

Paper 131-30

A Quick Tour of SAS® ETL Studio to Build a Data Mart

Kari Richardson, SAS Institute Inc., Cary, NC

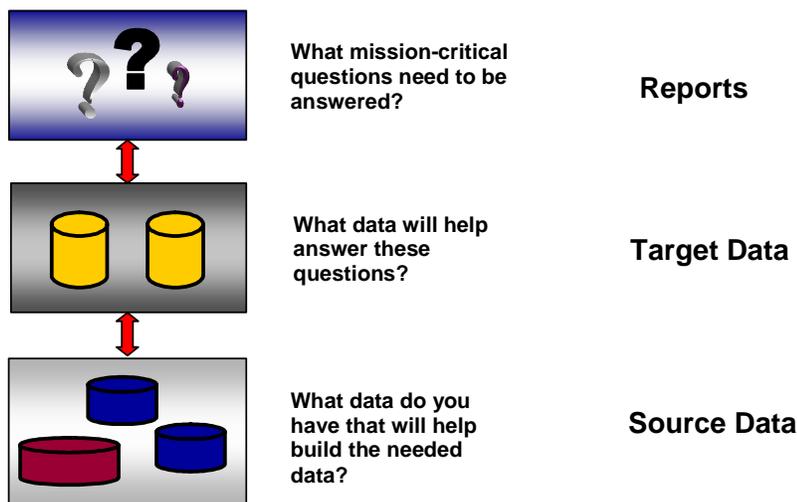
Eric Rosslund, SAS Institute Inc., Dallas, TX

ABSTRACT

This hands-on workshop will give users a basic tour through the functionality of SAS ETL Studio to help build a small data mart. The participants in this workshop will use SAS ETL Studio to define necessary library definitions, in addition to source and target tables definitions. Participants will create a process flow diagram using a simple transformation and load the target table. In the last step, participants will create two reports using the target table.

INTRODUCTION

Data marts are typically created by using a simple two-stage process: planning, which uses a top-down approach; and implementing, which is performed from the bottom up. The planning stage (top-down) begins with determining what mission-critical questions need to be answered, and what type of data is needed to build a report that will answer those questions. This data is often referred to as *target* data. Then, you determine what operational data will build the target data. This data is often referred to as *source* data. Careful planning will help you design a meaningful report that answers your mission-critical questions.



The implementation stage begins with defining the source data. Then, definitions are created for the target data, and the target data is loaded. Next, reports are generated using the target data.

SAS ETL Studio, a Java application, is a visual design tool that helps organizations quickly build, implement, and manage ETL processes from source to destination, regardless of the data sources or platforms. SAS ETL Studio enables you to perform the following tasks:

- **E**xtract data from operational data stores
- **T**ransform this data
- **L**oad the extracted data into your data warehouse or data mart.

BREAKDOWN OF STEPS FOR BUILDING A DATA MART

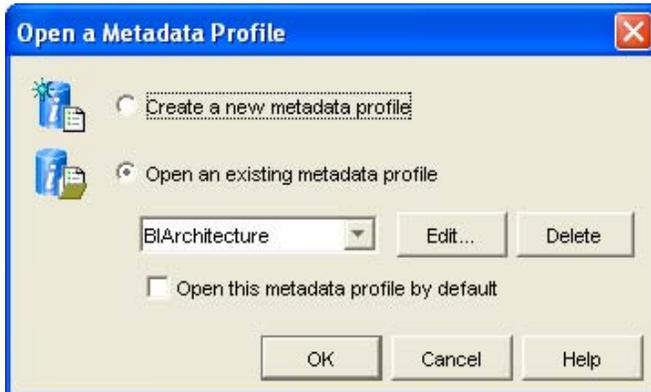
This paper navigates through five basic steps that can be used to build a data mart:

- Step 1: Define necessary libraries (for source and target data).
- Step 2: Define the source tables metadata.
- Step 3: Define the target table metadata.
- Step 4: Create and execute the job that will load the target from the sources.
- Step 5: Create the desired reports on the target data.

Launch SAS ETL Studio by selecting **Start** ⇒ **All Program** ⇒ **SAS** ⇒ **SAS ETL Studio**.

In the Open a Metadata Profile window, verify that **BIArchitecture** is selected and click

OK



Enter the Username and Password provided and click

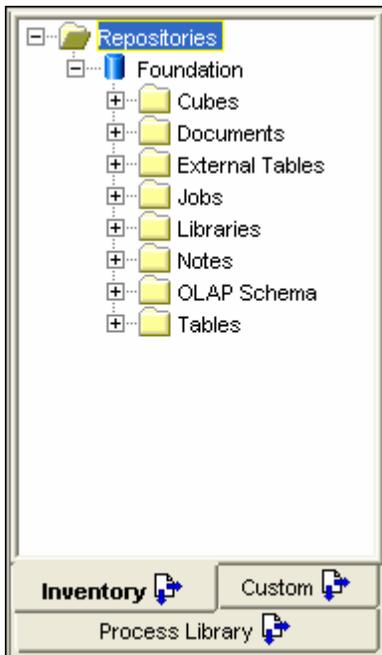
OK

STEP 1

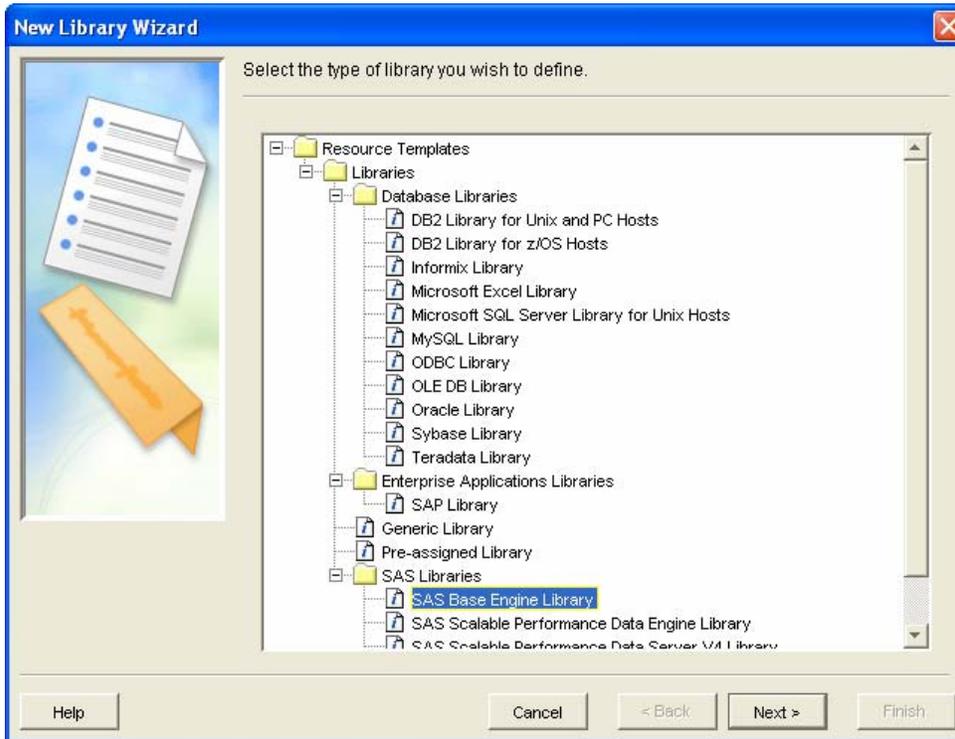
Define the necessary library objects. Two library definitions must be specified, one for the source tables and one for the destination table.

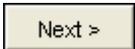
To define a library object in SAS ETL Studio:

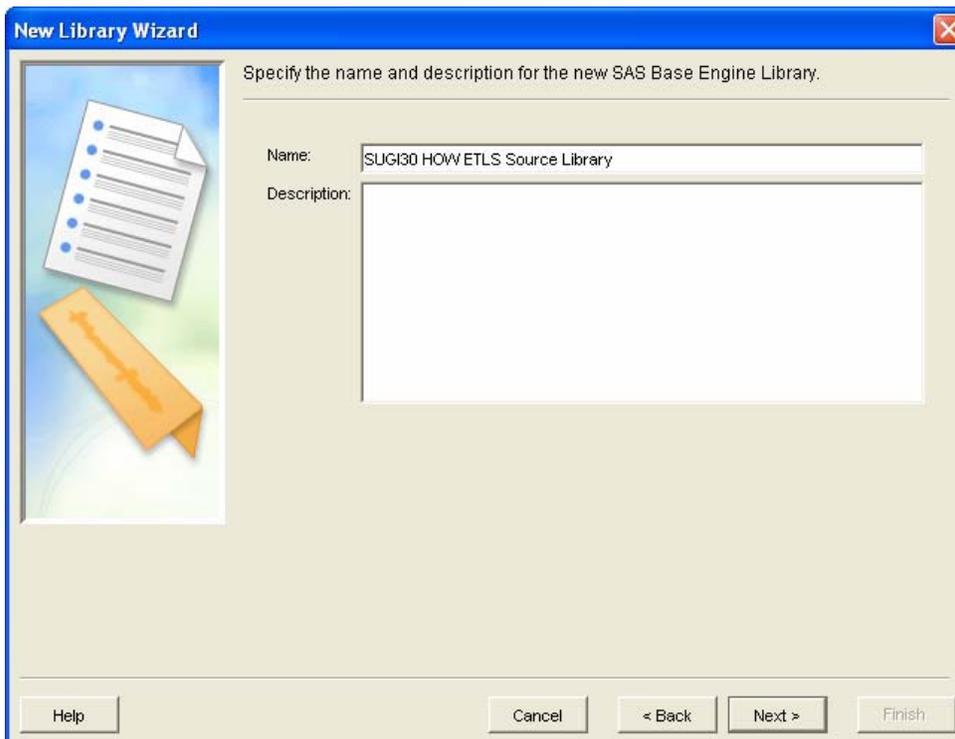
1. Choose the **Inventory** tab and expand the **Foundation** repository.

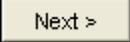


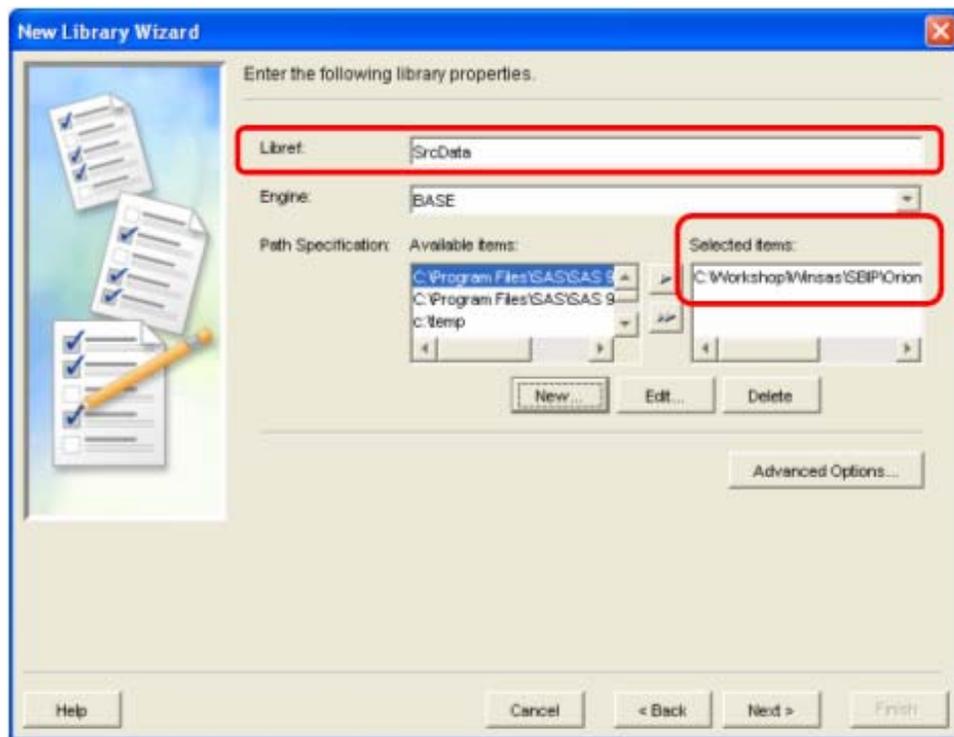
2. Right-click the **Libraries** folder and select **New...**. This invokes the New Library wizard.
3. Select **SAS Base Engine Library** as the type, then click .

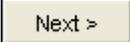


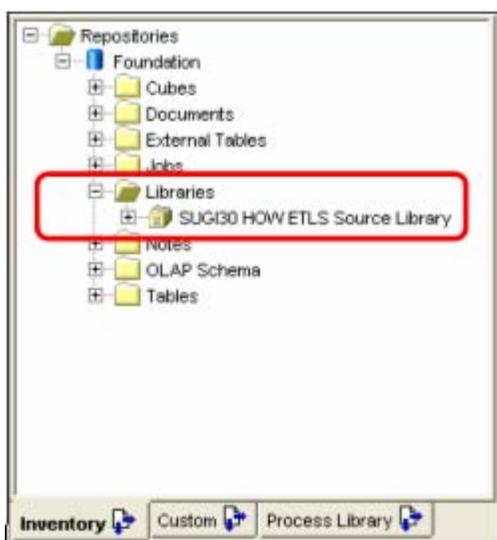
4. Enter **SUGI30 HOW ETLs Source Library** as the Name, then click .



- Specify a Libref of **SrcData** and a path of **C:\Workshop\Winsas\SBIP\Orion Star 15Aug03\ordetail** (the path should be available in the Available Items list). Click .



- Select **SASMain** as the server where the library is to be assigned. Then, click .
- Review the specifications and click .
- The Libraries folder should now have the newly defined library (Note: Other libraries will be defined on the workshop image).



9. Repeat Steps 1-8 to define an additional library with the following specifications:

Type:	SAS Base Engine Library
Name:	SUGI30 HOW ETLs Target Library
Libref:	Target
Path:	C:\Workshop\Winsas\SBIP\Warehouse
SAS Server:	SASMain

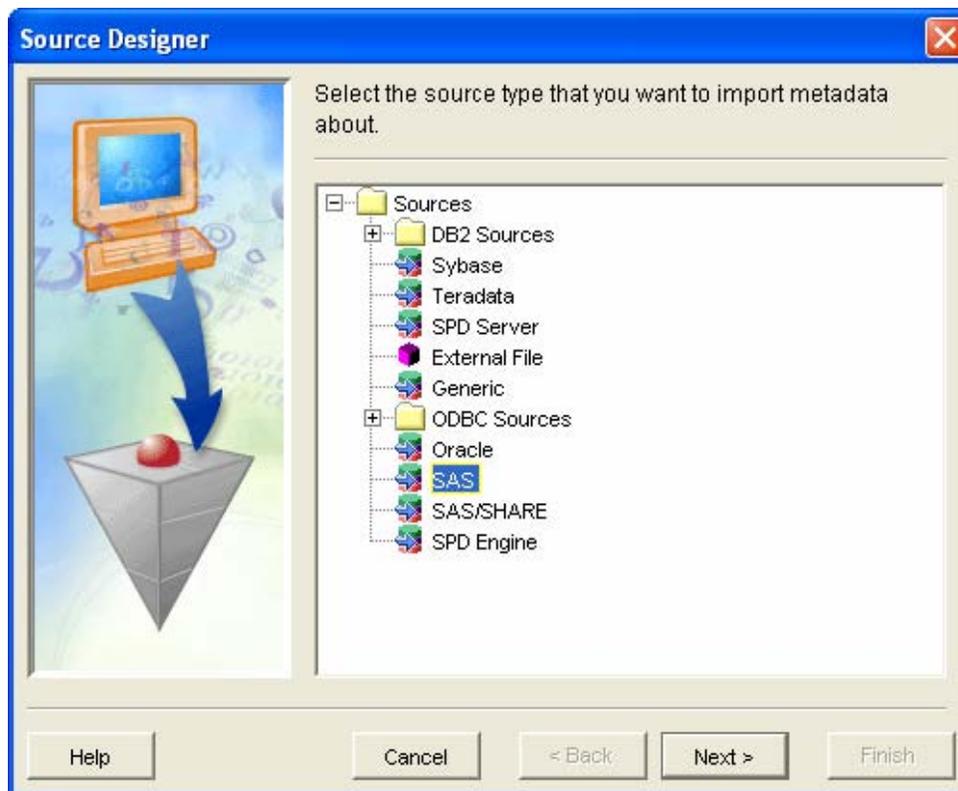
STEP 2

Define the table objects that represent source tables—two tables must be specified, the STAFF and ORGANIZATION tables.

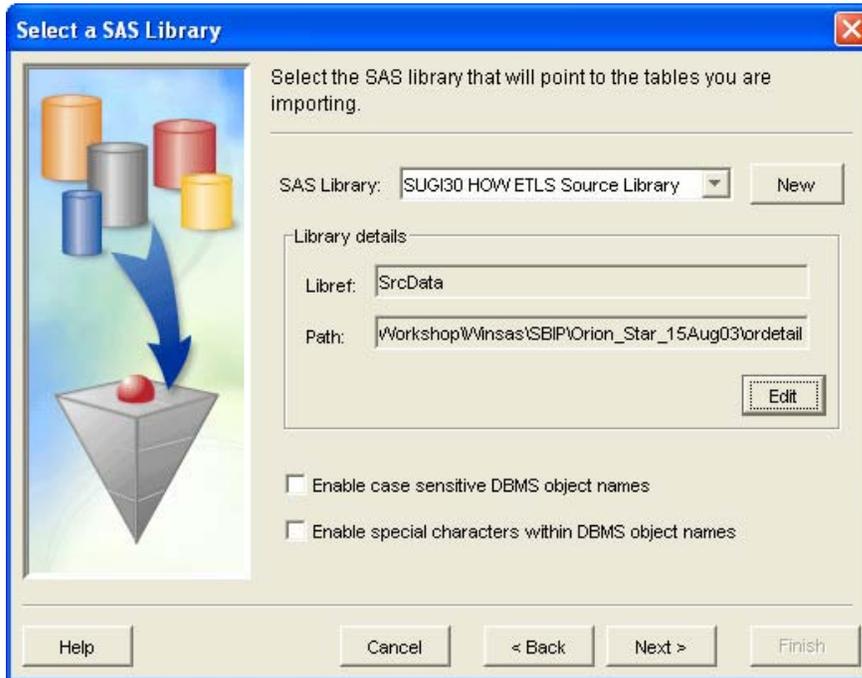
1. Select **Tools** ⇒ **Source Designer** from the menu bar.

2. Choose **SAS** as the source type that you want to import metadata about, then click

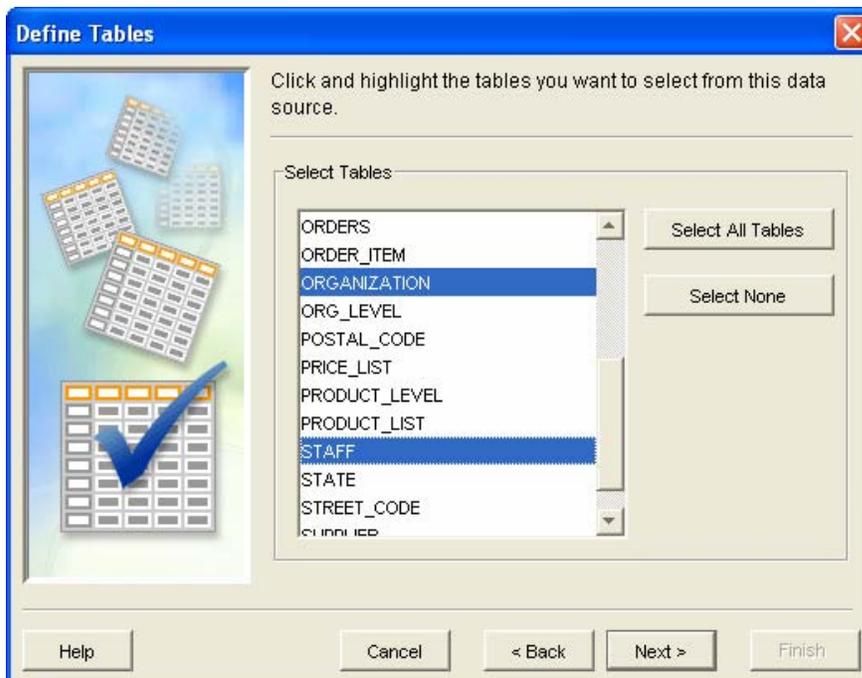
Next >



3. Select **SUGI30 HOW ETL Source Library** as the library that contains the tables to import metadata from, then click **Next >**.



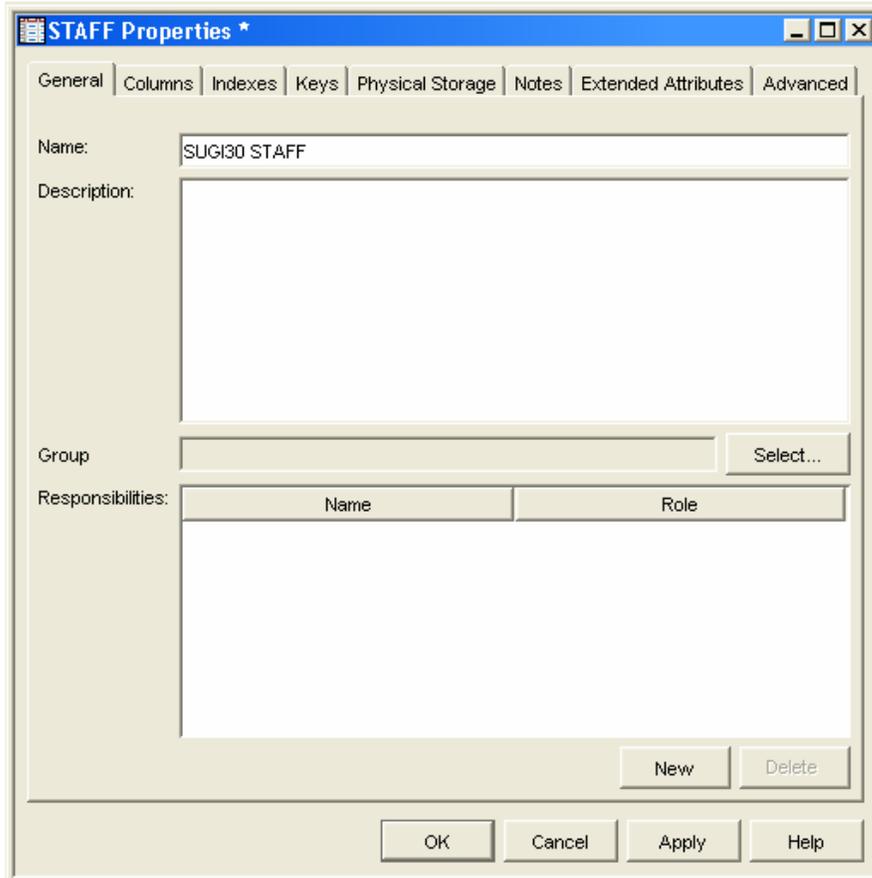
4. Select both **STAFF** and **ORGANIZATION** tables, then click **Next >**.



5. Keep the default selection for grouping on the Custom Tree. Click **Next >**.
6. Review the selection information and click **Finish**.

So that you don't confuse these newly defined table objects with others, the next step will be to rename the table objects:

7. Select the **Inventory** tab and expand the **Tables** folder.
8. Right-click the **STAFF** table and select **Properties**.
9. On the **General** tab (in the Properties window), change Name to **SUGI30 STAFF**.



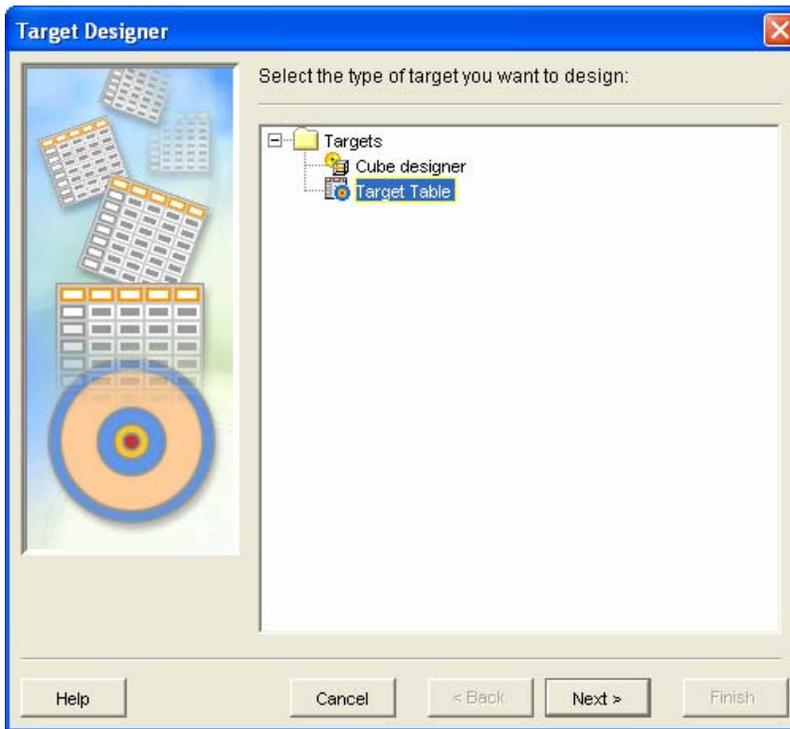
10. Click to save the change.

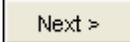
11. Repeat steps 7-10 to rename the **ORGANIZATION** table to **SUGI30 ORGANIZATION**.

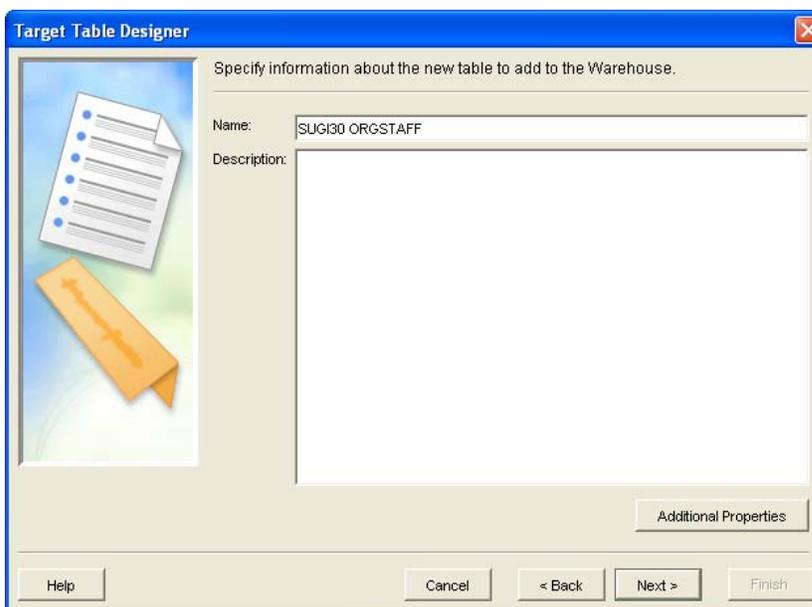
STEP 3

Define the table object that represents the target table.

1. Select **Tools** ⇒ **Target Designer** from the menu bar.
2. Choose **Target Table** as the target type. Click .



3. Specify **SUGI30 ORGSTAFF** as the Name of the object. Click .



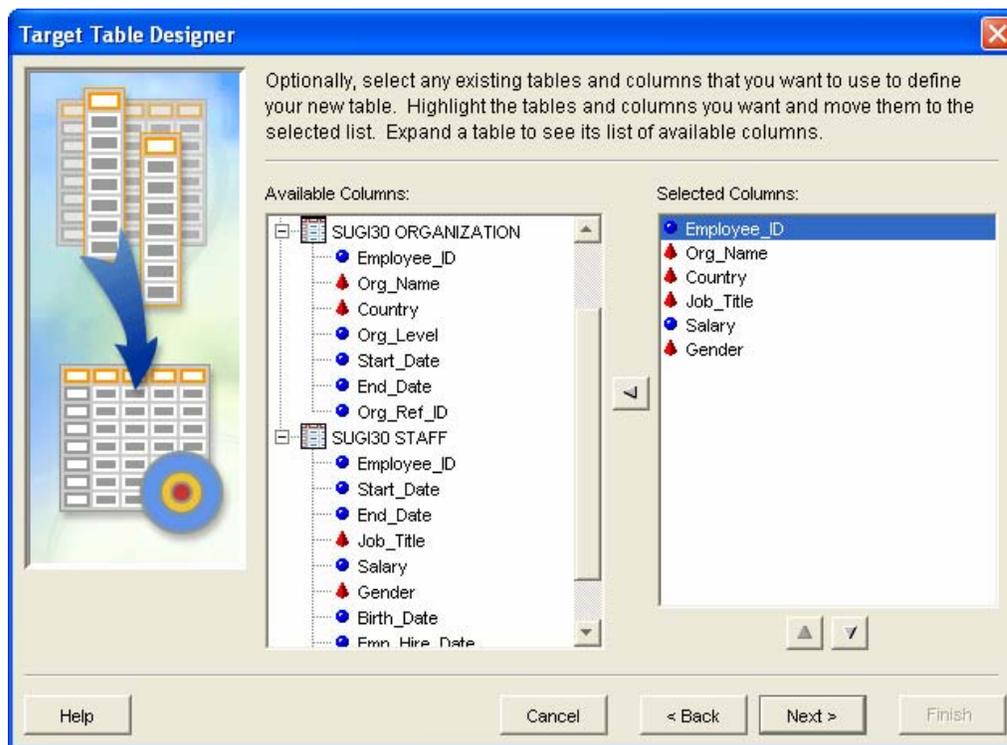
4. Expand the **SUGI30 ORGANIZATION** table and select the following variables:

- Employee_ID
- Org_Name
- Country

Similarly, from the **SUGI30 STAFF** table, select the following variables:

- JobTitle
- Gender
- Salary

When you've finished selecting the above columns, click .

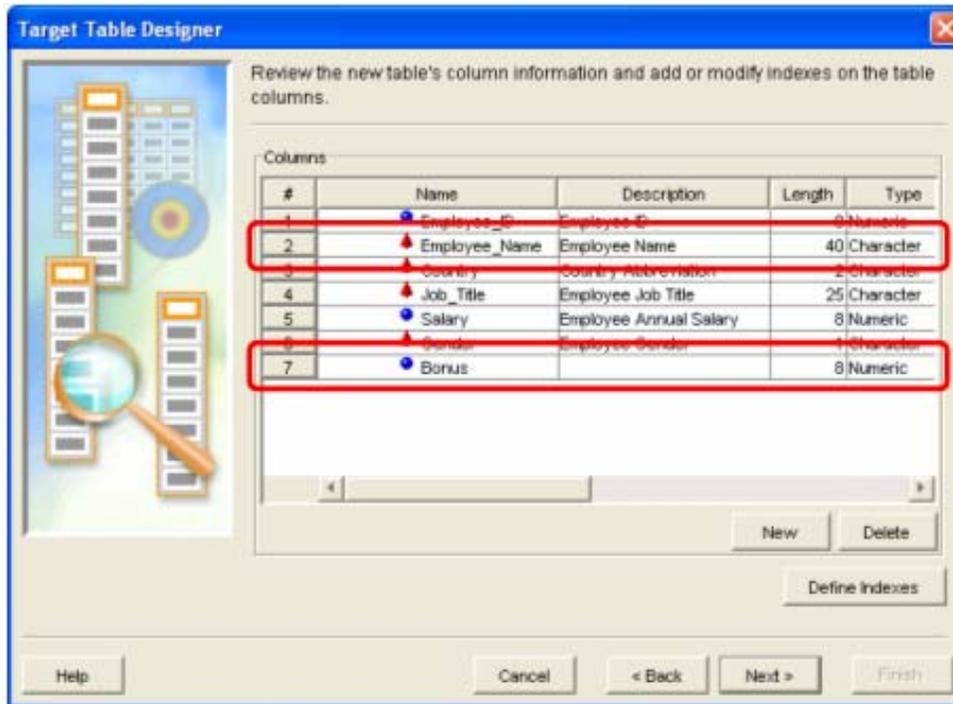


5. Change the attributes for ORG_NAME column:
 - ORG_NAME should be EMPLOYEE_NAME
 - Description should be Employee Name

Also, define a new column, Bonus, that has the following attributes:

- Numeric
- Default length of 8
- Format of Dollar8

When you've finished modifying the column information, click Next >.



6. Specify a library of **SUGI30 HOW ETLs Target Library**. Click **Next >**.

Target Table Designer

Specify physical storage information about your new target table.

DBMS: SAS

Library: SUGI30 HOW ETLs Target Library

Name: SUGI30_ORGSTAFF

Schema:

Enable case sensitive DBMS object names

Enable special characters within DBMS object names

Table Options

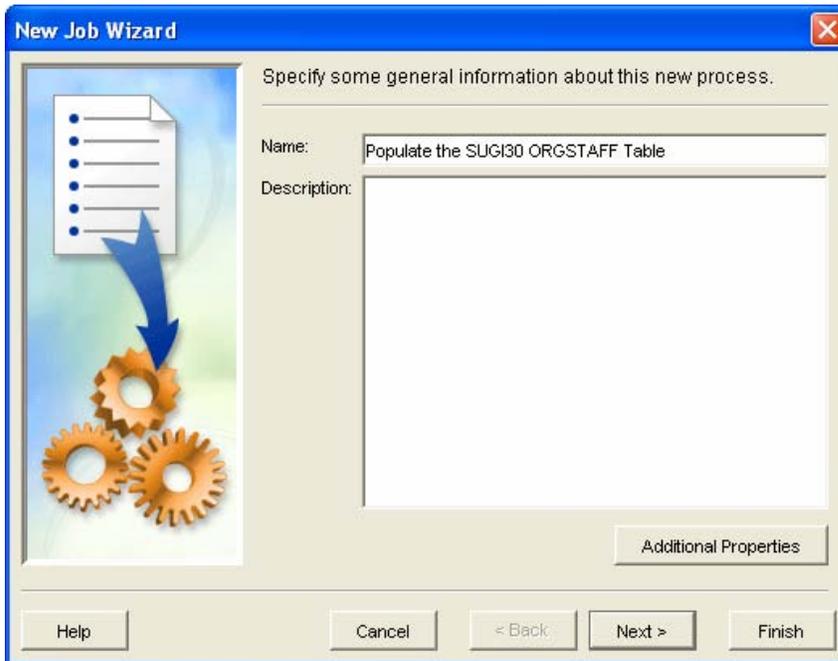
Help Cancel < Back Next > Finish

7. Review the specifications and click **Finish**.

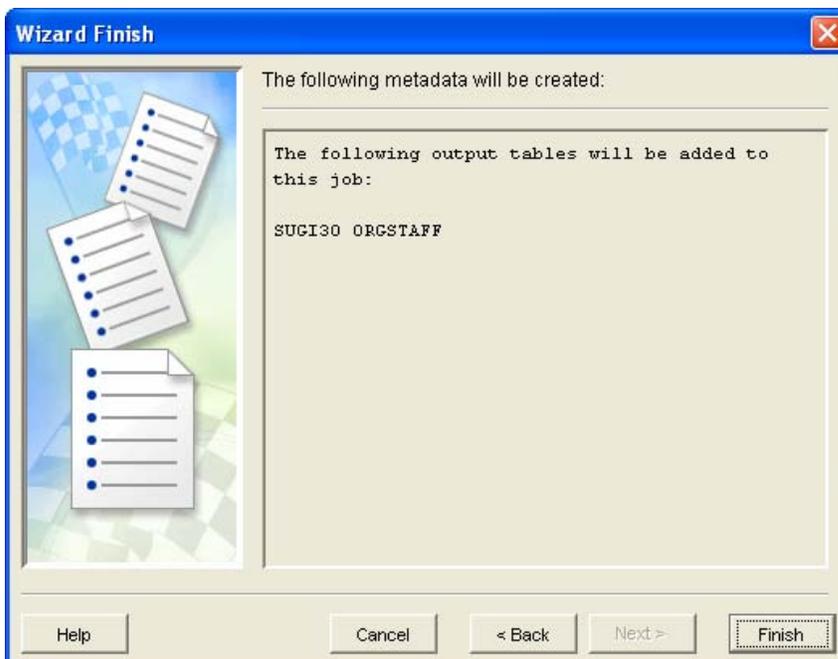
STEP 4

Specify metadata for the process flow that's needed for loading the target table:

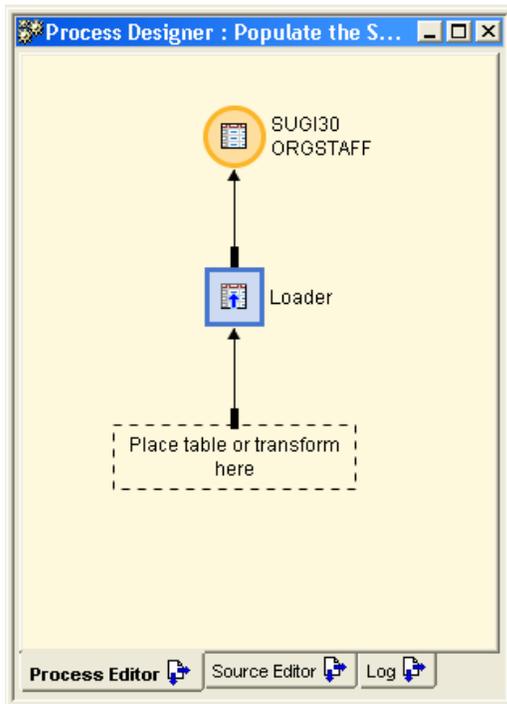
1. Select **Tools** ⇒ **Process Designer** from the menu bar to launch the New Job Wizard.
2. Enter a Name of **Populate the SUGI30 ORGSTAFF Table**, then click .



3. Select **SUGI30 ORGSTAFF** as the table to load, then click .
4. Review the specifications and click .

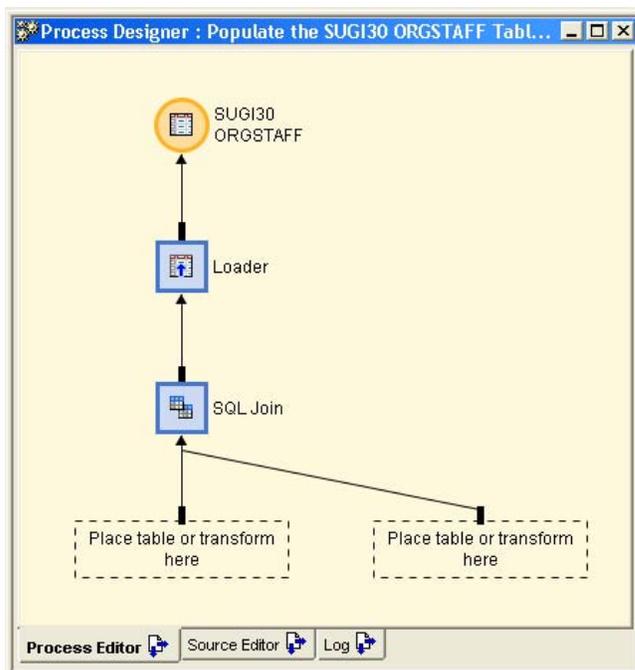


This launches the Process Designer window:

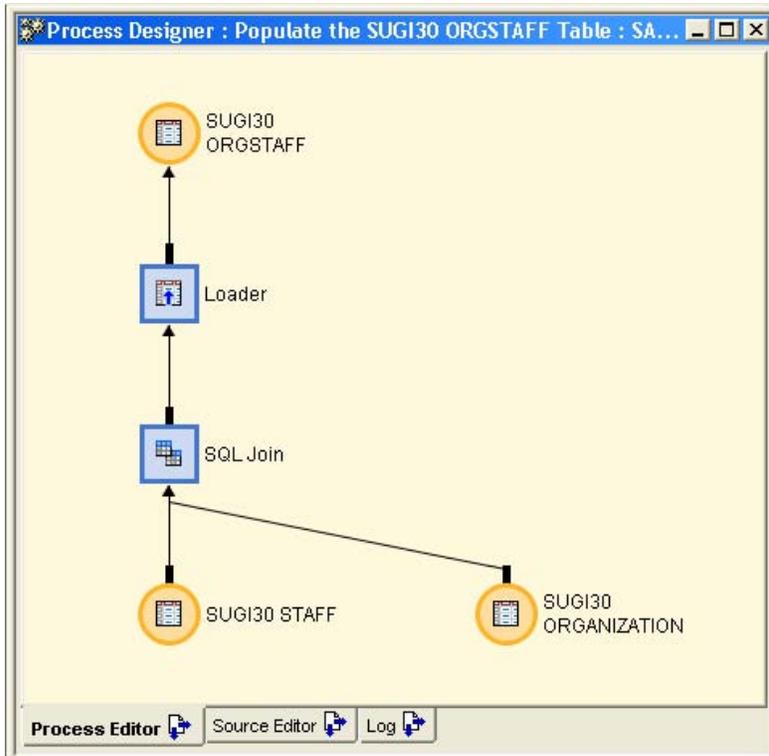


To complete the process flow diagram:

5. Select the **Process Library** tab from the Tree View area of SAS ETL Studio and expand the **Data Transforms** folder.
6. Locate the **SQL Join** transformation. Drag the **SQL Join** transformation to the process flow diagram, and drop it onto the template icon labeled "**Place table or transform here**".



7. Select the **Inventory** tab and expand the **Tables** folder.
8. Locate the table object for **SUGI39 STAFF** and drag it to one of the “**Place table or transform here**” templates.
9. Locate the table object for **SUGI30 ORGANIZATION** and drag it to the remaining “**Place table or transform here**” template.



Attributes now need to be specified for the SQL Join transformation and for the Loader process.

To specify attributes for the SQL Join transformation:

10. **Right-click** the **SQL Join** icon and select **Properties**.
11. Select the **Mapping** tab. On the Target table side of the Mapping tab, **right-click** and select **Import Columns**.
12. In the Import Columns window, expand **Foundation** and the **Ungrouped** group.
13. Expand the **SUGI30 ORGSTAFF** table, and move the **Bonus** column to the *Selected* area.

To specify the calculation for the Bonus column:

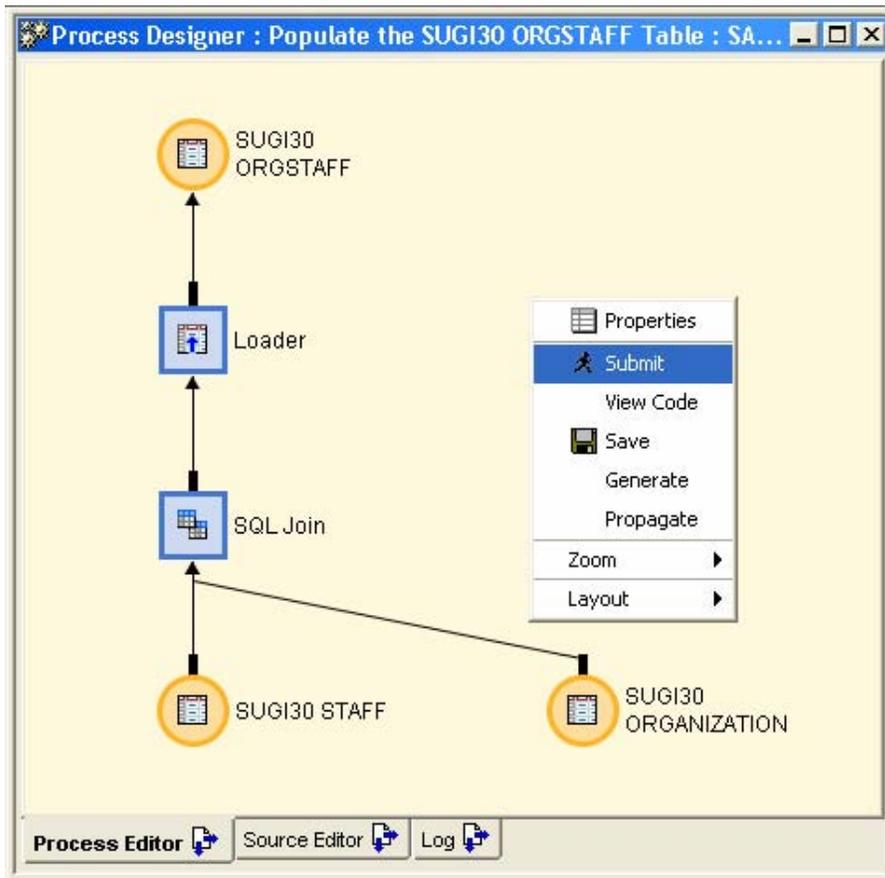
14. **Right-click** on the newly imported **Bonus** column and select **Expression**.
15. In the Expression Text area, enter **.06*salary**.
16. Close the Properties window to save the changes.

To specify attributes for the Loader process:

17. **Right-click** the **Loader** icon and select **Properties**.
18. Select the **Mapping** tab. On the Target table side of the Mapping tab, **right-click** and select **Quick Map** (this should map the Bonus column).
19. Map the **Org_Name** column to the **Employee_Name** column by dragging.
20. Close the Properties window to save the changes.

To run the job to load the SUGI30 ORGSTAFF table:

21. **Right-click** in the background of the Process Editor tab in the Process Designer window and click **Submit**.



22. When the job finishes executing, select the **Log** tab in the Process Designer window.
23. Verify that the job ran successfully and that TARGET.SUGI30_ORGSTAFF with 1048 Observations and 7 variables was created.

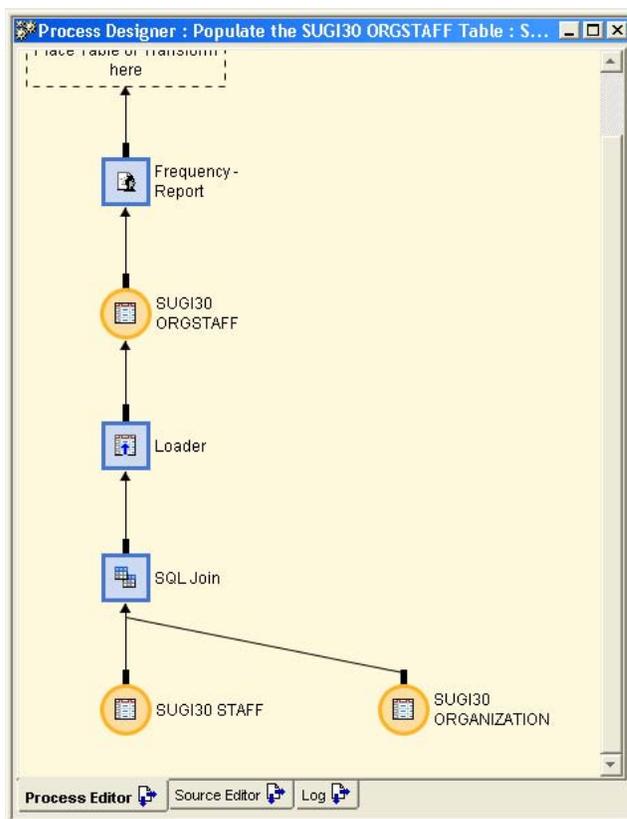
STEP 5

Human Resources needs two reports created.

- The first report should show the breakdown by gender of the employees across the countries. Also, the total number of employees in each country is needed, in addition to the total number of females and males.
- The second report analyzes salary and bonus for the gender values in each country. Minimum, maximum, and average statistics have been requested.

To create the first report of frequencies:

1. Select the **Process Library** tab from the Tree View area of SAS ETL Studio.
2. Expand the **Analysis** folder and locate the **Frequency - Report** transformation.
3. Drag the **Frequency - Report** transformation to the process flow diagram and drop it into empty space.
4. Drag the **SUGI30 ORGSTAFF** icon from the process flow diagram onto the “**Place table or transform here**” icon that leads to the **Frequency - Report** transformation.



To specify attributes for the frequency report:

5. **Right-click** the **Frequency - Report** icon and select **Properties**.
6. Select the **Options** tab.
7. Specify **Country*Gender** for the **TABLE statement** option.
8. Specify **NOCUM** for the **Options on the TABLE statement** option.
9. Select the **Column Options** tab.
10. Select the **Frequency Columns (TABLE statement)** option.
11. Move both **Country** and **Gender** to the Selected Columns list.
12. Close the Properties window.

To view the report, display the Output tab in the Process Designer window by performing the following tasks:

13. Click the **Options** icon from the Shortcut bar.
14. On the General tab, select **Show Output tab in Process Designer**.
15. Close the Options window.
16. Re-run the job (right-click in the background of the Process Editor tab and click **Submit**).
17. View the report by selecting the Output tab.

The SAS System 23

The FREQ Procedure

Table of Country by Gender

Country(Country Abbreviation)	Frequency	Percent	Row Pct	Col Pct	Female	Male	Total
Australia	48	4.58	44.44	10.15	60	5.73	108
Belgium	27	2.58	45.00	5.71	33	3.15	60
Germany	45	4.29	45.92	54.08	53	5.06	98

Process Editor | Source Editor | Log | Output

The last step is to create the report that analyzes salary and bonus. For this report, it will be necessary to sort the data first, and then create the desired report on the sorted data. So this report writing task will be broken into three parts:

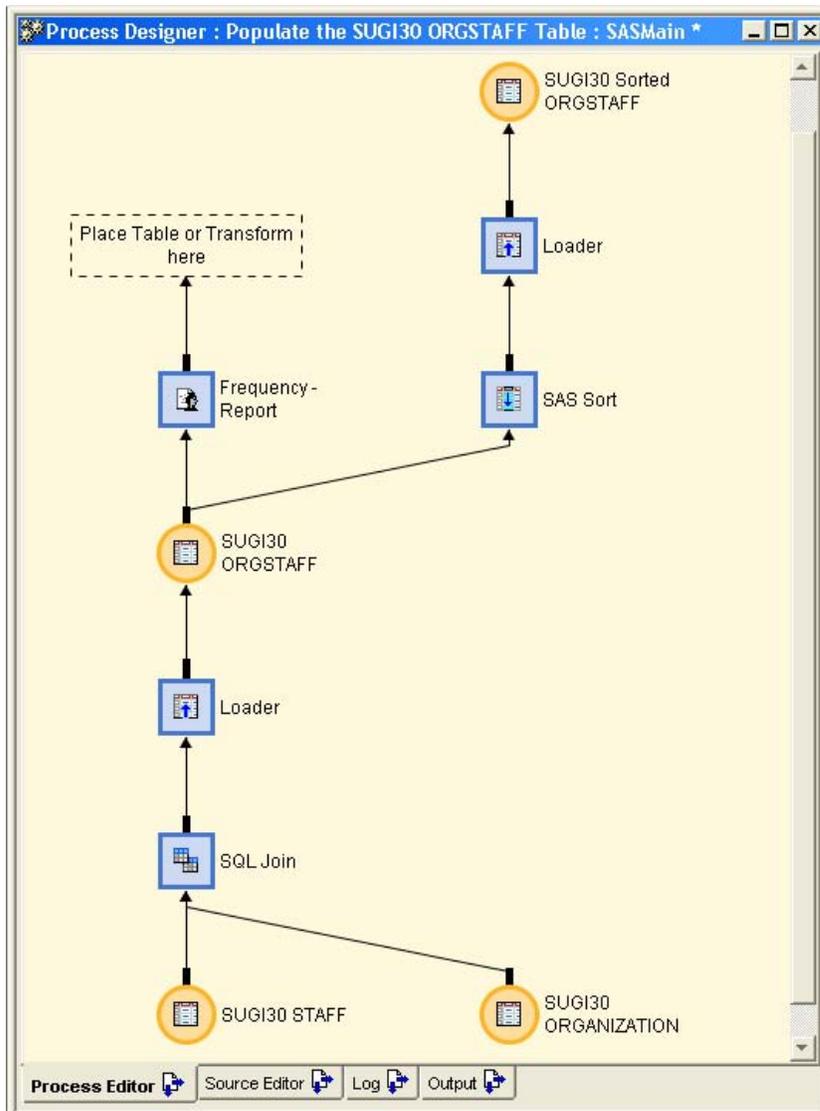
- Part 1** – Create the target table for the sorted data.
- Part 2** – Add the **SAS Sort** transformation.
- Part 3** – Add the **Summary Statistics – Report** transformation.

To create the needed target table:

1. Click the **Target Designer** icon from the Shortcut bar.
2. Select **Target Table** as the target type.
3. Specify **SUGI30 sorted ORGSTAFF** for the Name of the object.
4. Select all the variables from the **SUGI30 ORGSTAFF** table.
5. Select **SUGI30 HOW ETL Target Library** as library.
6. Specify **sortedOrgStaff** as the Name for the physical table.
7. Review the results and Finish.

To add the **SAS Sort** transformation:

1. Drag the **SUGI30 Sorted OrgStaff** table object to the process flow diagram and drop it into empty space.
2. Select the **Process Library** tab from the Tree View area of SAS ETL Studio.
3. Expand the **Data Transforms** folder and locate the **SAS Sort** transformation.
4. Drag the **SAS Sort** transformation to the process flow diagram and drop it onto the template icon labeled “**Place table or transform here**” (the template flowing into the Loader for the **SUGI30 Sorted OrgStaff** table).
5. Drag the **SUGI30 OrgStaff** table object (already in the process diagram) to the “**Place table or transform here**” template icon that flows into the **SAS Sort** transformation.

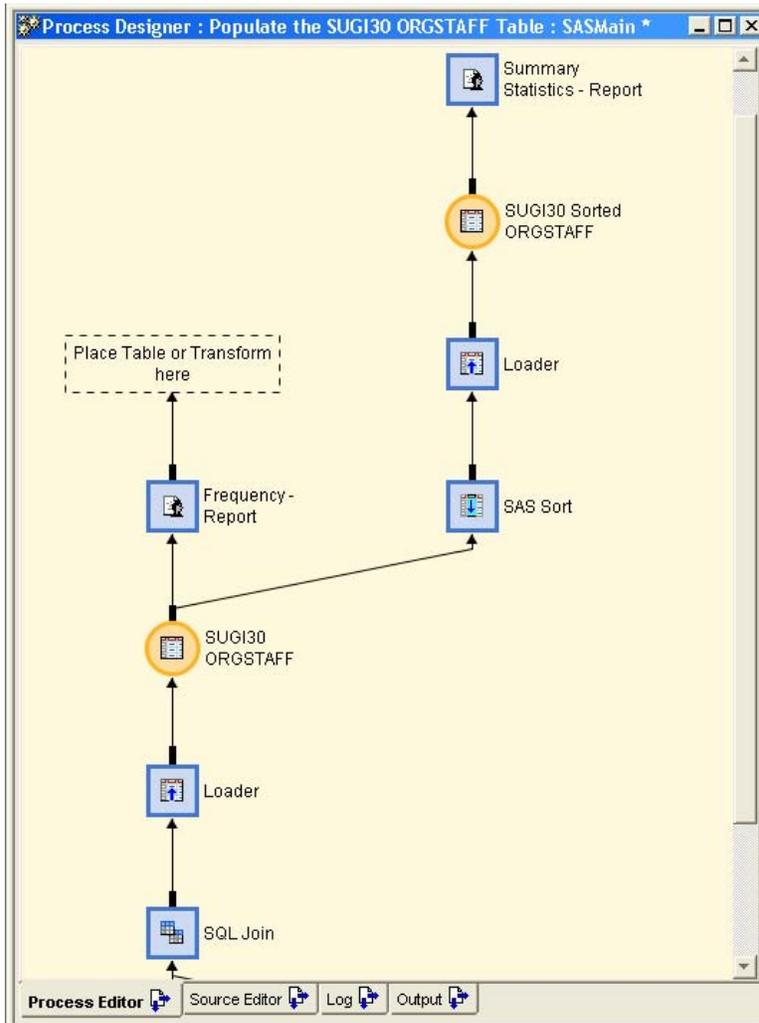


To specify properties for the **SAS Sort** transformation:

6. **Right-click** the **SAS Sort** icon and select **Properties**.
7. Select the **Sort By Columns** tab.
8. Move **Country** and **Gender** to the selected list (verify that **Country** is first in the selected list).
9. Close the Properties window.

To add the **Summary Statistics - Report** transformation:

1. Select the **Process Library** tab from the Tree View area of SAS ETL Studio.
2. Expand the **Analysis** folder and locate the **Summary Statistics - Report** transformation.
3. Drag the **Summary Statistics - Report** transformation to the process flow diagram and drop into empty space.
4. Drag the **SUGI30 Sorted OrgStaff** table object (already in the process flow diagram) to the "Place table or transform here" template icon that flows into the **Summary Statistics - Report** transformation.



5. **Right-click** the **Summary Statistics - Report** icon and select **Properties**.
6. Select the **Options** tab.
7. Specify **MIN MAX MEAN** for the **Statistics (separated by spaces)** option.
8. Select the **Column Options** tab.
9. Select the **Grouping columns (BY statement)** option.
10. Move both **Country** and **Gender** to the Selected Columns list (place **Country** first!).
11. Select the **Analysis columns (VAR statement)** option.
12. Move both **Salary** and **Bonus** to the Selected Columns list.
13. Close the Properties window.

To view the report:

1. Re-run the job (right-click and **Submit**).
2. View the report(s) on the Output tab.

The SAS System 23:45 Sunday

----- Country Abbreviation=Australia Employee Gender=Female -----

The MEANS Procedure

Variable	Label	Minimum	Maximum	Mean
Salary	Employee Annual Salary	24015.00	46230.00	27961.25
Bonus		1440.90	2773.80	1677.68

----- Country Abbreviation=Australia Employee Gender=Male -----

Variable	Label	Minimum	Maximum	Mean
Salary	Employee Annual Salary	24100.00	163040.00	33288.75
Bonus		1446.00	9782.40	1997.33

----- Country Abbreviation=Belgium Employee Gender=Female -----

Variable	Label	Minimum	Maximum	Mean
Salary	Employee Annual Salary	25240.00	156245.00	32342.22
Bonus		1514.40	9374.70	1940.53

Process Editor Source Editor Log Output

CONCLUSION

This hands-on workshop provides an overview of SAS ETL Studio and how it can be used to create a data mart.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors:

Kari Richardson
 SAS Institute Inc.
 SAS Circle
 Cary, NC 27513
 E-mail: kari.richardson@sas.com

Eric Rossland
 SAS Institute Inc.
 15455 N. Dallas Parkway
 Suite 1300
 Addison, TX 75001
 E-mail: eric.rossland@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.