TEN TIPS FOR EFFICIENT SAS CODE
Host

Vanessa Porter
Presenter

Peter Hobart
• In Listen Mode

• Control bar opened with the white arrow in the orange box
Efficiency Overview

- Optimisation has four competing factors
  - CPU
  - Memory
  - I/O (disk and network)
  - Disk space
EFFICIENCY

OVERVIEW

• Basic principles:
  • Don't do more work than you need to
  • Optimise for your environment
  • Go for the quick wins first
    • Jobs which take the most time
    • Jobs which are run most often
  • Benchmark after each change
Data Integration Studio produces detailed information on resources used and time taken by each step.

In base SAS and SAS Enterprise Guide, use the `options fullstimer` to get more information on the log.
NOTE: The data set WORK.SAMPLE2 has 317223 observations and 27 variables.

NOTE: DATA statement used (Total process time):
real time 2.10 seconds
cpu time 0.28 seconds

NOTE: The data set WORK.SAMPLE2 has 317223 observations and 27 variables.
NOTE: DATA statement used (Total process time):
real time 0.27 seconds
cpu time 0.28 seconds

NOTE: The data set WORK.SAMPLE2 has 317223 observations and 27 variables.
NOTE: DATA statement used (Total process time):
real time 0.27 seconds
cpu time 0.28 seconds

On the second and third runs the input data set was cached so the runtime was much lower. On a larger data set this effect will be reduced.
EFFICIENCY PROGRAMMER TIME VS. PROGRAM TIME

Simple:
Reads every row, outputs every 3rd row

```
Data sample1;
  set orion.customer_orders;
  if mod(_n_,3)=1;
run;
```

Efficient:
Reads every 3rd row

```
Data sample2;
  do i = 1 to rows by 3;
    set orion.customer_orders point=i nob=rows;
    output;
  end;
  stop;
run;
```

Both produce a 1/3 sample
data profits;
    set orion.customer_orders;
    sale = quantity*retailPrice;
run;

proc means data=profits nonobs maxdec=2;
    where month=12;
    class continent;
    var sale;
run;

TIP 2  SELECT INPUT COLUMNS WITH DROP AND KEEP

data profits;
  set orion.customer_orders;
  where month=12;
  sale = quantity*retailPrice;
run;

proc means data=profits nonobs maxdec=2;
  class continent;
  var sale;
run;

data profits;
  set orion.customer_orders(keep= month quantity retailPrice continent);
  where month=12;
  sale = quantity*retailPrice;
run;

proc means data=profits nonobs maxdec=2;
  class continent;
  var sale;
run;
## TIP 3  SELECT OUTPUT COLUMNS

```sas
data profits;
   set orion.customer_orders
       (keep= month quantity retailPrice continent);
   where month=12;
   sale = quantity*retailPrice;
run;

proc means data=profits nonobs maxdec=2;
   class continent;
   var  sale;
run;
```

```sas
data profits (keep=sale continent);
   set orion.customer_orders
       (keep= month quantity retailPrice continent);
   where month=12;
   sale = quantity*retailPrice;
run;

proc means data=profits nonobs maxdec=2;
   class continent;
   var  sale;
run;
```
**TIP 4  WHERE VS IF**

Read everything then select

```sas
data profits;
  set orion.customer_orders;
  sale = quantity*retailPrice;
  if month=12;
run;
```

Read only the required rows

```sas
data profits;
  set orion.customer_orders;
  sale = quantity*retailPrice;
  where month=12;
run;
```
Read everything then select

```sas
data profits;
    set orion.customer_orders;
    sale = quantity*retailPrice;
    if sale > 200;
run;
```

sale is calculated then evaluated

Read only the required rows

```sas
data profits;
    set orion.customer_orders;
    sale = quantity*retailPrice;
    where sale > 200;
run;
```

sale does not exist in the source table

But...we could use

```sas
where quantity*retailPrice > 200
```
TIP 5  USE INDEXES

No index

data profits;
set orion.customer_orders;
sale = quantity*retailPrice;
where month=12;
run;

Index on month

data profits;
set orion.customer_orders;
sale = quantity*retailPrice;
where month=12;
run;

NOTE: DATA statement used (Total process time):
  real time       2.43 seconds
  cpu time        0.26 seconds

NOTE: DATA statement used (Total process time):
  real time       0.24 seconds
  cpu time        0.07 seconds
Recreates the data and builds an index

data orion.customer_orders
   (index=(month));
set orion.customer_orders;
r

Reads the data and builds an index

proc datasets lib=orion noprint;
   modify customer_orders;
   index create month;
r
quit;
TIP 7  YOU CAN SUBSET DURING A SORT

```sas
proc sort data=orion.customer_orders
   out=orders_sorted
   (drop=continent);
   by month;
   where continent="Europe";
run;
```

Danger!
Sorting a data set without specifying `out=` can result in data loss.
TIP 8  OPTIMISE CONDITIONAL LOGIC

Tests every condition

```sas
data profits;
  set orion.customer_orders;
  if quantity = 1 then order="Sml";
  if quantity = 2 then order="Med";
  if quantity = 3 then order="Lrg";
  if quantity > 3 then order="XXL";
run;
```

Tests only until a condition is TRUE

```sas
data profits;
  set orion.customer_orders;
  if quantity = 1 then order="Sml";
  else if quantity = 2 then order="Med";
  else if quantity = 3 then order="Lrg";
  else if quantity > 3 then order="XXL";
run;
```

All four tests are performed on every row

If on the current row quantity=1 then only one test is performed
Subset

```sas
data profits;
  set orion.customer_orders;
  if quantity = 1 then order="Sml";
  else if quantity = 2 then order="Med";
  else if quantity = 3 then order="Lrg";
  else if quantity > 3 then order="XXL";
run;
```

Subset early

```sas
data profits;
  set orion.customer_orders;
  if quantity = 2 then order="Med";
  else if quantity = 1 then order="Sml";
  else if quantity = 3 then order="Lrg";
  else if quantity > 3 then order="XXL";
run;
```

Most orders in this data contain 2 items. By testing for this value first, less tests will be processed overall.
TIP 10  CONSIDER COMPRESSION

• Compressing data:
  • Reduces disk space used
  • Increases CPU
  • *Decreases IO*
    • May introduce changes in behaviour in "edit in place"

• SAS default compression is deliberately "light touch"
  • Long character variables will usually compress well
  • Numbers will not compress with the default algorithm
  • Compression introduces a space overhead: the compression achieved should outweigh this
  • Look at the log to check the compression achieved
  • Poor candidates for compression can grow instead
  • Check `bufsize` is not excessively large, if compression is less than expected.
TIP 10  CONSIDER COMPRESSION

data customer_orders (compress=yes);
    set orion.customer_orders;
run;

NOTE: Compressing data set WORK.CUSTOMER_ORDERS decreased size by 41.70 percent.
Compressed is 3603 pages; un-compressed would require 6180 pages.

NOTE: DATA statement used (Total process time):
    real time 2.57 seconds
    cpu time 1.59 seconds

Compress = yes or compress= char uses run length encoding
TIP 10  CONSIDER COMPRESSION

```sas
data customer_orders (compress=binary);
  set orion.customer_orders;
run;
```

NOTE: Compressing data set WORK.CUSTOMER_ORDERS decreased size by 46.21 percent.
Compressed is 3324 pages; un-compressed would require 6180 pages.

NOTE: DATA statement used (Total process time):
  real time       2.62 seconds
  cpu time        2.27 seconds

Compress = binary uses Ross data compression
- Slightly better compression
- Slightly increased CPU
• Check the results are from a successful run, not a cached data set
• Every site has a different mix of constraints
• Measurements can be distorted by other running jobs competing for resources
• Benchmark changes by running several times in real-world situations
• Test each change in isolation
THANK YOU FOR ATTENDING
Free SAS Tutorials

View SAS training "how to" videos, tutorials, and demos to learn tips and tricks for working with SAS software.
### Latest Activity

#### Featured Posts

1. **Hypothesis testing with ASE?**
   - In: SAS Data Mining
   - Latest post by: Benjamin8
   - New: 0, Replies: 7, Likes: 0

2. **SAS High-Performance Analytics Tip #1: How it diff...**
   - In: SAS Communities Library
   - Latest post by: RadhikaMyneni
   - New: 0, Replies: 0, Likes: 3

3. **New releases for SAS Data Management products**
   - In: SAS Communities Library
   - Latest post by: RonAgrega
   - New: 0, Replies: 0, Likes: 0

4. **How to create a connection to SAS Data Sets in SAS...**
   - In: SAS Communities Library
   - Latest post by: Audrey
   - New: 0, Replies: 0, Likes: 0

5. **How to call a SAS Data Management 2.6 Web Service ...**
   - In: SAS Communities Library
   - Latest post by: StephenFoerster
   - New: 0, Replies: 0, Likes: 1
SAS Technical Support Form

1. Basic Information
   Use this form to create a track with SAS Technical Support. Click here to update an existing track.

2. Problem Description
   Before you proceed you should:
   - Report critical problems by telephone
   - Verify that any SPAM software on your machine will not block our e-mail responses

   Shortly after you submit the form, you will receive an automatic e-mail that:
   - Confirms that you have submitted the form successfully
   - Provides a tracking number that has been assigned to your e-mail request

   After you supply the basic information, click Next Page to proceed. Otherwise click Reset to start over.

Basic Information

Have a SAS profile? Log in to prefill the form or click here to create one.
WHERE TO GO FOR HELP

SAS 9.4 Product Documentation

Starting Points
- Product Index A-Z
- Programmer’s Bookshelf
- What’s New in SAS
- Documentation by Title

Syntax Shortcuts
- Syntax Lookup
- SAS Procedures by Name and Product
- SAS Language Elements by Name, Product, and Category

Search
- Enter search term

Product
- All Products

Display
- All topics
- Examples only
- Syntax only
- What’s New topics

Submit
• www.sas.com/uk/usergroups

Platform Administration
23rd February 2016 – London
2nd March 2016 – Manchester
SAVE THE DATE

June 14th, 15th 16th