

# Using SAS Enterprise Guide as a MS SQL interface for data analytics

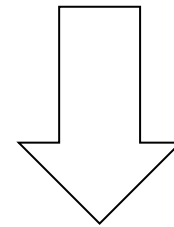
Paulus Mau, BC Hydro, Regulatory and Rates  
SAS EG User for data analytics

May 2013

1. Comparison

2. How - to

The SAS logo, consisting of a blue stylized 'S' followed by the lowercase letters 'sas' in a black sans-serif font.



The Microsoft logo, featuring the word 'Microsoft' in a bold, italicized, black sans-serif font with a registered trademark symbol (®) at the end.

## Common Analytics using Data from SQL database:

- Filtering
  - Grouping
  - Summary Statistics (Totals averages, Min, Max)
  - Joining Tables
  - Looking at differences
  - Transposing data (“Pivot Tables”)
- 
- Illustrated through a common business questions:  
“Please send me a one-page summary about our customers.”

# The Datasets and Structure

2012 and 2011 energy consumption data by account (fictional, annual tables)

	Random_ID	MONTH	YEAR	days_billed	Monthly_consumpti	Billing_Flag
▶	5483	4	2012	30	13320	N
	5483	5	2012	31	101412.413795	N
	5483	6	2012	30	70245.51724	N
	5483	7	2012	31	109570.909092	N
	5483	8	2012	31	136636.363635	N
	5483	9	2012	30	113880	N
	5483	10	2012	31	108054.375	N
	5483	11	2012	30	159168.75	N
	5483	12	2012	31	79350	N
	5483	1	2012	31	297716.25	N
	5483	2	2012	29	36000	N
	5483	3	2012	31	117855	N
	5484	4	2012	30	85800	N

## Features:

Each Row represents a monthly bill for each account.

There are 12 monthly bills in 1 year.

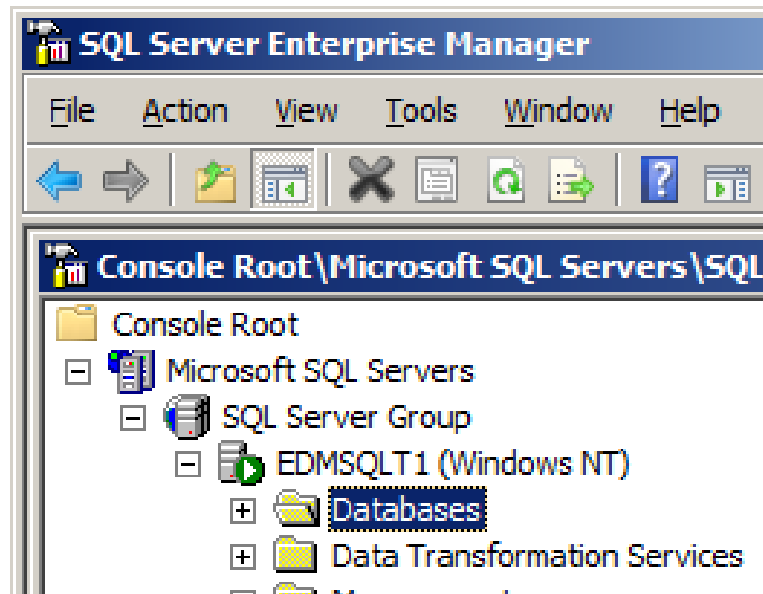
“Random\_ID” = Customer Account Number (fictional).

Other Columns: Month, Fiscal Year, Days billed, consumption amount, and a billing flag

## Microsoft

### SQL: SQL Server Enterprise manager

#### 1. Connect



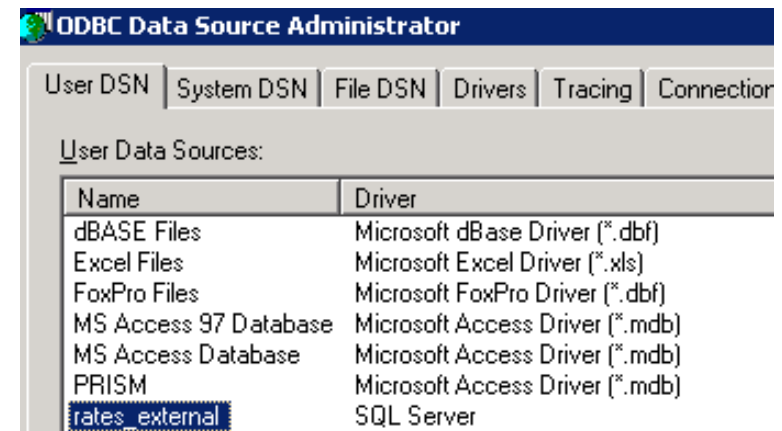
#### 2. Go down the tree

#### 3. Use the GUI provided by Microsoft to do the Querying

## SAS

### SAS via Open Data Base Connectivity

#### 1. Setup ODBC using SQL Server Driver



#### 2. Open EG, assign dataset to project library

#### Code window:

```
LIBNAME <any name> ODBC datasrc=<name created in ODBC> schema= <database owner>
```

```
LIBNAME FREDAPMA ODBC datasrc=rates_external  
schema= dbo
```

#### 3. Run the code via process flow

#### 4. Data linked as a library.

# Comparison of Business Analytics

SAS EG without coding  
vs.  
MS SQL Server Enterprise Studio GUI

	Random_ID	MONTH	YEAR	days_billed	Monthly_consumpti	Billing_Flag
▶	5483	4	2012	30	13320	N
	5483	5	2012	31	101412.413795	N
	5483	6	2012	30	70245.51724	N
	5483	7	2012	31	109570.909092	N
	5483	8	2012	31	136636.363635	N
	5483	9	2012	30	113880	N
	5483	10	2012	31	108054.375	N
	5483	11	2012	30	159168.75	N
	5483	12	2012	31	79350	N
	5483	1	2012	31	297716.25	N
	5483	2	2012	29	36000	N
	5483	3	2012	31	117855	N
	5484	4	2012	30	85800	N

# “quick summary of our 2012 customers”



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit
Total population consumption	Y	Y	
Average Consumption of each account by Month	Y	Y	
Top 10 accounts by annual consumption	Y	Need code editing to sum up months	Y (subquery required)
Max Customer	Y	Need code editing to sum up months	Y (subquery required)
Min Customer	Y	Need code editing to sum up months	Y (subquery required)

# “Please Compare. What’s the Yr/Yr trend for accounts that exist in both years?” (join tables)



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit
Total Consumption of population	Y	Y (create views)	
Average Consumption by Month	Y	Y (create views)	
consumption change for the top 10 accounts in 2012	Y		Y (views and/or complex subquery required)
Max Customer	Y		Y (views and/or complex subquery required)
Min Customer	Y		Y (views and/or complex subquery required)



# “Generate Pivot table so I can see monthly consumptions across the columns.”



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit	Export to Excel
Pivot Data	Y via Transpose	No	Very complex, need to have prior knowledge of columns	Small data yes. Big data, no.

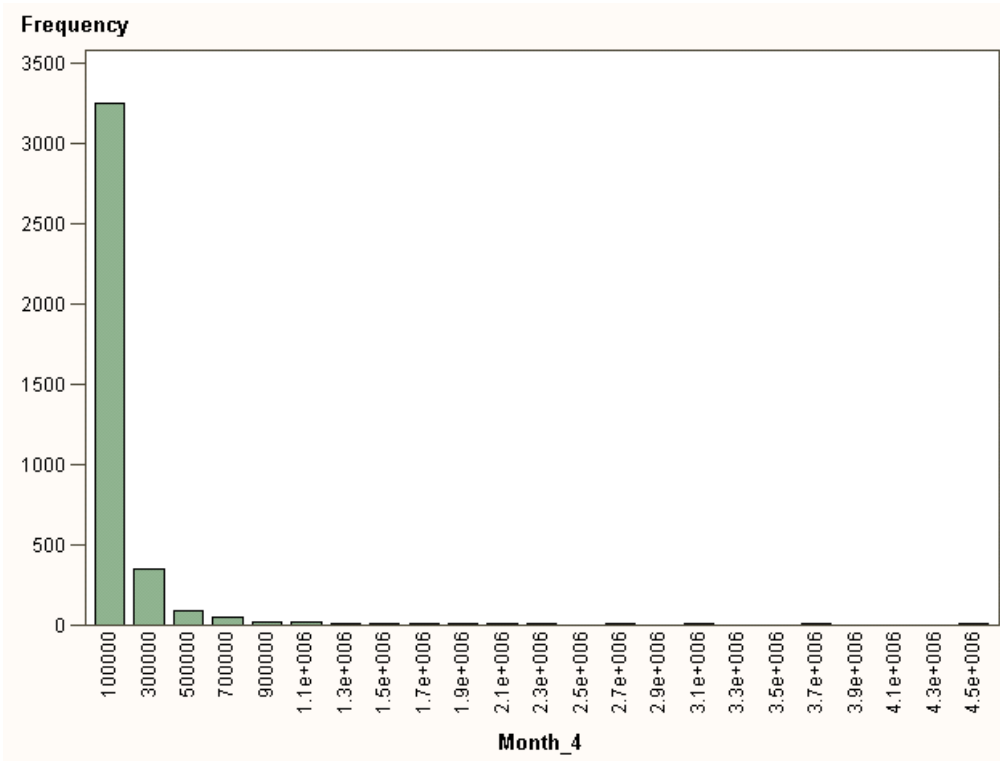
Random_ID	MONTH	YEAR	days_billed	Monthly_consumpti	Billing_Flag
5483	4	2012	30	13320	N
5483	5	2012	31	101412.413795	N
5483	6	2012	30	70245.51724	N
5483	7	2012	31	109570.909092	N
5483	8	2012	31	136636.363635	N
5483	9	2012	30	113880	N

Random_ID	Month_4	Month_5	Month_6	Month_7	Month_8	Month_9	Month_10
1	124560.00000	136327.50000	138985.60345	152822.89655	155631.27273	140432.72727	133165.71429
2	64320.00000	65408.00000	58372.00000	31068.00000	34569.93103	48992.06897	59606.00000
3	38880.00000	40485.51724	41081.37931	44134.92163	44193.77504	42833.89831	43710.50847
4	141600.00000	153348.38710	144731.61290	144913.54839	145200.00000	150526.45161	154080.00000

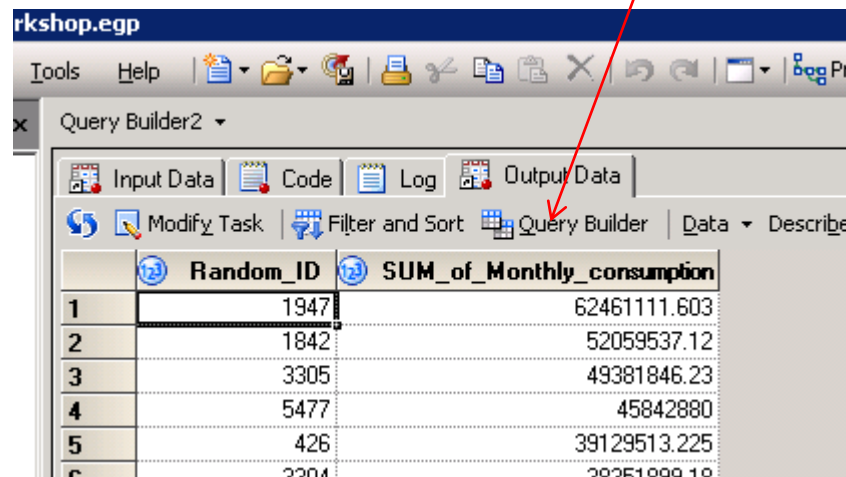
# “Please Graph the annual trends”



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit	Export to Excel
Direct Graphing	Y	Query and Export	2 hrs later	Graph Data



Tool to use: [Query Builder](#)



The screenshot shows the SAS EG Query Builder interface. The window title is 'rkshop.egp'. The menu bar includes 'Tools' and 'Help'. The toolbar contains various icons for file operations and execution. The main area has tabs for 'Input Data', 'Code', 'Log', and 'Output Data'. Below the tabs are icons for 'Modify Task', 'Filter and Sort', 'Query Builder', and 'Data'. A table is displayed with the following data:

	Random_ID	SUM_of_Monthly_consumption
1	1947	62461111.603
2	1842	52059537.12
3	3305	49381846.23
4	5477	45842880
5	426	39129513.225
6	2204	20251000.10

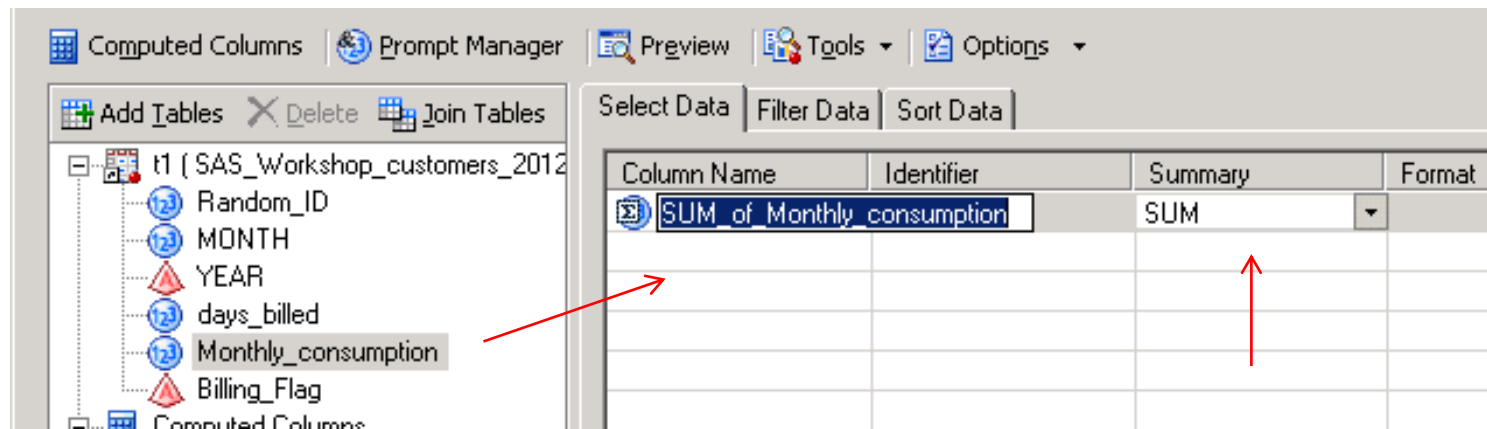
1. Descriptive Stats – Sum
2. Descriptive Stats – Average by Month
3. Calculating change, requiring table joins
4. Pivot Table

# “quick summary of our 2012 customers”

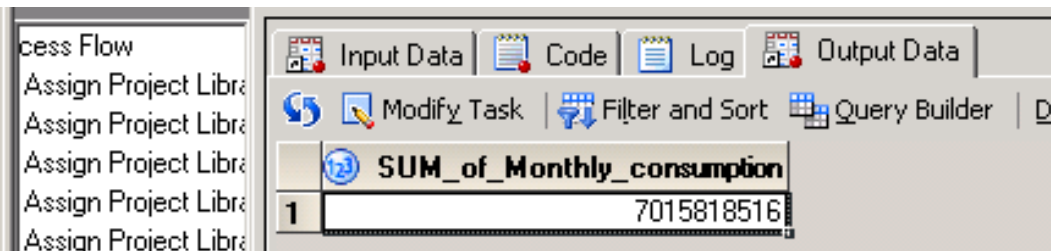


Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit
Total population consumption	Y	Y	

Query Builder:



Run!



# “quick summary of our 2012 customers”



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit
Average Consumption by Month	Y	Y	

Using Query Builder: Select Data:

The screenshot shows the SAS Query Builder interface. On the left, a tree view shows the table 't1 (SAS\_Workshop\_customers\_2012)' with columns: Random\_ID, MONTH, YEAR, days\_billed, Monthly\_consumption, and Billing\_Flag. Below it are 'Computed Columns' including '\_Calculation', '\_Calculation1', and '\_Calculation2'. On the right, the 'Select Data' tab is active, displaying a table with columns: Column Name, Identifier, and Summary. The table contains two rows: 'MONTH (MONTH)' with identifier 't1.MONTH' and 'AVG\_of\_Monthly...' with identifier '\_Calculation2' and summary 'AVG'. At the bottom, the 'Summary groups' section has a checked box for 'Automatically select groups' and a list containing 't1.MONTH'. Red arrows point from the 'MONTH' column in the tree to the 'MONTH (MONTH)' row, from 'Monthly\_consumption' to the 'AVG\_of\_Monthly...' row, and from the 'Automatically select groups' checkbox to the 't1.MONTH' entry in the summary groups list.

# “quick summary of our 2012 customers”



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit
Average Consumption by Month	Y	Y	

Sort

The screenshot shows the SAS Query Builder interface for a local connection. The 'Sort Data' tab is active, and the 'MONTH' column is selected for sorting in ascending order. Red arrows point to the 'Sort Data' tab, the 'MONTH' column in the table, and the 'Ascending' dropdown menu.

Column Name	Identifier	Sort Direction
MONTH	t1.MONTH	Ascending

# “quick summary of our 2012 customers”



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit
Average Consumption by Month	Y	Y	

Run!

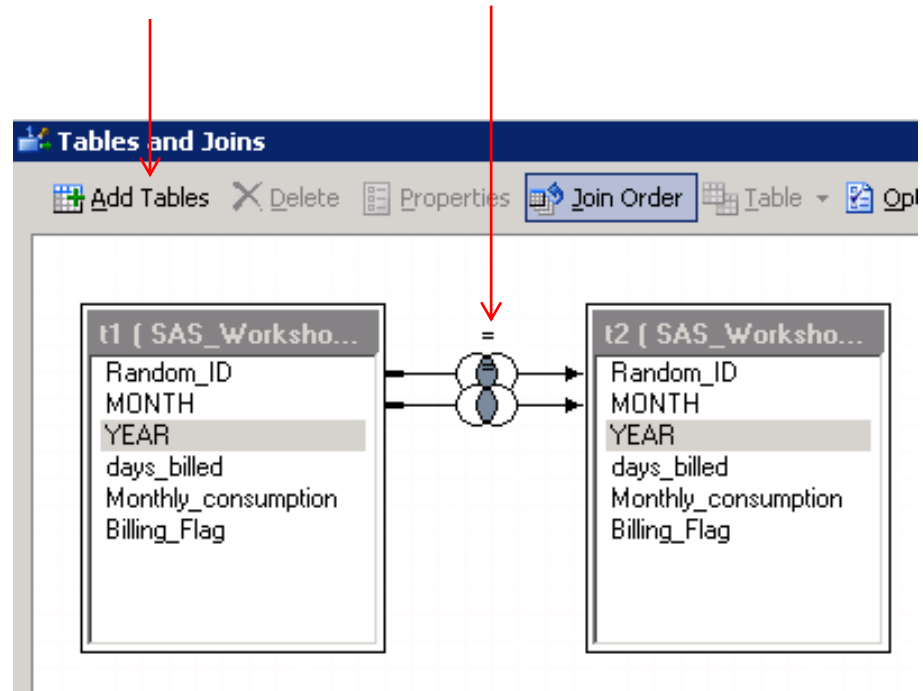
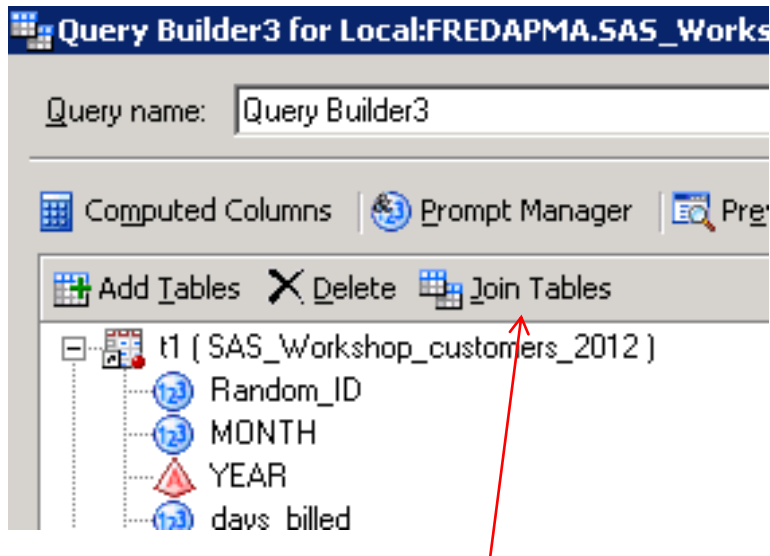
	MONTH	AVG_of_Monthly_consumption
1	1	155263.17677
2	2	145427.61133
3	3	153691.82435
4	4	143409.41963
5	5	146503.86517
6	6	141036.00453
7	7	142943.1248
8	8	146316.11801
9	9	143907.35655
10	10	150341.21567
11	11	150198.99932
12	12	156454.90423

“Please Compare. What’s the Yr/Yr trend?”  
 (Join 2 tables, one from this year and one from last year)



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit
how did the consumption change for the <u>top 10</u> accounts in 2012?	Y		Y (views and/or complex subquery required)

Query Builder: Join Tables





# “Please Compare. What’s the Yr/Yr trend?” (Join 2 tables, one from this year and one from last year)



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit
how did the consumption change for the top 10 accounts in 2012?	Y		Y (views and/or complex subquery required)

Query Builder: Drag and Drop to get the annual consumption of each account

Query name: Query Builder3      Output name: WORK.QUERY\_FOR\_SAS\_WORKSHOP\_

Computed Columns    Prompt Manager    Preview    Tools    Options

Add Tables    Delete    Join Tables

Column Name	Identifier	Summary
Random_ID (Random_ID)	t1.Random_ID	
Consumption_2012	_Calculation	SUM
Consumption_2011	_Calculation1	SUM

Summary groups  
 Automatically select groups  
 t1.Random\_ID

# “Please Compare. What’s the Yr/Yr trend?” (Join 2 tables, one from this year and one from last year)



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit
how did the consumption change for the top 10 accounts in 2012?	Y		Y (views and/or complex subquery required)

Query Builder: Sort the 2012 annual consumption, largest to smallest, Run

The screenshot shows the SAS Query Builder interface. On the left, two tables are joined: t1 (SAS\_Workshop\_customers\_2012) and t2 (SAS\_Workshop\_customers\_2011). Both tables have columns for Random\_ID, MONTH, YEAR, days\_billed, Monthly\_consumption, and Billing\_Flag. A 'Computed Columns' section at the bottom shows two calculated columns: '\_Calculation' and '\_Calculation1'. The main window displays the 'Sort Data' tab, which is highlighted with a red arrow. The table below shows the sort configuration:

Column Name	Identifier	Sort Direction
Consumption_2012	_Calculation	Descending

“Please Compare. What’s the Yr/Yr trend?”

(Join 2 tables, one from this year and one from last year)



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit
how did the consumption change for the top 10 accounts in 2012?	Y		Y (views and/or complex subquery required)

Intermediate Result > Query Builder

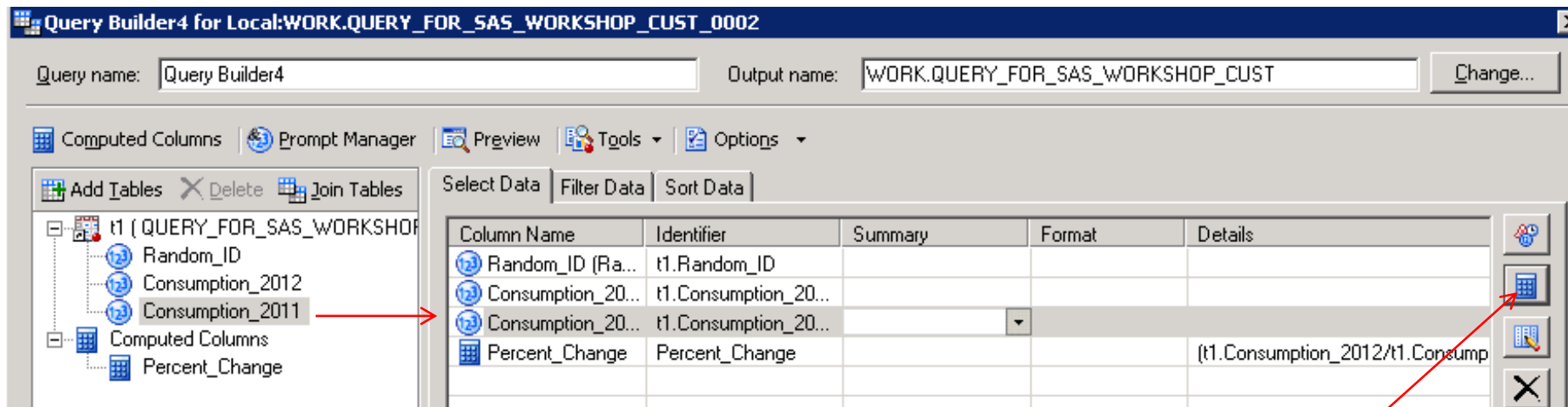
	Random_ID	Consumption_2012	Consumption_2011
1	1947	62461111.603	57612541.00000
2	1842	52059537.12	42324524.99999
3	3305	49381846.23	46051200.00000
4	5477	45842880	43828800.00000
5	426	39129513.225	34815251.00000
6	3304	38351899.18	41356800.00000
7	2268	37116454.465	34136826.00000
8	404	35671441.68	32885186.99999
9	5977	30381903.934	30654720.00000
10	6491	30155954.414	28523099.00000
11	551	27171688.837	27308945.00000

“Please Compare. What’s the Yr/Yr trend?”  
 (Join 2 tables, one from this year and one from last year)



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit
how did the consumption change for the top 10 accounts in 2012?	Y		Y (views and/or complex subquery required)

Query Builder: Drag and Drop, then Calculate new column Percent\_Change



“Please Compare. What’s the Yr/Yr trend?”  
(Join 2 tables, one from this year and one from last year)



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit
how did the consumption change for the top 10 accounts in 2012?	Y		Y (views and/or complex subquery required)

Drag and Drop, then Calculate new column Percent\_Change

Column Name:

Label:

Format:

Summary:  Length (in bytes):

Expression:  ←

“Please Compare. What’s the Yr/Yr trend?”  
 (Join 2 tables, one from this year and one from last year)



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit
how did the consumption change for the top 10 accounts in 2012?	Y		Y (views and/or complex subquery required)

Hit Run!

	Random_ID	Consumption_2012	Consumption_2011	Percent_Change
1	1947	62461111.603	57612541.00000	8.4158249559
2	1842	52059537.12	42324524.99999	23.000877434
3	3305	49381846.23	46051200.00000	7.2324852119
4	5477	45842880	43828800.00000	4.5953345745
5	426	39129513.225	34815251.00000	12.391874542
6	3304	38351899.18	41356800.00000	-7.265796242
7	2268	37116454.465	34136826.00000	8.7284871315
8	404	35671441.68	32885186.99999	8.4726739728
9	5977	30381903.934	30654720.00000	-0.889964307
10	6491	30155954.414	28523099.00000	5.7246774403
11	551	27171688.877	27308945.00000	-0.502605001

# Please Produce a pivot Table



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit	Export to Excel
Pivot Data	Y via Transpose	No	Very complex, need to have prior knowledge of columns	Small data yes. Big data, no.

Random_ID	MONTH	YEAR	days_billed	Monthly_consumpti	Billing_Flag
5483	4	2012	30	13320	N
5483	5	2012	31	101412.413795	N
5483	6	2012	30	70245.51724	N
5483	7	2012	31	109570.909092	N
5483	8	2012	31	136636.363635	N
5483	9	2012	30	113880	N

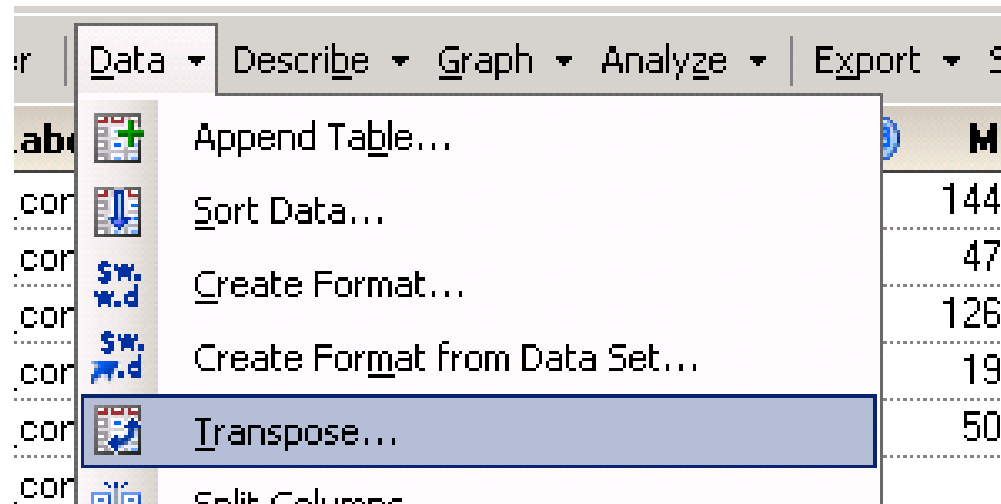
Random_ID	Month_4	Month_5	Month_6	Month_7	Month_8	Month_9	Month_10
1	124560.00000	136327.50000	138985.60345	152822.89655	155631.27273	140432.72727	133165.71429
2	64320.00000	65408.00000	58372.00000	31068.00000	34569.93103	48992.06897	59606.00000
3	38880.00000	40485.51724	41081.37931	44134.92163	44193.77504	42833.89831	43710.50847
4	141600.00000	153348.38710	144731.61290	144913.54839	145200.00000	150526.45161	154080.00000

# Please Produce a pivot Table



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit	Export to Excel
Pivot Data	Y via Transpose	No	Very complex, need to have prior knowledge of columns	Small data yes. Big data, no.

Data > Transpose





# Please Produce a pivot Table



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit	Export to Excel
Pivot Data	Y via Transpose	No	Very complex, need to have prior knowledge of columns	Small data yes. Big data, no.

Transpose > Data  
Drag and drop variables:

The screenshot shows the SAS Transpose dialog box. The 'Data' tab is selected. The 'Data source' is 'Local:FREDAPMA.SAS\_Workshop\_customers\_2011' and the 'Task filter' is 'None'. The 'Variables to assign' list contains: Random\_ID, MONTH, YEAR, days\_billed, Monthly\_consumption, and Billing\_Flag. The 'Task roles' list contains: Transpose variables, Monthly\_consumption, Copy variables, New column names (Limit: 1), MONTH, Group analysis by, Random\_ID, and New column labels (Limit: 1). Red arrows point to 'Monthly\_consumption', 'MONTH', and 'Random\_ID' in the 'Task roles' list.

# Please Produce a pivot Table



**Microsoft**

Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit	Export to Excel
Pivot Data	Y via Transpose	No	Very complex, need to have prior knowledge of columns	Small data yes. Big data, no.

Transpose > Options

Transpose\_Consumption for Local:FREDAPMA.SAS\_Workshop\_cust

Data  
Options  
Results  
Properties

**Options**

Source column

Name: Source

Label: Label

Column name prefix

Use prefix: Month\_ ←

Note: The prefix check box is checked and "New column names" tab

# Please Produce a pivot Table



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit	Export to Excel
Pivot Data	Y via Transpose	No	Very complex, need to have prior knowledge of columns	Small data yes. Big data, no.

Run!

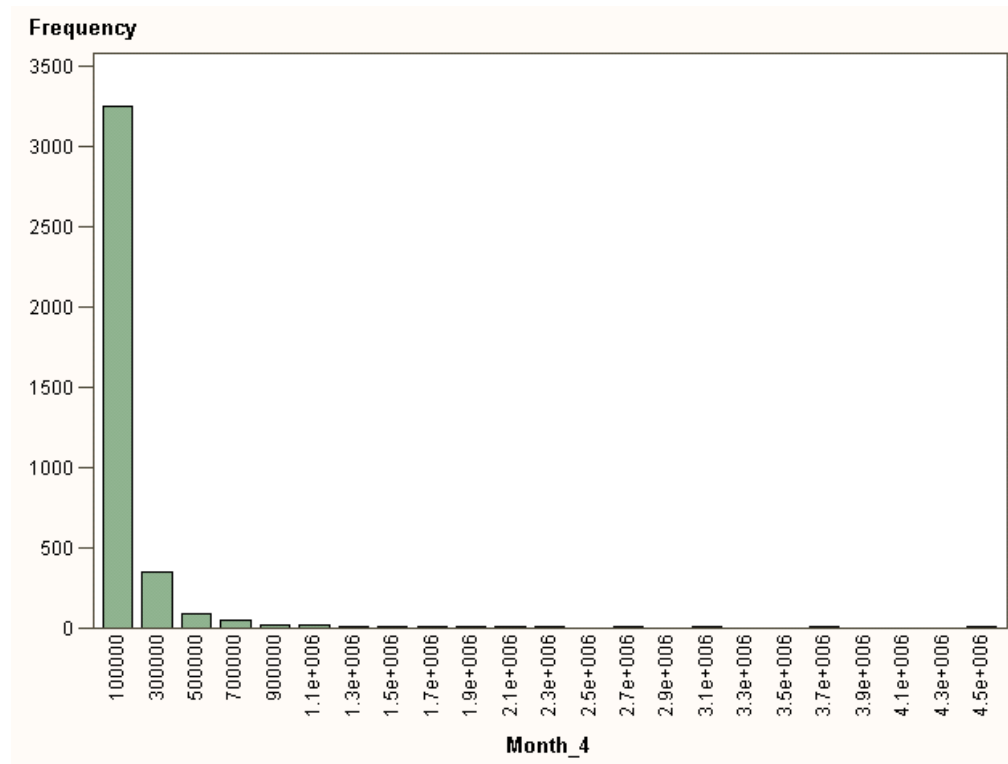
Random_ID	Month_4	Month_5	Month_6	Month_7
1	124560.00000	136327.50000	138985.60345	15
2	64320.00000	65408.00000	58372.00000	3
3	38880.00000	40485.51724	41081.37931	4
4	141600.00000	153348.38710	144731.61290	14
5	59040.00000	54856.55172	47047.44828	4
6	140160.00000	135510.00000	126378.00000	13
7	5400.00000	19042.75862	19848.49138	1
8	48480.00000	50344.00000	50795.31034	5

# “Can you do a histogram of April?”



Statistic	SAS EG GUI	SQL Server GUI	SQL Server GUI + code edit	Excel
Graph Pivot Data	Y	N	Need help	Y, after getting the data

Graph!



# Questions?

I'm still a novice – so I may not be able to answer all of your questions ;)

Paulus Mau

BC Hydro

Paulus.mau@bchydro.com