Ask the Experts
Leveraging SAS® Viya to improve Base SAS processes

Steven Sober - Advisory Solutions Architect, Global Technology Practice, DataOps

§sas
THE POWER TO KNOW®

Leveraging SAS® Viya to improve Runtimes of SAS processes
Agenda

1. **Terminology**
2. Findings from SAS Viya Readiness Utility
3. Customer User Cases
4. Reading Material

SAS

Leveraging SAS® Viya to improve Runtimes of SAS processes
Terminology

- SAS Viya
  - Two execution engines
    - SAS Workspace Server
      - Data (Disk) and Code can be processed here
        - LIBNAME
    - SAS Cloud Analytic Services (CAS = In-Memory)
      - Data (In-Memory) and Code can be processed here
        - CASLIB

- Deployment options
  - On-premises
  - Cloud
    - Public
    - Private
    - Hosted by SAS
  - SMP
    - Single CAS Worker
    - Linux
    - Windows on roadmap
  - Distributed
    - Multiple CAS Workers
    - Resiliency
    - **Linux**
    - Co-Located on Hadoop

Leveraging SAS® Viya to improve Runtimes of SAS processes
Terminology

- ## Findings from SAS Viya Readiness Utility
  - 309 SAS Jobs with 2,629 Steps
    - **Code from the Base SAS Core samples that are delivered with SAS 9**

- ## CAS Enabled (fast)
  - In-Memory processing of data (fast)
  - Process is ran on multiple threads on multiple CAS worker nodes (distributed)

- ## Candidate for CAS
  - PROC SQL (Workspace Server)
    - Convert to PROC FEDSQL (CAS Enabled)
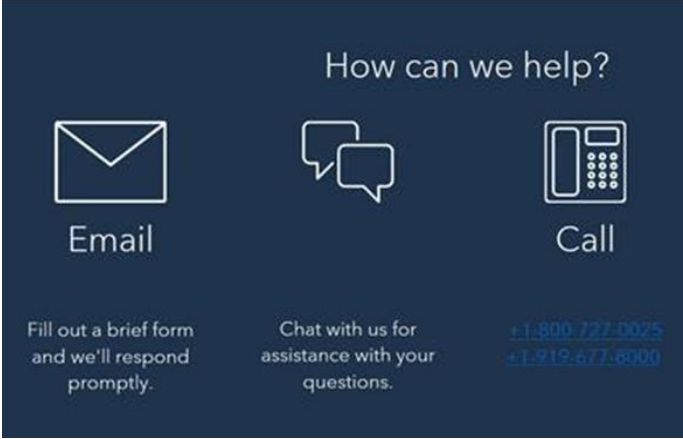
- ## Workspace Server
  - Process is ran on SAS Viya Workspace Server
    - Stand Alone SAS Viya
  - Process is ran on SAS 9.4 M5 or higher Workspace Server
    - SAS Viya and SAS 9.4M5

§.sas

Leveraging SAS® Viya to improve Base SAS processes
Terminology

1. **SAS Viya Readiness Utility**
   a) A code review service from SAS
      i. Contact your SAS Account Executive
      ii. **Contact the SAS Customer Success Organization**
         - www.sas.com/contact
            - I suggest scrolling down to the "Chat with …."
            - Ask for your your Customer Success Manager

How can we help?

Email
Fill out a brief form and we'll respond promptly.

Chat with us for assistance with your questions.

Call
+1-800-727-0025
+1-919-677-8000

Leveraging SAS® Viya to improve Runtimes of SAS processes
Agenda

1. Terminology

2. **Findings from SAS Viya Readiness Utility**

3. Customer User Cases

4. Reading Material

Leveraging SAS® Viya to improve Base SAS processes
DATA Step is CAS Enabled

Procedures That Use CAS Actions

The following Base SAS procedures can run CAS actions:

Procedure	Description
PROC APPEND	Adds rows from a CAS table to the end of a SAS data set, and adds rows from a SAS data set to the end of a CAS table.
PROC CONTENTS	Shows the contents of a CAS table and prints the directory of the caslib.
PROC COPY	Copies entire SAS libraries or specific members of the library.
PROC DATASETS	Manages CAS tables.
PROC DELETE	Deletes SAS data sets and CAS tables.
PROC DS2 1	Manipulates data with DS2 language statements.
PROC FCMP	Enables you to create, test, and store SAS functions, CALL routines, and subroutines before you use them in other SAS procedures or in DATA steps.
PROC FEDSQL 2	Manipulates data and performs reporting with FedSQL language statements.
PROC FORMAT	Creates user-defined informats to read data and user-defined formats to display data.
PROC LUA	Enables you to run statements from the Lua programming language within SAS code.
PROC MEANS	Computes descriptive statistics; can produce printed output and output data sets. By default, PROC MEANS produces printed output.
PROC REPORT	Combines features of the PRINT, MEANS, and TABULATE procedures with features of the DATA step in a single report-writing tool that can produce a variety of reports.
PROC SCOREACCEL	Provides an interface to the CAS server for DATA step and DS2 model publishing and scoring.
PROC SUMMARY	Computes descriptive statistics; can produce a printed report and create an output data set. By default, PROC SUMMARY creates an output data set.
PROC TABULATE	Displays descriptive statistics in tabular format, using some or all of the variables in a data set.
PROC TRANSPOSE	Transforms SAS data sets so that observations become variables and variables become observations.

1   The DS2 procedure does not use the CAS LIBNAME engine to access in-memory tables. Instead, the procedure accesses tables by caslib and name. For information and limitations, see *DS2 in CAS: Concepts* in *SAS DS2 Programmer's Guide*.

2   The FEDSQL procedure does not use the CAS LIBNAME engine to access in-memory tables. Instead, the procedure accesses tables by caslib and name. For information and limitations, see *SAS Viya: FedSQL Programming for SAS Cloud Analytic Services*

Leveraging SAS® Viya to improve Base SAS processes
CAS Enabled

SAS Viya Readiness Utility
309 SAS Jobs with 2,629 Steps

CAS Enabled - PROC FORMAT	68	2.59	68	2.59
CAS Enabled - PROC MEANS	38	1.45	106	4.03
CAS Enabled - PROC REPORT	14	0.53	120	4.56
CAS Enabled - PROC SUMMARY	21	0.80	141	5.36
CAS Enabled - PROC TABULATE	50	1.90	191	7.27
CAS Enabled - PROC TRANSPOSE	20	0.76	211	8.03

1. 8.03% of the code is CAS enabled provided
   a) Source and target tables are CAS tables
   b) Functionality of MEANS, REPORT, SUMMARY, TABULATE is the same as the In-Database capabilities of these procedures

Leveraging SAS® Viya to improve Base SAS processes
Candidate for CAS

SAS Viya Readiness Utility
309 SAS Jobs with 2,629 Steps

Candidate for CAS - DATA STEP		667	25.37	878	33.40

1. 25.37% of the code is DATA Step which is CAS enabled provided
   a) Source and Target tables are in CAS
   b) Review category_summary.pdf to gain an understanding of which key words are not CAS enabled and which SAS job they appear in
      a) WHERE clauses can push down to CAS provided the WHERE is on the Source(s) table(s) to the DATA Step
      b) CARDS, EXECUTE, FILE, INFILE, INPUT, MODIFY, REPLACE, SYMGET, SYMPUT, SYMPUTX….
         a) Leave these DATA Steps "as is" because these statements do not execute in CAS
      c) %include statements
         a) Can be in the body of a DATA Step and still run in CAS provided all the statements in the include file are CAS enabled

Leveraging SAS® Viya to improve Base SAS processes

Candidate for CAS

A Service SAS Offers
SAS Viya Code Readiness
309 SAS Programs

biorythm.sas

Category	Keyword	Viya	CAS
statement	TITLE	0	2
statementDataStep	CARDS	0	1
	INPUT	0	1

Leveraging SAS® Viya to improve Base SAS processes
Candidate for CAS

SAS Viya Readiness Utility
309 SAS Jobs with 2,629 Steps

Candidate for CAS - PROC ANOVA	2	0.08	880	33.47
Candidate for CAS - PROC CORR	21	0.80	901	34.27
Candidate for CAS - PROC GAM	1	0.04	902	34.31
Candidate for CAS - PROC GLM	22	0.84	924	35.15
Candidate for CAS - PROC PRINCOMP	1	0.04	925	35.18
Candidate for CAS - PROC REG	7	0.27	932	35.45
Candidate for CAS - PROC STANDARD	5	0.19	937	35.64
Candidate for CAS - PROC UNIVARIATE	67	2.55	1004	38.19
Workspace Server - PROC IML	1	0.04	2070	78.74

1. 4.85% of the code are analytical procedures
   a) Have a SAS analytical coders review these and determine if they can simulate these procedures using CASL or other VDMML procedures.

§sas

Leveraging SAS® Viya to improve Base SAS processes
Candidate for CAS

SAS Viya Readiness Utility
309 SAS Jobs with 2,629 Steps

odsexcl3.sas

Line Number / Step	Viya	CAS	Macro	Include
1 OPEN CODE	0	2	0	0
6 PROC SGPLOT	0	0	0	0
14 OPEN CODE	0	2	0	0
17 PROC PRINT	0	0	0	0
19 OPEN CODE	0	1	0	0
20 PROC GPLOT	0	0	0	0
23 OPEN CODE	0	1	0	0
27 PROC GLM	1	1	0	0
30 OPEN CODE	0	0	0	0

1. 4.85% of the code are analytical procedures
   a) Search the step_summary.pdf to figure out which SAS jobs contain these procedures

§sas

§.sas

Leveraging SAS® Viya to improve Base SAS processes
Candidate for CAS

SAS Viya Readiness Utility
309 SAS Jobs with 2,629 Steps

Candidate for CAS by Converting to PROC FEDSQL - PROC SQL	96	3.65	1100	41.84

1. 3.65% of the code is PROC SQL which is not CAS enabled
   a) Convert these into PROC FEDSQL which is CAS enabled provided
      i. Source and Target tables are in CAS
   b) For PROC SQL that is not converted into FEDSQL
      I. SQL will run on SAS workspace server
      II. You may have to specify the SAS system option in order to transfer large CAS table to the SAS Workspace Server
         a. OPTIONS CASDATALIMIT=ALL;
   c) In code reviews I have learned PROC SQL is typically 15% to 25% of the code

§sas

Leveraging SAS® Viya to improve Base SAS processes
Candidate for CAS

SAS Viya Readiness Utility
309 SAS Jobs with 2,629 Steps

Candidate for Commenting Out - PROC SORT	109	4.15	1209	45.99

1. 4.15% of the code is PROC SORT which is not CAS enabled
   a) Majority of PROC SORT code can be commented out because the BY statement of DATA Step or CAS enabled procedures will happen on the fly
   b) Consider simulating PROC SORT code that contains the NODUPKEY statement with a DATA Step
   c) In code reviews I have learned PROC SORT is typically 15% to 25% of the code

```
/*PROC SORT DATA=&SOURCE OUT=&TARGET NODUPKEY;*/
/*  BY VAR1 VAR2 VAR3;*/
/*RUN;*/
DATA &TARGET;
    SET &SOURCE;
    BY VAR1 VAR2 VAR3;
    IF FIRST.VAR3 THEN OUTPUT;
RUN;
```

§sas

Leveraging SAS® Viya to improve Base SAS processes
Workspace Server

SAS Viya Readiness Utility
309 SAS Jobs with 2,629 Steps

Workspace Server - OPEN CODE		646	24.57	1907	72.54

1. 24.57% of the code is OPEN CODE(code between DATA or PROC boundaries)
   a) LIBNAME, FILENAME
      a) Run on SAS workspace server
   b) Macros statements
      a) Statements that start with %
      b) %DO …., %IF …., %MACRO MYMAC(VAR1,VAR2…)….
         i. Majority of these % statements can be used with code that runs in CAS

§sas

Leveraging SAS® Viya to improve Base SAS processes
Workspace Server

SAS Viya Readiness Utility
309 SAS Jobs with 2,629 Steps

Workspace Server - PROC APPEND	8	0.30	1915	72.84

1. 0.30% of the code is PROC APPEND which is not CAS enabled
   a) Consider simulating with DATA Step that has multiple tables on the SET statement
      a) Allows you to add transformations to the data
   b) Consider PROC CASUTIL
   c) Consider addTable Action
   d) **In code reviews I have learned PROC APPEND is typically 1% to 5% of the code**

```
/*PROC APPEND BASE=&SOURCE OUT=&TARGET;*//*RUN;*/
DATA &TARGET;
    SET &TARGET(IN=S1)  &SOURCE2(IN=S2)  &SOURCE3(IN=S3);
    IF S1 THEN…;
    IF S2 THEN…;
    IF S3 THEN…;
RUN;
```

§sas

Leveraging SAS® Viya to improve Base SAS processes
Workspace Server

SAS Viya Readiness Utility

309 SAS Jobs with 2,629 Steps

In-Database / MapReduce - PROC FREQ	24	0.91	1233	46.90
In-Database / MapReduce - PROC RANK	24	0.91	1257	47.81
Workspace Server - LIBNAME	4	0.15	1261	47.97
Workspace Server - PROC CALENDAR	24	0.91	1939	73.75
Workspace Server - PROC CHART	52	1.98	1991	75.73
Workspace Server - PROC COMPARE	8	0.30	1999	76.04
Workspace Server - PROC CONTENTS	8	0.30	2007	76.34
Workspace Server - PROC DATASETS	17	0.65	2024	76.99
Workspace Server - PROC DISPLAY	2	0.08	2026	77.06
Workspace Server - PROC DOCUMENT	3	0.11	2029	77.18
Workspace Server - PROC FORMS	11	0.42	2040	77.60

Note: PROC FREQ is not CAS enabled. Consider PROC FREQTAB which is CAS enable

§sas

Leveraging SAS® Viya to improve Base SAS processes
Workspace Server

SAS Viya Readiness Utility
309 SAS Jobs with 2,629 Steps

Workspace Server - PROC GCHART	5	0.19	2045	77.79
Workspace Server - PROC GEOCODE	8	0.30	2053	78.09
Workspace Server - PROC GMAP	6	0.23	2059	78.32
Workspace Server - PROC GPLOT	2	0.08	2061	78.39
Workspace Server - PROC GPROJECT	5	0.19	2066	78.59
Workspace Server - PROC GSLIDE	3	0.11	2069	78.70
Workspace Server - PROC NICKNAME	1	0.04	2071	78.78
Workspace Server - PROC ODSLIST	2	0.08	2073	78.85
Workspace Server - PROC ODSTEXT	2	0.08	2075	78.93
Workspace Server - PROC PLOT	60	2.28	2135	81.21
Workspace Server - PROC PRINT	433	16.47	2568	97.68
Workspace Server - PROC PRINTTO	2	0.08	2570	97.76
Workspace Server - PROC SGMAP	5	0.19	2575	97.95
Workspace Server - PROC SGPANEL	4	0.15	2579	98.10
Workspace Server - PROC SGPLOT	12	0.46	2591	98.55
Workspace Server - PROC SGRENDER	3	0.11	2594	98.67
Workspace Server - PROC SGSCATTER	4	0.15	2598	98.82
Workspace Server - PROC SHEWHART	2	0.08	2600	98.90
Workspace Server - PROC TEMPLATE	20	0.76	2620	99.66
Workspace Server - PROC TIMEPLOT	9	0.34	2629	100.00

1. **41.50%** of the code is SAS Viya Friendly
   a) In code **reviews I have observed 50% to 70%** of the code is SAS Viya friendly
2. 27.91% of the code must run on a SAS workspace server

§sas

Leveraging SAS® Viya to improve Runtimes of SAS processes
Agenda

1. Terminology
2. Findings from SAS Viya Readiness Utility
3. **Customer User Cases**
4. Reading Material

Leveraging SAS® Viya to improve Base SAS processes
Conclusion - CAS is Fast

- **Financial User Case**
  - **Monte Carlo Simulations for Delinquency**
    - DATA Step
      - ~20,000 line
      - CPU intensive code
      - Source table ~1TB
      - Target table ~50GB
    - Multiple Iteration of DATA Step
      - **SAS Workspace Server - 1 hour per iteration**
  - **Business Benefit**
    - Quicker time to offer
    - No other processes are allowed to run while Monte Carlo simulation runs
      - **26 Hours (Workspace Server)**
      - **2 ½ Hours (CAS)**

- **To CAS Enable DATA Step**
  1. Changed multiple RANUNI functions (Workspace Server) to RAND function (CAS Enabled)
  2. Changed source and target tables in DATA Step to use CASLIB instead of 9.4 LIBREF
  3. Commented out PROC APPEND
  4. Created CAS enabled DATA Step to emulate append process
  - Multiple iteration of CAS enabled DATA Step
    - **CAS - 6 minutes per iteration**

§.sas

Leveraging SAS® Viya to improve Base SAS processes

Conclusion - CAS is Fast

- **US Government User Case**
  - <u>51 Iterations</u> of program for Analytical Base Table creation for Modeling and Scoring
    - **DATA Step and Base PROCS**
  - Source tables: ~220GB – 230GB
  - Target tables: ~44GB – 46GB

- **Business Benefit**
  - Faster time to results which allows for increased frequency of process execution.
  - Maintaining accuracy of results between 9.4 and Viya
  - Workspace server (SAS 9)
    - 56 hours
  - CAS
    - 9 hours

- **To CAS Enable the Process - Little to no changes required!**

  1. Changed source and target tables in DATA Step to use CASLIB instead of 9.4 LIBREF
  2. Ran BY statements with high cardinality variables *(~50 million unique ids)* in <u>*Workspace Server*</u>
     a) PROC SORT (workspace server)
     b) DATA Step (workspace server)
  3. Changed PROC SQL to PROC FEDSQL for CAS execution
  4. Changed PROC FREQ to PROC FREQTAB for CAS execution
  5. Converted PROC LOGISTIC to PROC LOGSELECT for CAS execution.
  6. Added subsequent DATA Step **to call score code** using %INCLUDE statement.

  - CAS Enabled DATA Step and Base PROCS
    - 18 hours -> 7 hours
  - CAS Enabled PROC LOGISTICS (PROC LOGSELECT)
    - 38 hours -> 2 hours

§.sas

Leveraging SAS® Viya to improve Base SAS processes
Conclusion - CAS is Fast

- ## Health Care User Case

  - Base SAS Process
    - 70% DATA Step
    - SAS Workspace server
      - Runtime ~ **4 weeks**

- ## Business Benefit

  - Moved from quarterly reporting to monthly reporting

- ## SAS Viya

  1. Changed source and target tables in DATA Step to use CASLIB instead of 9.4 LIBREF

  2. Changed PROC SQL to PROC FEDSQL for CAS execution

  - **CAS**

    - **Runtime 1 week**

§.sas

Leveraging SAS® Viya to improve Runtimes of SAS processes
Agenda

1. Terminology

2. Findings from SAS Viya Readiness Utility

3. Customer User Cases

4. **Reading Material**

Leveraging SAS® Viya to improve Base SAS processes
Reading Material

1. Parallel Programming with the DATA Step: Next Steps (2018)
2. The following Base SAS procedures are CAS Enabled (2018)
3. SAS® Viya™ 3.3: FedSQL Programming for SAS® Cloud Analytic Services (2018)
   a) FedSql coding examples (2018)
4. FedSQL Implicit Pass-Through Facility in CAS (2018)
5. What's New in SAS® Data Connectors for SAS® Viya® (2018)
6. How to Emulate DESCENDING BY Variables in DATA Step Code that Runs Distributed in SAS® Viya™ (2018)
7. DATA Step in SAS® Viya™: Essential New Features (2017)
8. How to Emulate PROC APPEND in CAS (2017)
9. Getting your SAS 9 code to run multi-threaded in SAS Viya 3.3 (2017)
10. Jedi SAS Tricks - FedSQL Dictionary Tables (2017)
11. Six reasons you should stop using the RANUNI function to generate random numbers (2013)

§.sas