SAS® GLOBAL FORUM 2017

April 2 - 5 | Orlando, FL

#SASGF

USERS PROGRAM

Presenter

Philip Mason

Phil has been a SAS programmer for 32 years, having spent 25 years as an independent SAS Consultant. Starting in Australia he then moved on to Europe working in Belgium, Netherlands and the U.K. where he is now based. He has written 2 books on SAS Tips and is nearing completion of a book on SAS Stored Processes for SAS Publishing. He has also written several SAS courses, founded some user groups, helped run SAS Global Forum, and so on.

He still enjoys working with SAS each day and is currently helping Lloyds Banking Group.

Make Data Flow diagrams from SAS programs automatically

Phil Mason

Agenda

Overview

PROC SCAPROC

GraphViz Overview

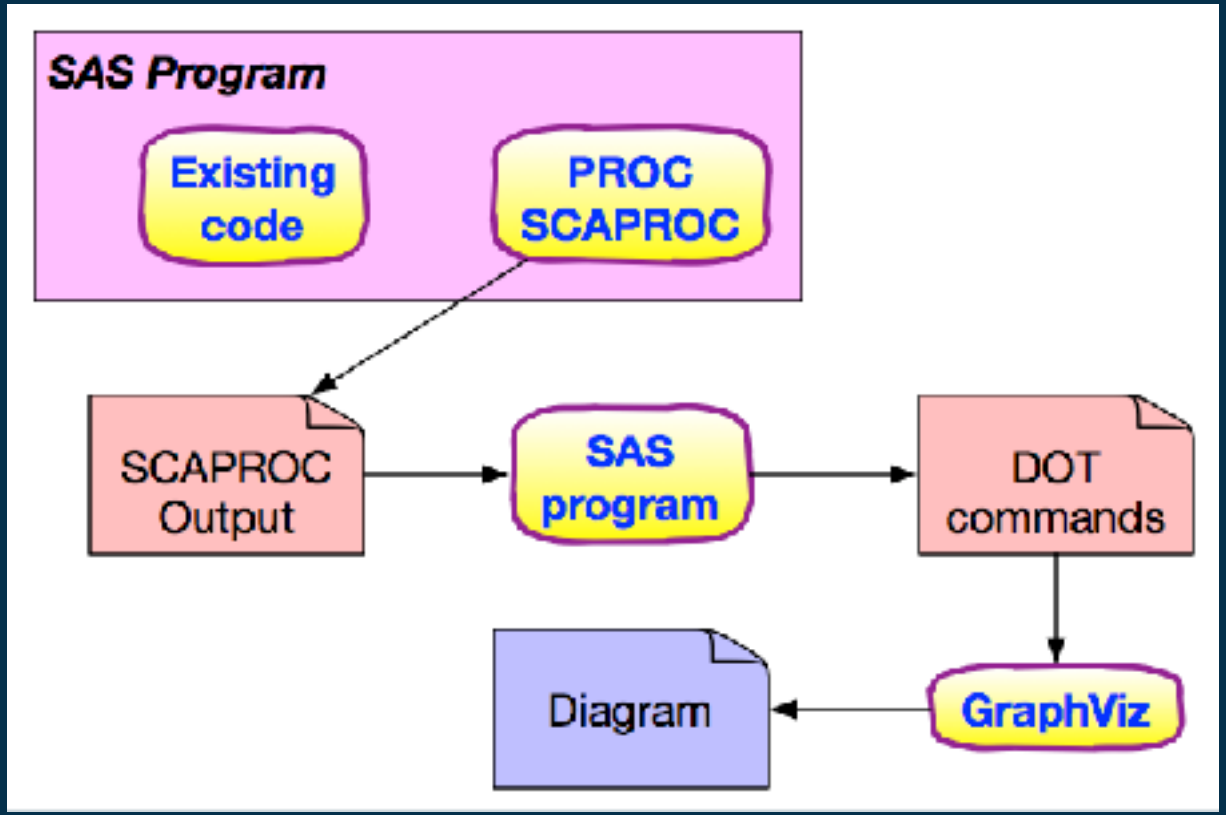
Generating GraphViz from SCAPROC

Automating with macros

Future Enhancements

Conclusions

Overview


SAS Program
Existing code
PROC SCAPROC
SCAPROC Output
SAS program
DOT commands
Diagram
GraphViz

PROC SCAPROC

Syntax

PROC SCAPROC;

RECORD *filespec* <ATTR> <OPENTIMES> <INTCON> <EXPANDMACROS>
<GRID *filespec* <RESOURCE "*resource name*"> <INHERITLIB> <NOOPTIMIZE> >;

WRITE;

```
proc scaproc ;
    record "c:\users\phil\scaproc.txt" ;
run ;

data x ;
    set sashelp.prdsale sashelp.prdsal2 ;
run ;

proc scaproc ;
    write ;
run ;
```

- SAS Code Analyzer, captures info about:
  - input, output & macro symbols
- ATTR - variable attribute info
- OPENTIMES - open time, size & physical filename
- INTCON - integrity constraint info
- EXPANDMACROS - expand macro calls into separate tasks
- GRID - help grid enable code

SCAPROC Output

```
/* JOBSPLIT: JOBSTARTTIME 27JAN2017:17:07:04.57 */
/* JOBSPLIT: TASKSTARTTIME 27JAN2017:17:07:04.57 */
/* JOBSPLIT: DATASET INPUT SEQ #C00001.PRDSALE.DATA */
/* JOBSPLIT: LIBNAME #C00001 V9 'C:\Program Files\SASHome\SASFoundation\9.4\core\sashelp' */
/* JOBSPLIT: CONCAT#LM #C00001 SASHELP */
/* JOBSPLIT: LIBNAME SASHELP V9 '( 'C:\Program Files\SASHome\SASFoundation\9.4\nls\en\SASCFG' 'C:\Program Files\SASHome\SASFoundation\9.4\core\sashelp' 'C:\Program Files
\SASHome\SASFoundation\9.4\aacomp\sashelp' 'C:\Program Files\SASHome\SASFoundation\9.4\af\sashelp' 'C:\Program Files\SASHome\SASFoundation\9.4\cmp\sashelp' 'C:\Program Files
\SASHome\SASFoundation\9.4\connect\sashelp' 'C:\Program Files\SASHome\SASFoundation\9.4\dmine\sashelp' 'C:\Program Files\SASHome\SASFoundation\9.4\ets\sashelp' 'C:\Program
Files\SASHome\SASFoundation\9.4\graph\sashelp' 'C:\Program Files\SASHome\SASFoundation\9.4\iml\sashelp' 'C:\Program Files\SASHome\SASFoundation\9.4\inttech\sashelp' 'C:\Program
Files\SASHome\SASFoundation\9.4\mddbserv\sashelp' 'C:\Program Files\SASHome\SASFoundation\9.4\or\sashelp' 'C:\Program Files\SASHome\SASFoundation\9.4\spdsclient\sashelp' 'C:
\Program Files\SASHome\SASFoundation\9.4\stat\sashelp' )' */
/* JOBSPLIT: DATASET INPUT SEQ #C00001.PRDSAL2.DATA */
/* JOBSPLIT: LIBNAME #C00001 V9 'C:\Program Files\SASHome\SASFoundation\9.4\core\sashelp' */
/* JOBSPLIT: CONCAT#LM #C00001 SASHELP */
/* JOBSPLIT: LIBNAME SASHELP V9 '( 'C:\Program Files\SASHome\SASFoundation\9.4\nls\en\SASCFG' 'C:\Program Files\SASHome\SASFoundation\9.4\core\sashelp' 'C:\Program Files
\SASHome\SASFoundation\9.4\aacomp\sashelp' 'C:\Program Files\SASHome\SASFoundation\9.4\af\sashelp' 'C:\Program Files\SASHome\SASFoundation\9.4\cmp\sashelp' 'C:\Program Files
\SASHome\SASFoundation\9.4\connect\sashelp' 'C:\Program Files\SASHome\SASFoundation\9.4\dmine\sashelp' 'C:\Program Files\SASHome\SASFoundation\9.4\ets\sashelp' 'C:\Program
Files\SASHome\SASFoundation\9.4\graph\sashelp' 'C:\Program Files\SASHome\SASFoundation\9.4\iml\sashelp' 'C:\Program Files\SASHome\SASFoundation\9.4\inttech\sashelp' 'C:\Program
Files\SASHome\SASFoundation\9.4\mddbserv\sashelp' 'C:\Program Files\SASHome\SASFoundation\9.4\or\sashelp' 'C:\Program Files\SASHome\SASFoundation\9.4\spdsclient\sashelp' 'C:
\Program Files\SASHome\SASFoundation\9.4\stat\sashelp' )' */
/* JOBSPLIT: DATASET OUTPUT SEQ WORK.X.DATA */
/* JOBSPLIT: LIBNAME WORK V9 'C:\Users\phil\AppData\Local\Temp\2\SAS Temporary Files\_TD21120_D351TQ92_' */
/* JOBSPLIT: FILE OUTPUT c:\users\phil\scaproc.txt */
/* JOBSPLIT: ELAPSED 28  */
/* JOBSPLIT: SYSSCP WIN */
/* JOBSPLIT: PROCNAME DATASTEP */
/* JOBSPLIT: STEP SOURCE FOLLOWS */

data x ;
    set sashelp.prdsale sashelp.prdsal2 ;
run ;


/* JOBSPLIT: JOBENDTIME 27JAN2017:17:07:04.68 */
/* JOBSPLIT: END */
```

Reading output from PROC SCAPROC

```
data scaproc_parsed(keep=in out step) ;
 retain step 1 ;
 infile "c:\users\phil\scaproc.txt" ;
 input ;
 word1=scan(_infile_,1,' ') ;
 word2=scan(_infile_,2,' ') ;
 word3=scan(_infile_,3,' ') ;
 word4=scan(_infile_,4,' ') ;
 word5=scan(_infile_,5,' ') ;
 word6=scan(_infile_,6,' ') ;
 word7=scan(_infile_,7,' ') ;
 if word2='JOBSPLIT:' & word3='DATASET' & word4='INPUT' then in=word6 ;
 if word2='JOBSPLIT:' & word3='DATASET' & word4='OUTPUT' then out=word6 ;
 if word2='JOBSPLIT:' & word3='STEP' then step+1 ;
 if in>'' or out>'' then output ;
run ;
```

	step	in	out
1	1	#C00001.PRDSALE.DATA	
2	1	#C00001.PRDSAL2.DATA	
3	1		WORK.X.DATA

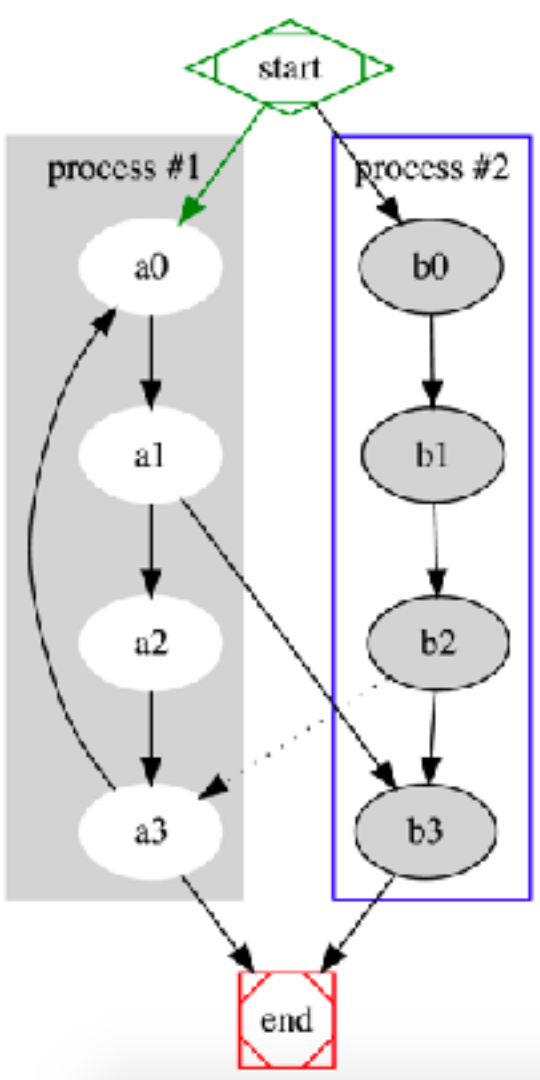
- #C00001 should be SASHELP, happens in EG

Graphviz Overview

- What is Graphviz?
  - open source graph visualisation software
  - represents structural info
  - can handle very large/complex graphs
  - graph is specified in simple text (DOT language)
  - run via Web (WebGraphViz), HTML 5 canvas JavaScript (CanViz), binaries on your platform

```
digraph G {
  "Welcome" -> "To"
  "To" -> "Web"
  "To" -> "GraphViz!"
}
```

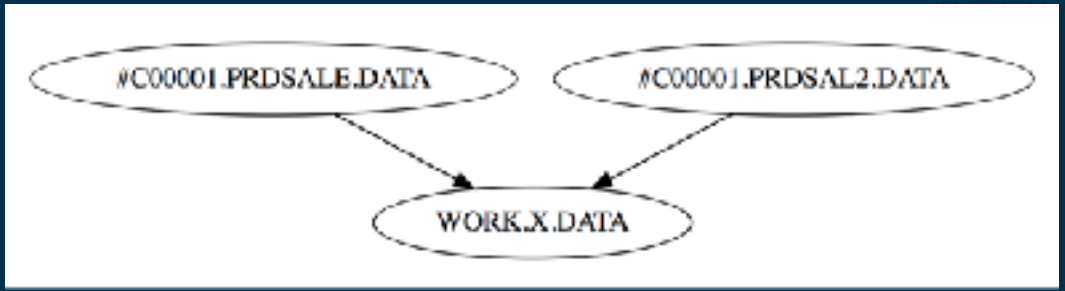
```
digraph G {subgraph cluster_0 {
                    style=filled;
                    color=lightgrey;
                    node [style=filled, color=white];
                    a0 -> a1 -> a2 -> a3;
                    label = "process #1";}
          subgraph cluster_1 {
                    node [style=filled];
                    b0 -> b1 -> b2 -> b3;
                    label = "process #2";
                    color=blue}
          start -> a0 [color=green];
          start -> b0;
          a1 -> b3;
          b2 -> a3 [style=dotted];
          a3 -> a0;
          a3 -> end;
          b3 -> end;
          start [shape=Mdiamond color=green];
          end [shape=Msquare color=red];   }
```



Making DOT commands from SCAPROC

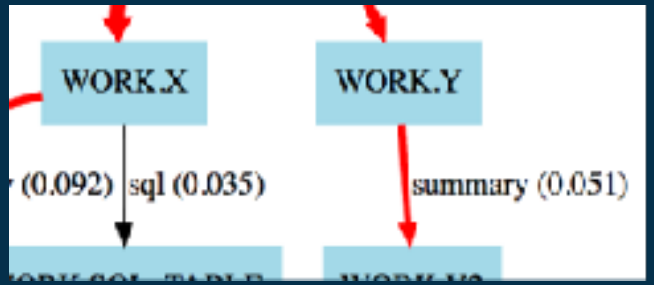
	step	in	out
1	1	#C00001.PRDSALE.DATA	
2	1	#C00001.PRDSAL2.DATA	
3	1		WORK.X.DATA

```
digraph {
"#C00001.PRDSALE.DATA" -> "WORK.X.DATA"
"#C00001.PRDSAL2.DATA" -> "WORK.X.DATA"
}
```



Adding performance information

- We can enhance diagram with extra information
  - Black arrow indicates step took less than median step run time
  - Red arrow indicates step took longer than median step run time
  - Thicker arrows indicate longer was taken (e.g. 75th, 90th percentile)
  - Name of step next to arrow
  - Actual real time next to arrow



Automating with macros

- PROC SCAPROC starts recording → %eanbegin
- Code runs → your SAS code
- PROC SCAPROC stops and writes file → %eanend
- SAS program processes SCAPROC file to produce DOT commands for Graphviz → scaproc_analyse
  - Percentiles calculated for steps and used to show performance info → %percentiles

Using the macros

- %eanbegin(*text to use as title for diagram*)
  - e.g. %eanbegin(Monthly report flow)
- %eanend
- %scaproc_analyse
  - %percentiles(*table*, *variables*, pctlpts=*percentiles you want*)
    - e.g. %percentiles(sashelp.air, air, pctlpts=25 50 75)

Example using macros

```
%* turn on SCAPROC and
   set verbose logging ;
%let _eandebug=scaproc,verbose;

%* Start recording SCAPROC data ;
%eanbegin(Sample 1)

**********************;
*** Sample program ***;
**********************;
data x ;
 set sashelp.class ;
run ;

data y ;
 set sashelp.class ;
run ;

proc summary data=x ;
 class sex ;
 var height ;
 output out=x2 mean= ;
run ;
```

```
proc summary data=y ;
 class sex ;
 var height ;
 output out=y2 mean= ;
run ;

proc sort data=x2 out=x3 ;
 by sex ;
run ;

proc sort data=y2 out=y3 ;
 by sex ;
run ;

data z ;
 merge x3 y3 ;
  by sex ;
run ;

proc print ;
run ;
```

```
proc sql ;
 create table sql_table as
 select *
 from x
 left join
 y
 on x.sex=y.sex ;
quit ;


****************************;
*** finish of sample program ;
****************************;

%* Finish recording SCAPROC data
    and write it out ;
%eanend


%* Generate the graphViz dot language
    to be used to make diagram ;
%scaproc_analyse
```

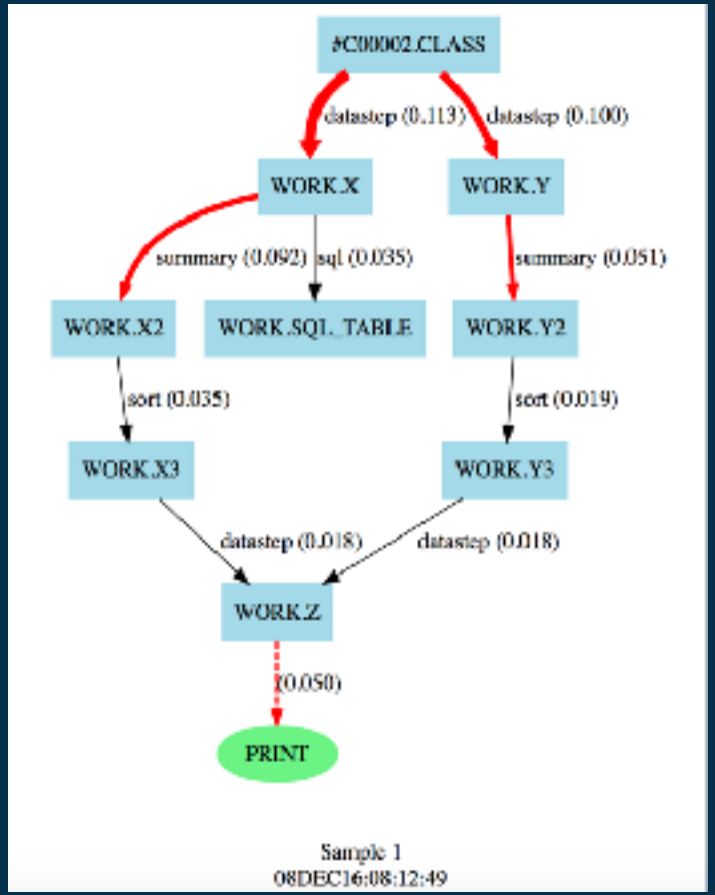
```
    /* JOBSPLIT: JOBSTARTTIME 18JAN2017:08:17:56.55 */
    /* JOBSPLIT: TASKSTARTTIME 18JAN2017:08:17:56.59 */
    /* JOBSPLIT: DATASET INPUT SEQ #C00002.CLASS.DATA */
    /* JOBSPLIT: LIBNAME #C00002 V9 '/config/sas_9.4/sas_9.4/RTL/SASFoundation/9.4/sashelp' */
    /* JOBSPLIT: CONCATMEM #C00002 SASHELP */
    /* JOBSPLIT: LIBNAME SASHELP V9 '( '/config/sas_9.4/sas_9.4/RTL/SASFoundation/9.4/nls/en/sascfg' '/config/sas_9.4/sas_9.4/RTL/SASFoundation/9.4/
nls/en/sascfg' '/config/sas_9.4/sas_9.4/RTL/SASFoundation/9.4/sashelp' )' */
    /* JOBSPLIT: OPENTIME #C00002.CLASS.DATA DATE:18JAN2017:08:17:56.59 PHYS:/config/sas_9.4/sas_9.4/RTL/SASFoundation/9.4/sashelp/class
    .sas7bdat SIZE:131072 */
    /* JOBSPLIT: DATASET OUTPUT SEQ WORK.X.DATA */
    /* JOBSPLIT: LIBNAME WORK V9 '/temp/work1/sas_9.4/SAS_work873E02930064_p24107prw310/SAS_work870802930064_p24107prw310' */
    /* JOBSPLIT: ATTR #C00002.CLASS.DATA INPUT VARIABLE:Name TYPE:CHARACTER LENGTH:8 LABEL: FORMAT: INFORMAT: */
    /* JOBSPLIT: ATTR #C00002.CLASS.DATA INPUT VARIABLE:Sex TYPE:CHARACTER LENGTH:1 LABEL: FORMAT: INFORMAT: */
    /* JOBSPLIT: ATTR #C00002.CLASS.DATA INPUT VARIABLE:Age TYPE:NUMERIC LENGTH:8 LABEL: FORMAT: INFORMAT: */
    /* JOBSPLIT: ATTR #C00002.CLASS.DATA INPUT VARIABLE:Height TYPE:NUMERIC LENGTH:8 LABEL: FORMAT: INFORMAT: */
    /* JOBSPLIT: ATTR #C00002.CLASS.DATA INPUT VARIABLE:Weight TYPE:NUMERIC LENGTH:8 LABEL: FORMAT: INFORMAT: */
    /* JOBSPLIT: ATTR WORK.X.DATA OUTPUT VARIABLE:Name TYPE:CHARACTER LENGTH:8 LABEL: FORMAT: INFORMAT: */
    /* JOBSPLIT: ATTR WORK.X.DATA OUTPUT VARIABLE:Sex TYPE:CHARACTER LENGTH:1 LABEL: FORMAT: INFORMAT: */
    /* JOBSPLIT: ATTR WORK.X.DATA OUTPUT VARIABLE:Age TYPE:NUMERIC LENGTH:8 LABEL: FORMAT: INFORMAT: */
    /* JOBSPLIT: ATTR WORK.X.DATA OUTPUT VARIABLE:Height TYPE:NUMERIC LENGTH:8 LABEL: FORMAT: INFORMAT: */
    /* JOBSPLIT: ATTR WORK.X.DATA OUTPUT VARIABLE:Weight TYPE:NUMERIC LENGTH:8 LABEL: FORMAT: INFORMAT: */
    /* JOBSPLIT: SYMBOL GET SCAPROC_LABEL */
    /* JOBSPLIT: SYMBOL GET USE_LABEL */
    /* JOBSPLIT: SYMBOL SET SCAPROC_LABEL */
    /* JOBSPLIT: SYMBOL GET _EANDEBUG */
    /* JOBSPLIT: SYMBOL GET I */
    /* JOBSPLIT: SYMBOL SET PARM */
    /* JOBSPLIT: SYMBOL GET PARM */
    /* JOBSPLIT: SYMBOL SET I */
    /* JOBSPLIT: ELAPSED 117  */
    /* JOBSPLIT: SYSSCP AIX 64 */
    /* JOBSPLIT: PROCNAME DATASTEP */
    /* JOBSPLIT: STEP SOURCE FOLLOWS */


    ************************************************************************;
    *** This is the sample program we will measure and then make a flow chart ;
    ************************************************************************;
    data x ;
       set sashelp.class ;
    run ;


    /* JOBSPLIT: TASKSTARTTIME 18JAN2017:08:17:56.70 */
    /* JOBSPLIT: DATASET INPUT SEQ #C00002.CLASS.DATA */
```

SCAPROC output produced

DOT commands produced by SCAPROC_ANALYSE

```
// Generated by SAS for Sample 1
// Percentiles: 50:0.035 60:0.05 70:0.051 80:0.092 90:0.1 95:0.113 99:0.113
digraph test {
graph [label="\n\nSample 1\n08DEC16:08:12:49"]
node [shape=box color=lightblue style=filled]
"PRINT"[shape=ellipse color=lightgreen]
"#C00002.CLASS"->"WORK.X" [label=" datastep (0.113)" color=red penwidth=7 style=solid];
"#C00002.CLASS"->"WORK.Y" [label=" datastep (0.100)" color=red penwidth=5 style=solid];
"WORK.X"->"WORK.X2" [label=" summary (0.092)" color=red penwidth=4 style=solid];
"WORK.Y"->"WORK.Y2" [label=" summary (0.051)" color=red penwidth=3 style=solid];
"WORK.X2"->"WORK.X3" [label=" sort (0.035)" color=red  style=solid];
"WORK.Y2"->"WORK.Y3" [label=" sort (0.019)"   style=solid];
"WORK.X3"->"WORK.Z" [label=" datastep (0.018)"   style=solid];
"WORK.Y3"->"WORK.Z" [label=" datastep (0.018)"   style=solid];
"WORK.Z"->"PRINT" [label="(0.050)" color=red penwidth=2 style=dashed];
"WORK.X"->"WORK.SQL_TABLE" [label=" sql (0.035)" color=red  style=solid];
"WORK.Y"->"WORK.SQL_TABLE" [label=" sql (0.035)" color=red  style=solid];
}
```

Diagram produced



Future Enhancements

- Install Graphviz on local to SAS to automate graph generation
- Integrate info from SAS log such as number of rows, CPU time, recs in/out
- Fix # librefs in EG
- Make a version based on log analysis rather than Proc SCAPROC
  - attempt a version just based on SAS program code
- Integrate into stored process for running via web browser
- Store DOT code generated so it can be compared to future versions
  - allows showing program structure changes, comparing performance
- Generate web based diagrams and make us of tooltips and hyperlinks

Conclusions

- Start making flow diagrams from your code automatically (almost)
  - There's nothing to buy or license, you have it all right there
  - Make diagrams for easy documentation
  - Use diagrams to quickly understand complex program flows
  - Quickly spot performance issues in large programs

Questions?

- I'll be outside the room after this talk
- Give me your email address if you want updated macros sent to you.

Don't forget to provide feedback!

1. Go to the Agenda icon in the conference app.
2. Find this session title and select it.
3. On the sessions page, scroll down to Surveys and select the name of the survey.
4. Complete the survey and click Finish.