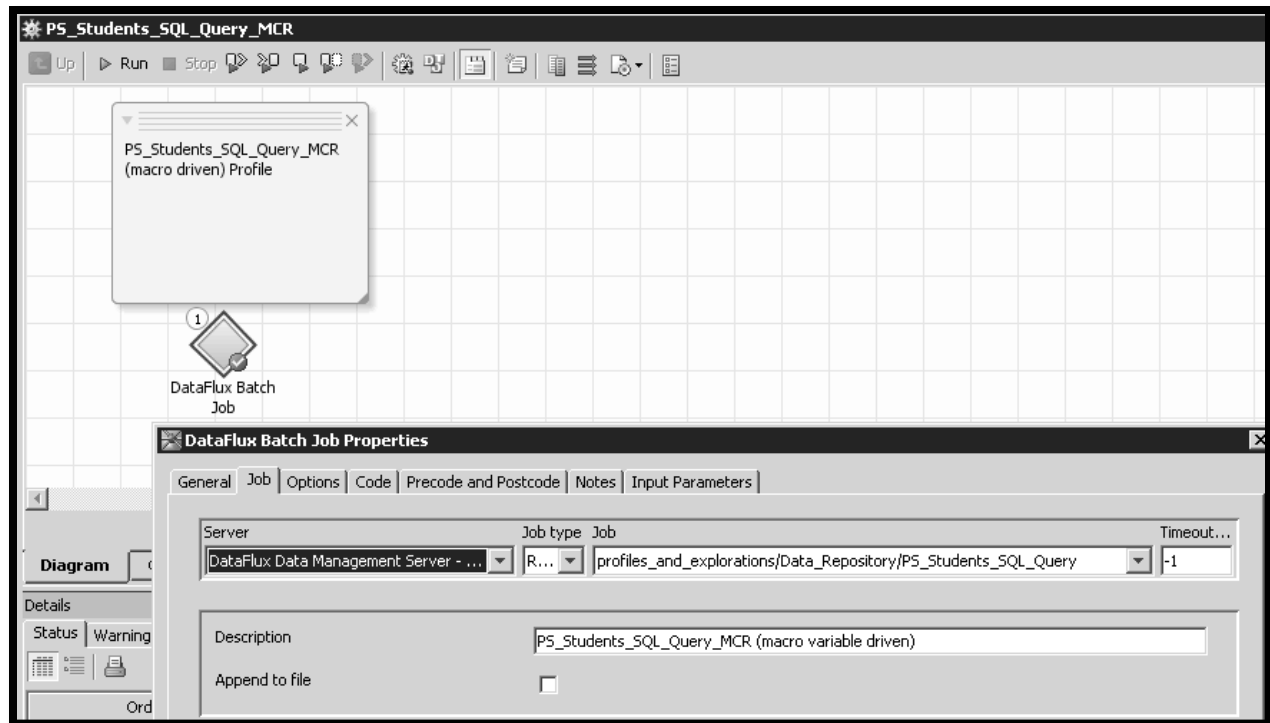


Figure 17. DataFlux® Batch Job Properties – Job Tab – new job configured

Details

Status Warnings and Errors Statistics Control Flow

Order	Name	Status
1	Precode	Completed successfully
2	DataFlux Batch Job	Completed successfully
3	Postcode	Completed successfully
	PS_Students_SQL_Query_MCR	Completed successfully

Figure 18. SAS® Data Integration Studio Job Status

DataFlux Data Management Studio 2.6

File View Tools Window Help New ?

Data Management Servers

- sasdev03
 - Batch Jobs
 - Profile Jobs
 - profiles_and_explorations
 - Data Repository
 - PS_Students_SQL_Query

PS_Students_SQL_Query PROFILE

Summary Run History

Job Owner	Name	Request Id	Type	Start Time	End Time	Run Time (minutes)	Status
whileman	PS_Students_SQL_Query	40	Profile	8/13/2018 11:59 AM	8/13/2018 12:00 PM	00:31:00	Completed Successfully
whileman	PS_Students_SQL_Query		File	8/13/2018 12:36 PM	8/13/2018 12:37 PM	00:33:00	Completed Successfully
whileman	PS_Students_SQL_Query		File	8/13/2018 12:39 PM	8/13/2018 12:41 PM	01:46:00	Completed Successfully

Figure 19. Run History Tab – SAS® Data Management Studio

LESSONS LEARNED

- Follow the steps in sequence.
- Do not wait until the end and then try to figure out the problem specific to your situation. Test your connections along the way (Steps 2,4,5,7,8,10).
- If you made it safely through Step 9 only to experience issues configuring the DataFlux® Batch Job in SAS® Data Integration Studio (Step 10), try using the sasinstaller account to logon to the SAS® compute tier where your Data Management Server is running. I had to resort to this approach to get any jobs to display in the Job Dropdown (see Figure 17). Also, make sure the SAS® Data Management service is started!!!
- If you do not succeed at first, contact SAS® Technical Support early on. They are great!

CONCLUSION

In this paper, I covered the 10 key steps in moving a data profile from SAS® Data Management Studio to SAS® Data Integration Studio. This move frees up resources on the client machine, resulting in resources being more readily available to SAS® Data Management Studio. With these available resources, one can focus on more productive tasks like developing new data profiles, explorations, batch jobs, processes, business rules, and monitoring your data quality across the organization. Also discussed are tips I learned along the way to make your first attempt at this endeavor more productive and enjoyable. There is nothing like learning by example.

REFERENCES

SAS Support, "DataFlux® Data Management Studio 2.7: User Guide - Deploying Profiles to a DataFlux® Data Management Server." Available at http://support.sas.com/documentation/onlinedoc/dfdmstudio/2.7/dmpdmsug/Content/dfDMStd_T_DMSErProfile.html

SAS Support, "DataFlux® Data Management Studio 2.7: User Guide - Accessing a DBMS Repository From the Server." Available at https://support.sas.com/documentation/onlinedoc/dfdmstudio/2.7/dmpdmsug/Content/dfU_T_ReposAdd_dbms_server.html

SAS Support, "DataFlux® Data Management Studio 2.6: User Guide - Performance Tips for Profiles." Available at http://support.sas.com/documentation/onlinedoc/dfdmstudio/2.6/dmpdmsug/Content/dfU_FAQ.html

SAS Support, "DataFlux® Data Management Server 2.7: Administrator's Guide." June 2017. Available at <http://support.sas.com/documentation/cdl/en/dmpdmsrvag/68260/PDF/default/dmpdmsrvag.pdf>

ACKNOWLEDGMENTS

The author thanks the SESUG chair and staff for this conference and the opportunity to present. The author also thanks Lisa Rich from SAS® Tech Support, for her support in helping me to overcome several obstacles encountered along the way, in getting my first data profile moved to and running on SAS® Data Management Server in both SAS® Data Management Studio and SAS® Data Integration Studio.

CONTACT INFORMATION

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

Wayne Hileman
whileman@ed.sc.gov

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.