



Global Solutions from the World's Largest Data-Focused CRO

## CT05 - Do Loops in Proc DS2

Presented by Fred Ross

# Agenda

- What is Proc DS2 & What can it do?
- Example of Multithreading
  - Dummy Data
  - Do Loop in a Data Step
  - Conversion to DS2
  - Conversion into Thread
  - Utilisation of Thread
  - Comparison of Run Times
- Limitations & Considerations for DS2
- Conclusion
- Questions

# What is Proc DS2 & What can it do?



Proc DS2 is a new procedure with full release in SAS 9.4. It's a completely new programming language within SAS.

- DS2 = Data step 2
- **Multi threading**
- Object oriented programming
- More advanced data processing and manipulation
- Embedded Proc FedSQL
- Additional data types

# Dummy Data

```
%let subjects=10000;  
%let visits=1000;  
data subjvisitdata;  
  do usubjid=1 to &subjects;  
    do avisitn=0 to &visits by 5;  
      output;  
    end;  
  end;  
run;
```

| usubjid | avisitn |
|---------|---------|
| 1       | 0       |
| 1       | 5       |
| 1       | 10      |
| ...     | ...     |
| 1       | 995     |
| 1       | 1000    |
| 2       | 0       |
| 2       | 5       |
| 2       | 10      |
| ...     | ...     |

# Do Loop in a Data Step

```
data dsetout1(drop=i);  
  set subjvisitdata;  
  do i=0 to &visits by 10;  
    if avisitn=i then anl01fl='Y';  
  end;  
run;
```

NOTE: DATA statement used (Total process time):

|           |              |
|-----------|--------------|
| real time | 1.68 seconds |
| cpu time  | 1.69 seconds |

| usubjid | avisitn | anl01fl |
|---------|---------|---------|
| 1       | 0       | Y       |
| 1       | 5       |         |
| 1       | 10      | Y       |
| ...     | ...     |         |
| 1       | 995     |         |
| 1       | 1000    | Y       |
| 2       | 0       | Y       |
| 2       | 5       |         |
| 2       | 10      | Y       |
| ...     | ...     |         |

# Conversion into DS2

```
proc ds2;  
  data dsetout2/overwrite=yes;  
    dcl char(2) anl01fl;  
    method run();  
      dcl int i;  
      set subjvisitdata;  
      do i=0 to &visits by 10;  
        if avisitn=i then anl01fl='Y';  
      end;  
    end;  
  enddata;  
run;  
quit;
```

|                                                |              |
|------------------------------------------------|--------------|
| NOTE: PROCEDURE DS2 used (Total process time): |              |
| real time                                      | 2.84 seconds |
| cpu time                                       | 2.23 seconds |

# Conversion into a Thread

```
proc ds2;
  data dsetout2/overwrite=yes;
    dcl char(2) anl01fl;
    method run();
      dcl int i;
      set subjvisitdata;
      do i=0 to &visits by 10;
        if avisitn=i then anl01fl='Y';
      end;
    end;
  enddata;
run;
quit;
```



```
proc ds2;
  thread thread1/overwrite=yes;
    dcl char(2) anl01fl;
    method run();
      dcl int i;
      set subjvisitdata;
      do i=0 to &visits by 10;
        if avisitn=i then anl01fl='Y';
      end;
    end;
  endthread;
```

# Utilisation of Thread

```
proc ds2;
  thread thread1/overwrite=yes;
    dcl char(2) anl01fl;
    method run();
      dcl int i;
      set subjvisitdata;
      do i=0 to &visits by 10;
        if avisitn=i then anl01fl='Y';
      end;
    end;
  endthread;
data dsetout3/overwrite=yes;
  dcl thread thread1 subjvisitdata;
  method run();
    set from subjvisitdata threads=4;
  end;
run;
quit;
```

NOTE: PROCEDURE DS2 used (Total process time):

|           |              |
|-----------|--------------|
| real time | 0.65 seconds |
| cpu time  | 2.57 seconds |

# Comparison of Run Times

Rough calculation of **Real/CPU** Time when using DS2

$$\begin{aligned} &= \\ &\quad \text{Time taken to call Proc DS2} \\ &\quad + \\ &\quad \text{Time taken to split data into threads} \\ &\quad + \\ &\quad \frac{\text{Time taken to process METHOD}}{\text{Number of threads}} \quad + \quad \text{Time taken to process METHOD across all threads} \\ &\quad + \\ &\quad \text{Time to merge data back together} \end{aligned}$$

|                                 | Real time | CPU time |
|---------------------------------|-----------|----------|
| Data step                       | 1 . 68    | 1 . 69   |
| Data step conversion to DS2     | 2 . 84    | 2 . 23   |
| DS2 + Multithreading 4 threads  | 0 . 65    | 2 . 57   |
| DS2 + Multithreading 18 threads | 0 . 62    | 3 . 12   |

# Limitations & Considerations for DS2



Quanticate  
The Clinical Data Experts

- System limitations
  - Does your system allow for multithreading?
- Conditional Do Loops
  - Solely conditional loops cannot be sped up through multithreading
- Recreating other procedures in DS2
  - Proc Freq
- Do you actually need it in your program?

# Conclusion

- Proc DS2 offers more advanced manipulation techniques
- Multithreading can allow for better utilisation of your system
- DS2 provides useful additional techniques to any SAS programmer
- Using the basic first steps outlined it is easy to pick up

# Questions



## Contact:

Name : Fred Ross

Company : Quanticate International Ltd.

Email : [Fred.Ross@Quanticate.com](mailto:Fred.Ross@Quanticate.com)