

A New Gateway to Open Source in SAS Viya

Jim Box

EU Connect 2025
CT07



Running Open-Source Code in SAS Programs



PROC IML in 9.4 / Viya Compute Context



PROC PYTHON in Viya Cloud Analytics Services

New Approach

Gateway CAS Action Set – Viya CAS Engine

Cross-Language Execution – Submit R or Python code in an external program or a submit block

Parallel Processing – utilize CAS parallel processing to run R or Python in a multi-threaded execution

Flexibility – Specify the number of threads to use, timeout settings, and logging levels

Single Threaded Approach

Run some R models

But First, what are we working with

Examine the R Environment

```
1  ⊖ proc iml;  
2      submit / R;  
3          ver = R.version.string  
4          vdf = data.frame(version = ver)  
5          pkg = installed.packages()[,c(1,3)]  
6      endsubmit;  
7  call ImportDataSetFromR("R_pkg", "pkg");  
8  call ImportDataSetFromR("R_ver", "vdf");  
9  quit;  
10  
11 ⊖ data _null_;  
12 set R_ver;  
13 CALL SYMPUTX("RVer",version);  
14 RUN;  
15  
16  
17 TITLE "R Version:  &Rver";  
18 proc print data = R_pkg; run;  
--
```

R Version: R version 4.3.3 (2024-02-29)

Obs	Package	Version
1	arrow	18.1.0
2	askpass	1.2.1
3	assertthat	0.2.1
4	backports	1.5.0
5	base	4.3.3
6	base64enc	0.1-3
7	BH	1.87.0-1
8	BiocManager	1.30.25
9	bit	4.5.0.1
10	bit64	4.5.2
11	blob	1.2.4
12	boot	1.3-29
13	broom	1.0.7
14	bslib	0.8.0
15	cachem	1.1.0
16	callr	3.7.6
17	caret	7.0-1
18	cellranger	1.1.0
19	chk	0.9.2
20	class	7.3-22
21	cli	3.6.3

PROC CAS Approach

Specific Call to Gateway Action

```
73  /* PROC CAS call to the model.r program */
74  %LET caslib = PUBLIC;
75  %LET castbl = HEART;
76  %let modelCode = glm(as.factor(Status) ~ AgeAtStart + Systolic + Diastolic + Weight + Cholesterol + Sex, data=df,family = binomial);
77
78  ⊖ proc cas;
79      filename rcode '/nfsshare/sashls2/home/jimbox/model.r'; ←-----→
80      code=readfile('rcode');
81      action gateway.runLang /
82          args={caslib="&caslib", castbl = "&castbl", mod="&modelCode"},
83          code=code,
84          lang="R",
85          log_level="INFO",
86          single=true;
87  run;
88
89  quit;
```

```
var <- gw$args
print(var)
```

```
caslib <- paste(var['caslib'])
castbl <- paste(var['castbl'])
```

```
df <- read_table(list(name= castbl,caslib = caslib))
print(head(df)) ←-----→
```

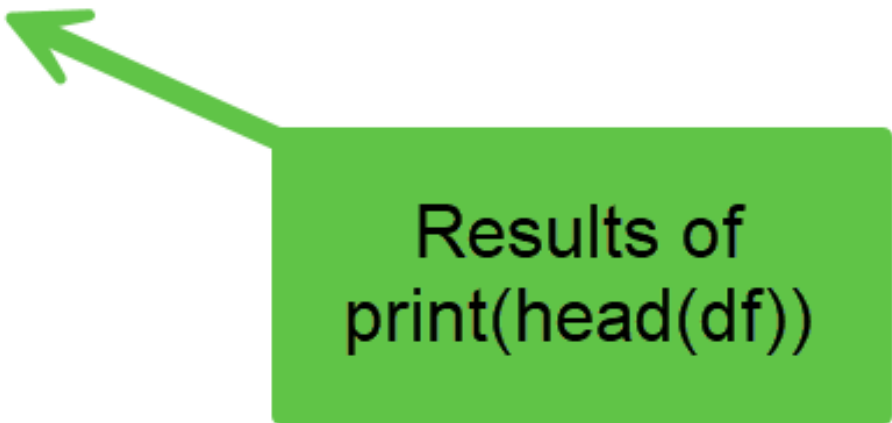
```
model = eval(parse(text = var['mod']))
```

```
print(summary(model)) ←-----→
```

PROC CAS Approach

Specific Call to Gateway Action

```
NOTE: $caslib
NOTE: [1] "PUBLIC"
NOTE: $castbl
NOTE: [1] "HEART"
NOTE: $mod
NOTE: [1] "glm(as.factor(Status) ~ AgeAtStart + Systolic + Diastolic + Weight + Cholesterol + Sex, data=df,family = binomial)"
NOTE:   Status DeathCause AgeCHDdiag   Sex AgeAtStart Height Weight Diastolic
NOTE: 1   Dead      Other      NA Female      29  62.50   140     78
NOTE: 2   Alive      <NA>      NA  Male      42  66.00   156     76
NOTE: 3   Alive      <NA>      NA  Male      35  71.00   194     68
NOTE: 4   Alive      <NA>      55  Male      33  70.00   174     90
NOTE: 5   Alive      <NA>      NA  Male      40  66.25   151     72
NOTE: 6   Alive      <NA>      NA Female      37  66.25   148     78
NOTE:   Systolic MRW Smoking AgeAtDeath Cholesterol Chol_Status BP_Status
NOTE: 1      124 121      0      55          NA      <NA>    Normal
NOTE: 2      110 116     20      NA        281      High    Optimal
NOTE: 3      132 124      0      NA        211  Borderline  Normal
NOTE: 4      142 114      0      NA        188  Desirable   High
NOTE: 5      132 112     30      NA        192  Desirable   Normal
NOTE: 6      110 112     15      NA        192  Desirable   Optimal
NOTE:   Weight_Status   Smoking_Status
NOTE: 1   Overweight      Non-smoker
NOTE: 2   Overweight      Heavy (16-25)
NOTE: 3   Overweight      Non-smoker
NOTE: 4   Overweight      Non-smoker
NOTE: 5   Overweight Very Heavy (> 25)
NOTE: 6   Overweight      Moderate (6-15)
NOTE: Call:
NOTE: glm(formula = as.factor(Status) ~ AgeAtStart + Systolic + Diastolic +
```



Results of
print(head(df))

PROC CAS Approach

Specific Call to Gateway Action

```
NOTE: Call:
NOTE: glm(formula = as.factor(Status) ~ AgeAtStart + Systolic + Diastolic +
NOTE:      Weight + Cholesterol + Sex, family = binomial, data = df)
NOTE: Coefficients:
NOTE:      Estimate Std. Error z value Pr(>|z|)
NOTE: (Intercept) -8.6616892  0.3393746 -25.523  < 2e-16 ***
NOTE: AgeAtStart   0.1068458  0.0044927  23.782  < 2e-16 ***
NOTE: Systolic     0.0151530  0.0024647   6.148 7.85e-10 ***
NOTE: Diastolic    0.0065554  0.0043500   1.507  0.132
NOTE: Weight      -0.0010776  0.0013689  -0.787  0.431
NOTE: Cholesterol  0.0019217  0.0007645   2.514  0.012 *
NOTE: SexMale      0.9025547  0.0764243  11.810  < 2e-16 ***
NOTE: ---
NOTE: Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
NOTE: (Dispersion parameter for binomial family taken to be 1)
NOTE:      Null deviance: 6708.0  on 5050  degrees of freedom
NOTE: Residual deviance: 5368.7  on 5044  degrees of freedom
NOTE:      (158 observations deleted due to missingness)
NOTE: AIC: 5382.7
NOTE: Number of Fisher Scoring iterations: 4
```

Results of
Print(summary(model))

Q4

PROC CAS Approach

Include R Code in PROC CAS Call

```
95 ① PROC CAS;  
96  externalsource Rprog1;  
97  
98  df <- read_table(list(name= "HEART",caslib = "PUBLIC"))  
99  
100 model <- glm(as.factor(Status) ~ AgeAtStart + Systolic + Diastolic + Weight + Cholesterol + Sex,  
101              data=df,  
102              family = binomial)  
103  
104 print(summary(model))  
105 mdf = as.data.frame(summary(model)$coefficients)  
106 mdf$Vars = rownames(mdf)  
107  
108 mdf <- mdf[, c("Vars", names(mdf)[-ncol(mdf)])]  
109 mdf  
110  
111 write_table(mdf, list(name = 'Heart_Model', caslib = 'casuser', replace = TRUE))  
112  
113 endexternalsource;  
114  
115     action gateway.runLang /  
116  
117         code=Rprog1,  
118         lang="R",  
119         log_level="INFO",  
120         single=true;  
121 run;  
122  
123 PROC PRINT data = casuser.heart_model; run;  
124
```

PROC CAS Approach

Model Output in a CAS Table for Printing

Obs	Vars	Estimate	Std. Error	z value	Pr(> z)
1	(Intercept)	-8.661689167	0.3393746156	-25.5225016	1.10932E-143
2	AgeAtStart	0.1068458465	0.004492739	23.781895035	5.14176E-125
3	Systolic	0.015152978	0.0024647219	6.1479464042	7.849255E-10
4	Diastolic	0.0065554329	0.0043499851	1.5070012315	0.1318103415
5	Weight	-0.001077557	0.001368858	-0.787194096	0.4311682528
6	Cholesterol	0.001921728	0.0007645438	2.513561558	0.0119518917
7	SexMale	0.9025547243	0.0764243497	11.809779576	3.47471E-32

PROC GATEWAY Approach

Cleaner Presentation

```
131
132 ⊖ PROC GATEWAY single lang = R;
133
134 submit;
135
136 df <- read_table(list(name= "HEART",caslib = "PUBLIC"))
137
138 model <- glm(as.factor(Status) ~ AgeAtStart + Systolic + Diastolic + Weight + Cholesterol + Sex,
139             data=df,
140             family = binomial)
141
142 mdf = as.data.frame(summary(model)$coefficients)
143 mdf$Vars = rownames(mdf)
144
145 mdf <- mdf[, c("Vars", names(mdf)[-ncol(mdf)])]
146
147 write_table(mdf, list(name = 'Heart_Model', caslib = 'casuser', replace = TRUE))
148
149 endsubmit;
150
151 run;
152
```

Multi-Threaded Approach

Run some R models in a threaded fashion

Multi Threading R-Code

Efficiency and Speed Gains, Especially with Large Data or Complex Tasks



Faster Execution via distributed compute workload



Better Resource Utilization – take advantage of your cores



Scalability as data accumulates over time



Leverage Iterative Processes – e.g. simulations or bootstrapping

Multi Threading R-Code

Key Code Changes in PROC GATEWAY

```
187  /** Submitting the code to be multi threaded Via Proc Gateway */
188
189  ⊖ PROC GATEWAY NTHREADS=2 lang=R;
190  submit;
191
192      df <- read_table(list(caslib= 'public', name= 'heart', groupby='Sex', groupbyMode = 'redistribute'))
193
194      print(gw$num_threads)
195
196      if(nrow(df)>0){
197
198          cat("Thread", gw$thread_id)
199
200          model <- glm(as.factor(Status) ~ AgeAtStart + Systolic + Diastolic + Weight + Cholesterol,
201                      data=df,
202                      family = binomial)
203
204          assign(paste("model_",df$Sex[1], sep = ""),model)
205
206          mdf = as.data.frame(summary(assign(paste("model_",df$Sex[1], sep = ""),model))$coefficients)
207          mdf$Vars = rownames(mdf)
208          mdf <- mdf[, c("Vars", names(mdf)[-ncol(mdf)])]
209
210          assign(paste("mdf_",df$Sex[1], sep = ""),mdf)
211
212          print(df$Sex[1])
213          head(assign(paste("mdf_",df$Sex[1], sep = ""),mdf), 10)
214      }
215
216  endsubmit;
217  RUN;
218
```

Multi Threading R-Code

Results in PROC GATEWAY

NOTE: Thread 5[1] "Male"						
NOTE:	<u>Vars</u>	Estimate	Std. Error	z value	Pr(> z)	
NOTE:	(Intercept)	(Intercept)	-1.011163e+01	0.605165457	-16.7088742	1.129478e-62
NOTE:	AgeAtStart	AgeAtStart	1.244938e-01	0.006471886	19.2360864	1.846336e-82
NOTE:	Systolic	Systolic	2.646623e-02	0.004062030	6.5155193	7.243861e-11
NOTE:	Diastolic	Diastolic	1.322024e-03	0.006372544	0.2074562	8.356536e-01
NOTE:	Weight	Weight	-4.192031e-04	0.002077309	-0.2018010	8.400723e-01
NOTE:	Cholesterol	Cholesterol	3.565731e-03	0.001175513	3.0333397	2.418632e-03

NOTE: Thread 4[1] "Female"						
NOTE:	<u>Vars</u>	Estimate	Std. Error	z value	Pr(> z)	
NOTE:	(Intercept)	(Intercept)	-7.5979906517	0.414528775	-18.3292237	4.837596e-75
NOTE:	AgeAtStart	AgeAtStart	0.0939128686	0.006724172	13.9664594	2.497369e-44
NOTE:	Systolic	Systolic	0.0112491877	0.003205295	3.5095637	4.488425e-04
NOTE:	Diastolic	Diastolic	0.0077942557	0.006067017	1.2846932	1.988995e-01
NOTE:	Weight	Weight	-0.0006085457	0.001855482	-0.3279717	7.429330e-01
NOTE:	Cholesterol	Cholesterol	0.0015808878	0.001049167	1.5068035	1.318610e-01
NOTE: WARNING: ignoring environment value of R_HOME						

Gateways to Open Source

Final Thoughts

Takeaways

Gateway CAS Action is a powerful way to Submit Open-Source Code

PROC GATEWAY is most elegant approach

Code can be in remote file or in Submit Block

Syntax is consistent across Python and R submissions

Ability to multi-thread is very impactful

Thanks!

EU Connect 2025
CT07

