



PROC DS2–A powerful tool you need to know

Weston Chen, Novartis

PharmaSUG China 2016
Paper 51X–F5J2A9E6C6

Agenda

- ▶ What's the DS2?
- ▶ Comparing to the traditional data step
- ▶ Examples using DS2
- ▶ Summary



What's the DS2

- ▶ SAS proprietary programming language for data manipulation and data modeling applications from SAS v9.2(TS2M3);
- ▶ Sharing core features similar to the DATA step;
- ▶ Having several features that are not available in the DATA step:
 - For example, variable scoping, user-defined methods, ANSI SQL data types, user-defined packages, and programming structure elements.

Benefit of using DS2

DS2 is particularly suited for the cases below:

- ▶ Requiring the precision from new supported data types;
- ▶ Using new expressions or writing methods or packages;
- ▶ Being able to execute SAS SQL from within the DS2 program;
- ▶ Taking advantage of threaded processing in products;



DS2 Syntax

– DS2 method

- ▶ Every DS2 program may contain these three methods:

```
method init();  
end;
```

← Initiate variables or open tables

```
method run();  
end;
```

← Be functional equivalent of the DATA step, running as an implicit loop

```
method term();  
end;
```

← Run once at the end of the DS2 program

DS2 Syntax

– DS2 DECLARE Statement and Scope

- ▶ When a DECLARE statement defines a variable, the variable assumes the scope of the programming block in which the variable is declared.

```
proc ds2;
```

```
declare int x;
```

```
method init();
```

```
declare double d;
```

```
...;
```

```
end;
```

```
...;
```

```
run;
```

```
quit;
```

← A variable that is declared outside a method has global scope within that DS2 programming block.

← A variable that is declared in a method has local scope and exists only when the method executes.

Comparing to the Traditional DATA Step

- ▶ Sharing many language elements and those elements behave in the same way:
 - Formats, functions, statements like DATA, SET, KEEP, DROP, RUN, BY, RETAIN, PUT, OUTPUT, DO, IF-THEN/ELSE, Sum, and others;
- ▶ Performing most of the DATA step tasks in the same manner:
 - variable arrays, multi-dimensional arrays, hash tables, expressions, date and time values, conversion between data types.



Comparing to the Traditional DATA Step

```
data new;  
  retain max;  
  keep max;  
  if _n_=1 then max=0;  
  set class end=last;  
  if weight>max then max=weight;  
  if last then output;  
run;
```

```
proc ds2;  
  data new2(overwrite=yes);  
  → dcl double max;  
  retain max;  
  keep max;  
  method init();  
    max=0;  
  end;  
  method run();  
    set class;  
    if weight>max then max=weight;  
  end;  
  method term();  
    output;  
  end;  
run;  
quit;
```

Comparing to the Traditional DATA Step

```
data new;  
  set class;  
  by descending age  
  height weight;  
  if _n_ = 10 then stop;  
run;
```

ERROR: BY variables are not properly sorted on data set WORK.CARS.

```
proc ds2;  
  data new2;  
  method run();  
    set class;  
    by descending age height  
    weight;  
    if _n_ = 10 then stop;  
  end;  
enddata;  
run;  
quit;
```

This program runs just fine.

More Examples using DS2

–using user–defined method

```
/* converting Celsius temperature to Fahrenheit temperature */
```

```
proc ds2;
```

```
data _null_;
```

```
method c_to_f(double Tc) returns double;  
  return (((Tc*9)/5)+32);  
end;
```

← Define use–defined method

```
method init();  
  dcl double Degc DegF;  
  do DegC=0 to 40 by 10;  
    DegF=c_to_f(DegC);  
    put DegC= DegF=;  
  end;  
end;
```

← Call use–defined method

```
enddata;
```

```
run;
```

```
quit;
```

More Examples using DS2

-Using parallel processing to reduce run time

- ▶ A testing on DATA step, DS2 without parallel processing, DS2 with parallel processing
 - data set: SRC with 1,000,000 observations
 - variables: id, wt, ht

More Examples using DS2

- ▶ DATA step:

```
data datastep;  
  array score[0:100];  
  set src end=last;  
  do i=lbound(score) to hbound(score);  
    score[i]= (sqrt(((id * wt * ht) / (id + ht +  
wt))))*(sqrt(((wt * ht) / (i + ht + wt))*id));  
  end;  
  count+1;  
  if last then put 'Data step process ' count 'obs.';  
  drop i count;  
run;
```

More Examples using DS2

- ▶ DS2 step:

```
proc ds2;  
data withds2/overwrite=yes;  
  dcl bigint count;  
  drop count;  
  vararray double score[0:100] score0–score100;  
  method run();  
    dcl int i;set src;  
    do i=lbound(score) to hbound(score);  
      score[i]= (sqrt(((id * wt * ht) / (id + ht + wt))))*(sqrt(((wt * ht) / (i + ht + wt))*id));  
    end;  
    count+1;  
  end;  
  method term();  
    put 'DS2 Data step process' count 'obs.';  
  end;  
enddata;  
run;  
quit;
```

More Examples using DS2

- ▶ DS2 step with parallel processing:

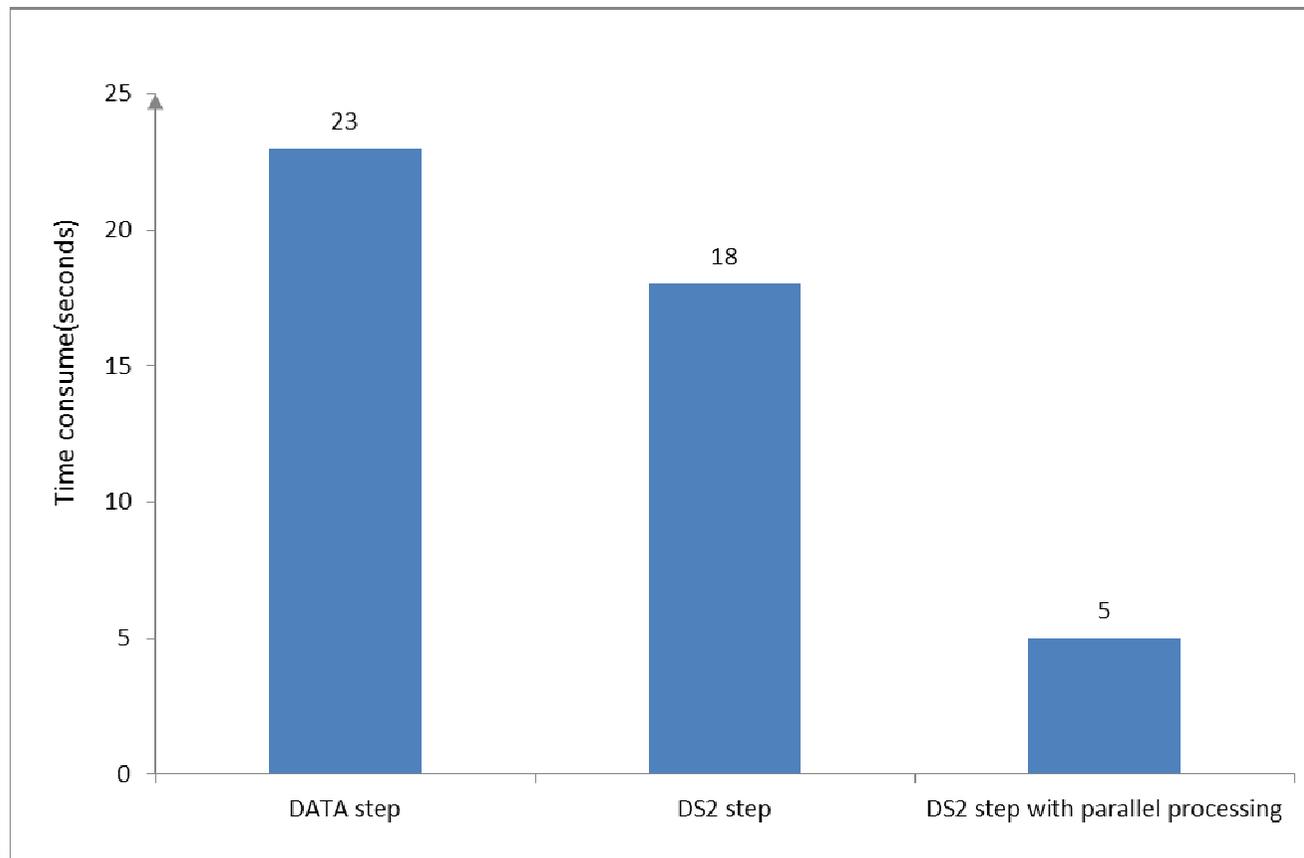
```
proc ds2;
```

```
  thread thrd/overwrite=yes;  
    dcl bigint count;drop count;  
    vararray double score[0:100] score0–score100;  
    method run();  
      dcl int i;set src;  
      do i=lbound(score) to hbound(score);  
        score[i]= (sqrt(((id * wt * ht) / (id + ht + wt))))*(sqrt(((wt * ht) / (i + ht + wt))*id));  
      end;  
      count+1;  
    end;  
    method term();  
      put 'Thread' _threadid_ ' parallel process' count 'obs.';  
    end;  
  endthread;
```

```
run;  
quit;
```

More Examples using DS2

- ▶ Comparison the running time



Summary

- ▶ DS2 is included with Base SAS and it shares some core features with the DATA step;
- ▶ Other than features sharing with DATA step, DS2 could provide more powerful features for advanced data manipulation;
- ▶ **For many programmers there is not a ‘need’ to move to DS2 programming, but the move will be worthwhile.**
 - Peter Eberhardt

Reference

- ▶ SAS® 9.4 DS2 Language Reference. SAS Institute Inc. 2013. Cary, NC: SAS Institute Inc.;
- ▶ Peter Eberhardt and Xue Yao “DS2 with Both Hands on the Wheel”, Proceedings of the SAS Global Forum 2015 Conference. Cary, NC: SAS Institute Inc.;
- ▶ Shaun Kaufmann “Object–Oriented Program Design in DS2”, Proceedings of the SAS Global Forum 2015 Conference. Cary, NC: SAS Institute Inc.;

Thank you!

Name: Weston Chen

Organization: Novartis

Address: Zhangjiang Hi-Tech Park

City, State ZIP: Shanghai

Work Phone: +86 21 61606410

E-mail: wei-2.chen@novartis.com

