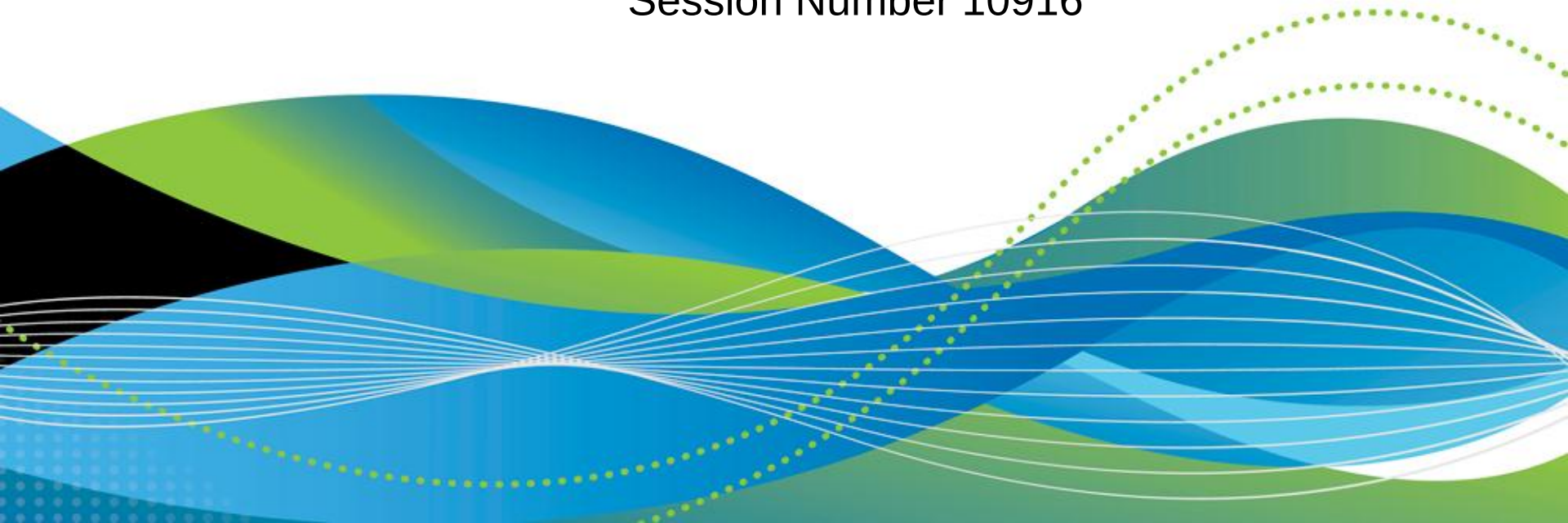
#SHAREorg

SHARE
Technology · Connections · Results

Capacity Planning Techniques for Growing SAS Workloads

Fred Forst - SAS Institute, Inc Cary, NC

March 16, 2012
Session Number 10916



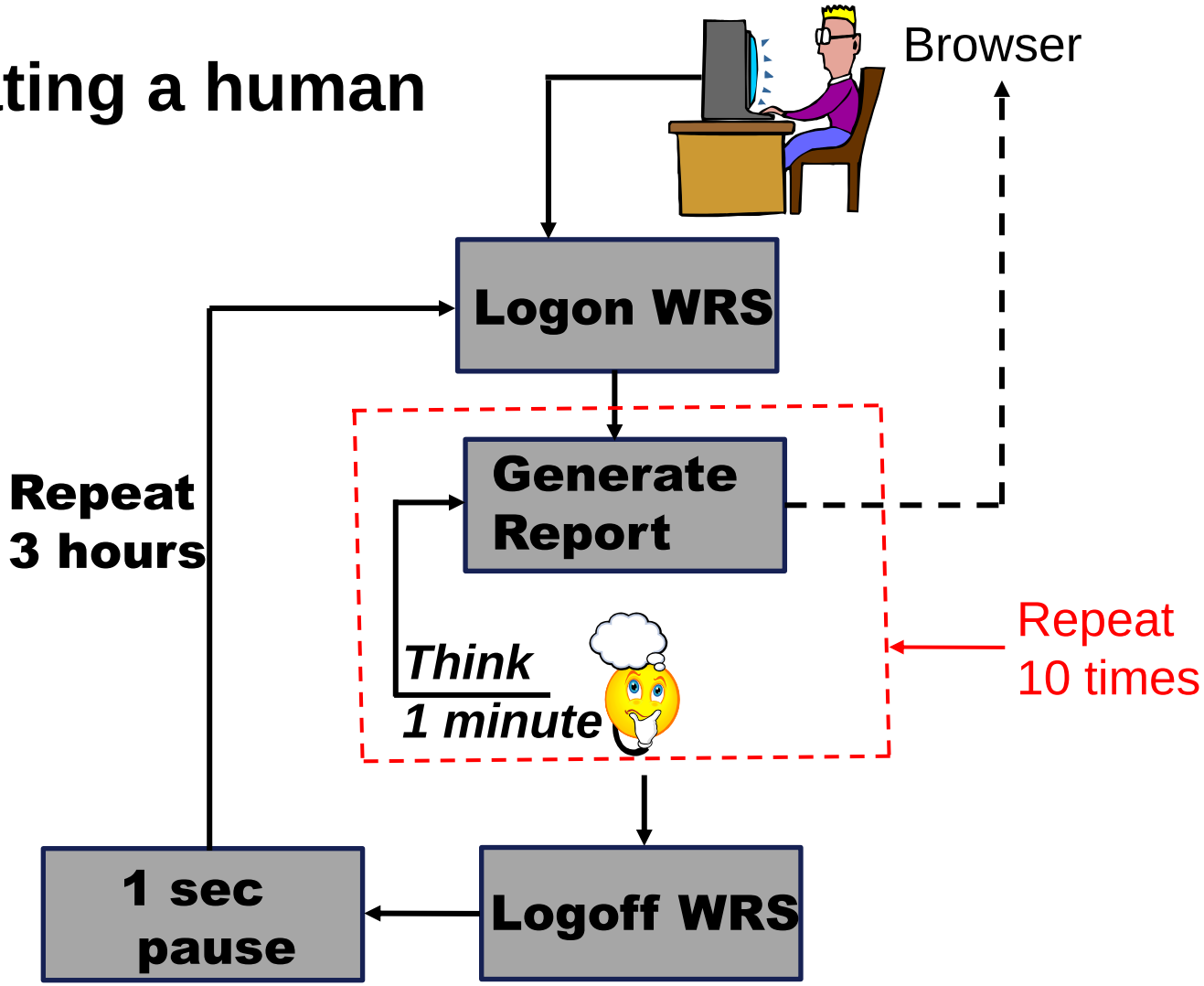
Capacity Planning Techniques for Growing SAS Workloads

- SAS Business Intelligence Workload
  - HP ALM Performance Center Load Generator
  - Analysis of results
- Non-BI Workloads
  - Know your workload
  - Data Gathering – What does workload look like?
    - SMF
    - ARM
  - SAS workload modeler macro. ‘What if’ scenarios.

BI Workload

- SAS Web Report Studio (WRS)
  - OLAP Cube data
  - Currently 25 users randomly chose amongst 90 reports
  - Each report submits at least one query to the SAS OLAP Server
  - Simulate a growing user base, 25 – 400 users
  - **What happens to report generation elapsed times?**
- HP ALM Performance Center®
  - Simulates the human generating random WRS reports
  - Produces summary & detail reports of WRS simulation

Simulating a human



z/OS

Logon WRS

Generate Report

Logoff WRS

HP Performance Center

Windows server

Pseudo C-code (script)

```
do user=1 to max_users
   logon WRS url
     do rpts=1 to num_reports
      select random report
      generate report
      calculate elapsed time
      think think_time
     end;

   logoff WRS url
     stop if > max_hours
     pause x_seconds
end;
```

Workflow Profile

- 90 SAS Web Report Studio Reports
- Four categories of reports:
  - Contract status reports (67% of users)
  - HR reports (18%)
  - IT Network reports (10%)
  - HR power user reports (5%)

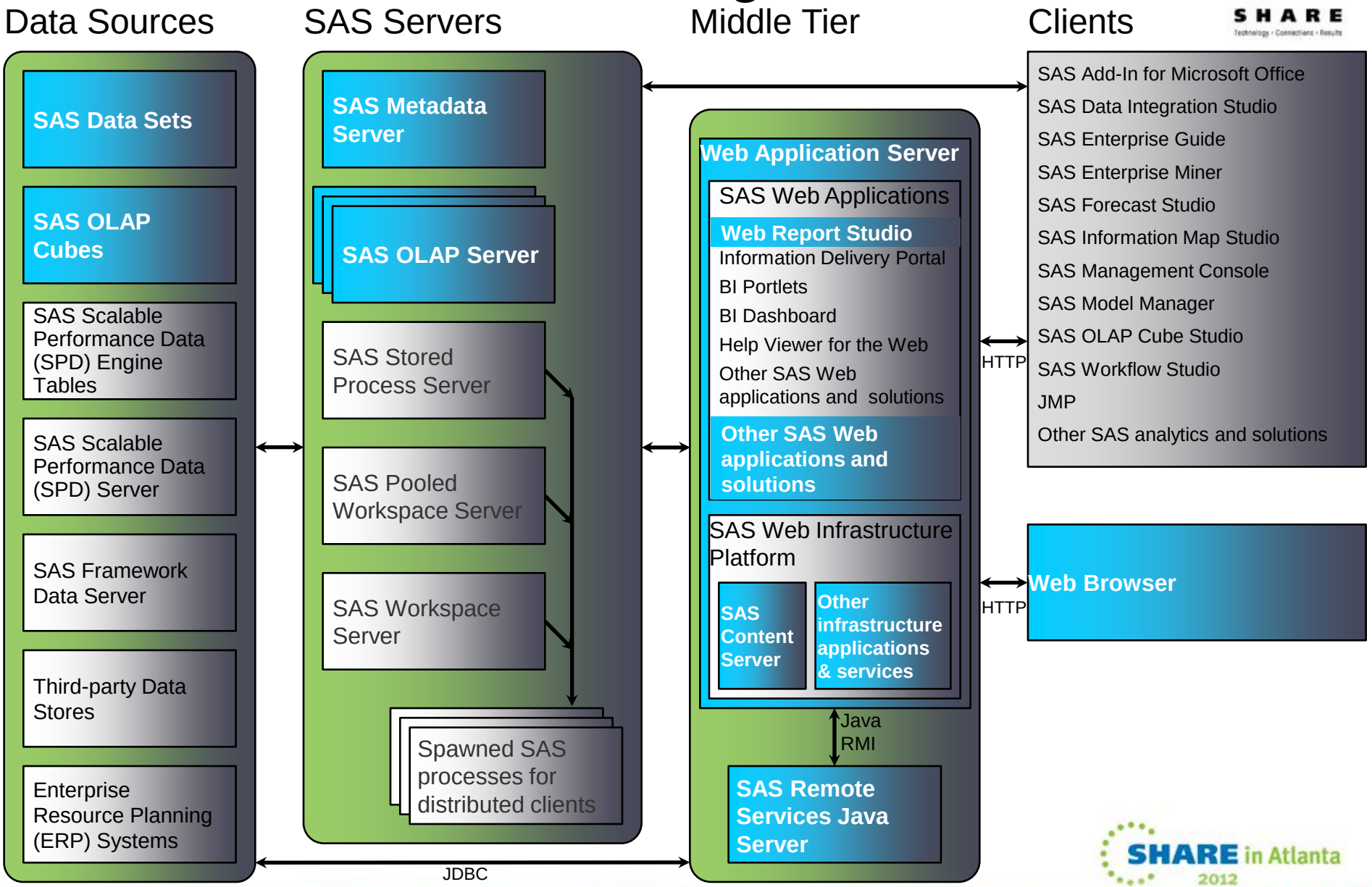
Hardware and OS Profile

- IBM® zEnterprise z196 model 710 5.2GHz CP
- Four dedicated CPs
- 48Gb Memory
- IBM System Storage DS8800 Subsystem
- Eight Ficon Express8 Channels
- z/OS 1.12
- IBM WebSphere Application Server V7
- JAVA V1.6

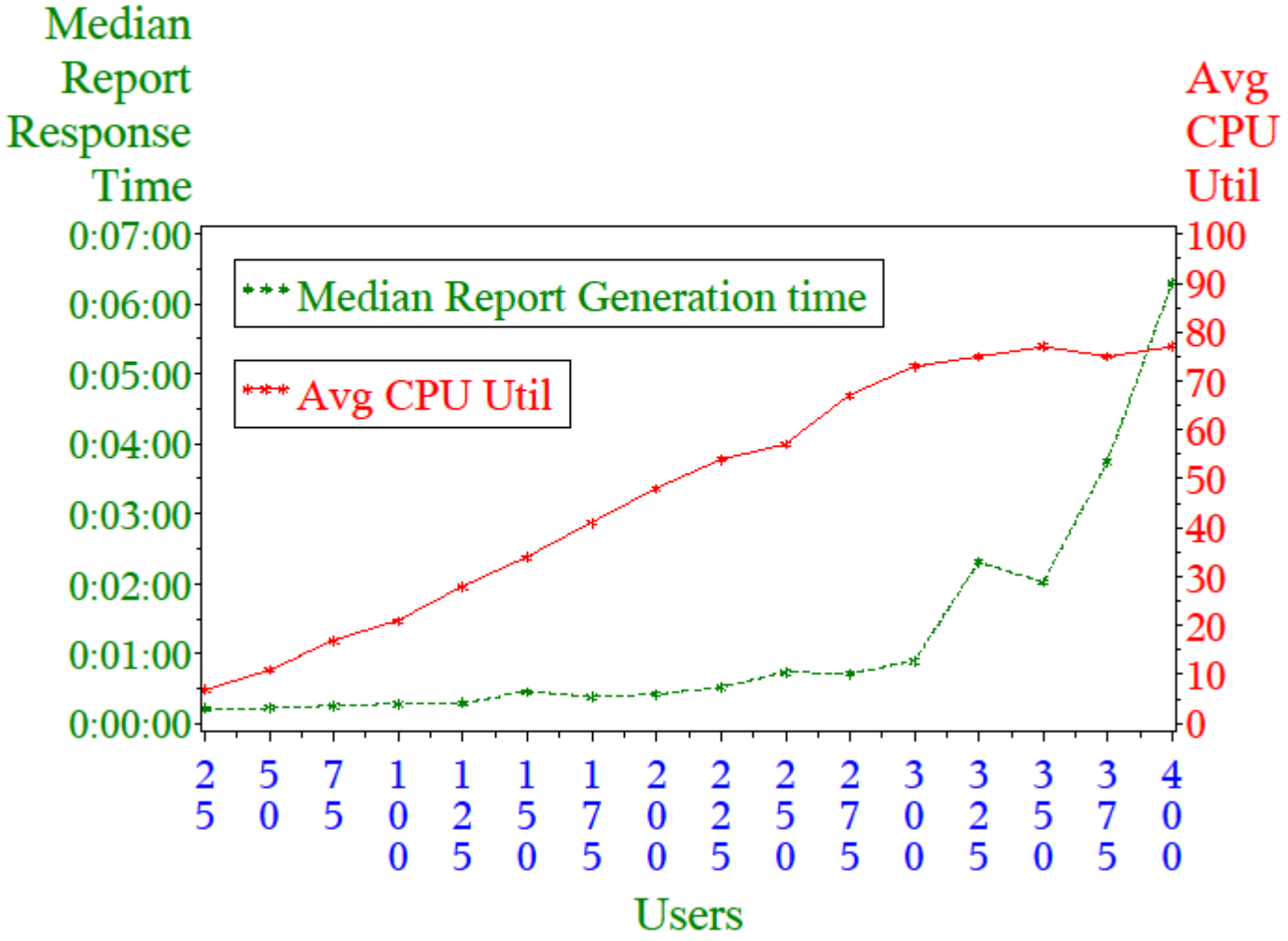
SAS Profile

- SAS V9.3
- Four SAS OLAP Servers load balanced
- SAS Metadata server (64 bit)
- SAS Web Report Studio

Architecture of SAS Intelligence Platform



Web Report Studio Simulations



CPU usage

- WebSphere – 62% of total (most eligible fo z/AAP)
- OLAP srv – 18%
- Metadata srv –17%
- Other – 2%

Analysis

- Common Sense approach
  - Elbow in elapsed time curve seems to be around 300 users

Users	Elapsed Time (avg report time)	
25	13.23sec	
300 (12x)	53.88 (4x)	← A 12x increase in users suffered only a 4x performance degradation
400 (16x)	378 (28x)	← A 16x increase in users suffered a 28x performance degradation

Conclusion: This hardware/software combination scales very well up to 300 users

Duncan Multiple Range Test

Good ‘ol T-Test

Hypothesis testing:

Avg Report time for 25 users same as for 50 users

H_0: $\mu_{25} = \mu_{50}$

H_1: $\mu_{25} \neq \mu_{50}$

Limitation: Tests the equality of only 2 groups

Duncan Multiple Range Test

H_0: $\mu_{25} = \mu_{50} = \mu_{75} \ldots = \mu_{400}$

H_1: $\mu_{25} \neq \mu_{50} \neq \mu_{75} \ldots \neq \mu_{400}$

Duncan Multiple Range test tests the hypothesis all groups are equal.

Duncan Multiple Range test gives statistical validity to results

Duncan Grouping	Mean	users
A	377.997	400
B	224.995	375
C	138.599	325
D	121.200	350
E	53.879	300
F	43.920	250
F		
F	42.659	275
G	31.319	225
H	27.643	150
H		
I H	24.959	200
I		
I	22.678	175
J	17.760	125
J		
J	17.098	100
J		
K J	15.413	75
K		
K	13.319	50
K		
K	13.230	25

75 users and 200 users members of 2 groupings

Equal groupings can overlap

Non-BI Workload

- Traditional SAS workload
- Don’t have HP Performance center
- Want to simulate growing # of users or growing data
- Step #1 – know your users
  - Batch or interactive?
  - What SAS PROCs do they use and how often?
  - Ad hoc usage or static jobs?
  - **What happens to SAS session elapsed times?**
  - **What happens to SAS PROC elapsed times?**

SAS log using option FULLSTIMER

```
NOTE: DATA statement used (Total process time):
      real time               2.97 seconds
      user cpu time           1.71 seconds
      system cpu time         0.68 seconds
      Memory                                  250k
      OS Memory                               6072k
      Timestamp                1/24/2012  9:12:35 AM
```

Want these historical metrics in a dataset

1. Parse user SAS logs
2. Use SMF option (z/OS only)
3. Use ARM (any OS)

Data Gathering using SMF

```
//STEPSAS EXEC SAS,
// OPTIONS='SMFEXIT=SMFEXIT SMF'
//SYSIN  DD *
     ...your SAS code…
//MXG     EXEC SAS9
//SMF      DD DISP=OLD,DSN=<SMF data>
//SOURCLIB DD DISP=SHR,DSN=MXG.SOURCLIB
//LIBRARY  DD DISP=SHR,DSN=MXG.LIBRARY
//SYSIN  DD *
%INC SOURCLIB(TYPESASU);
PROC PRINT  DATA=WORK.TYPESASU;
 VAR SASEXCPS SASCORE SASPROC SASTCBTM;
  FORMAT SASTCBTM 8.4;
```

Data Gathering using SMF

OBS	EXCPS USED BY PROC	MEMORY USED BY PROC	PROCEDURE NAME OR SASDATA	TCB TIME USED	ELAPSED
1	471	4766K	DATASTEP	0.0726	0.3500
2	834	3785K	SORT	0.1273	0.6600
3	11	4601K	DATASTEP	0.0012	0.0000
4	1587	13M	UNIVARIA	0.1513	0.6300

Data Gathering using ARM

```
%LET _ARMEXEC=1;
OPTIONS ARMLOC="/u/logs/user123.txt"     ;
options ARMSUBSYS=(ARM_PROC);

%PERFINIT;
%PERFSTRT(TXNNAME="PROC_GLM_TEST");
        …User code here…
%PERFSTOP;
```

Data Gathering using ARM

What is wrong with this picture?
Will users volunteer to code all the ARM stuff?

```
%LET _ARMEXEC=1;
OPTIONS
 ARMLOC="/u/logs/&sysuserid._&datetime.txt";
 ARMSUBSYS=(ARM_PROC);

%PERFINIT;
%PERFSTRT(TXNNAME="&get_creative");

        ...User code here...

%PERFSTOP;
```

Execute this via the `initstmt` option

Execute this via the `termstmt` option

```
I,1642709491.856633,1,0.049118,0.000000,SAS,FREFOR
G,1642709491.856716,1,1,PROCEDURE,PROC START/STOP,PROC_NAME,ShortStr,PROC_IO,Cou
nt64,PROC_MEM,Count64,PROC_LABEL,LongStr
I,1642709491.901724,2,0.054036,0.000000,SAS,FREFOR
G,1642709491.905480,2,2,share,,_IOCOUNT_,Count64,_MEMCURR_,Gauge64,_MEMHIGH_,Gau
ge64,_THREADCURR_,Gauge32,_THREADHIGH_,Gauge32
S,1642709491.905834,2,2,1,0.056871,0.000000,2370,21696512,21778432,1,1
S,1642709491.912629,1,1,2,0.057520,0.000000,DATASTEP,0,0,
P,1642709492.193007,1,1,2,0.130060,0.000000,0,DATASTEP,2835,24219648,
S,1642709492.199900,1,1,3,0.130486,0.000000,SORT    ,0,0,
P,1642709492.896566,1,1,3,0.259169,0.000000,0,SORT    ,3680,35074048,
S,1642709492.896874,1,1,4,0.259169,0.000000,DATASTEP,0,0,
P,1642709492.900411,1,1,4,0.260674,0.000000,0,DATASTEP,3689,35074048,
S,1642709492.925387,1,1,5,0.261046,0.000000,UNIVARIA,0,0,
P,1642709493.576853,1,1,5,0.412693,0.000000,0,UNIVARIA,5494,40583168,
P,1642709493.580379,2,2,1,0.415204,0.000000,0,5499,29945856,40583168,1,1
E,1642709493.594713,1,0.418874,0.000000
```

Use the supplied **ARMPROC** macro to read raw data

Obs	Timestamp	PROC	cputime	elapsed
1	20JAN12:16:33:44.0231	DATASTEP	0:00:00.0723	0:00:00.2810
2	20JAN12:16:33:44.7513	SORT	0:00:00.1259	0:00:00.7215
3	20JAN12:16:33:44.7549	DATASTEP	0:00:00.0013	0:00:00.0034
4	20JAN12:16:33:45.4218	UNIVARIA	0:00:00.1503	0:00:00.6417

common SAS config file

```
sasautos=('<sasautos_path>' sasautos)
INITSTMT="%arm_start;run;"
TERMSTMT="%arm_stop;RUN;"
```

or

```
sasautos=('<sasautos_path>' sasautos)
smf
smfexit=smfexit
```

After gathering data…

- 25 Interactive SAS sessions
- Average session has 40 PROCs/DATA
- Think time between PROCs is 60 seconds
- No pattern: Ad hoc workload

Distribution of PROCs

proc	Frequency	Percent	Cumulative Frequency	Cumulative Percent
DATASTEP	13917	41.30	13917	41.30
SORT	8200	24.33	22117	65.63
REPORT	1026	3.04	23143	68.67
FREQ	1020	3.03	24163	71.70
MEANS	1019	3.02	25182	74.72
UNIVARIA	977	2.90	26159	77.62
SUMMARY	969	2.88	27128	80.50
CONTENTS	674	2.00	27802	82.50
CPORT	668	1.98	28470	84.48
PRINT	667	1.98	29137	86.46
GLM	663	1.97	29800	88.43
TRANSPOS	650	1.93	30450	90.36
APPEND	649	1.93	31099	92.28
COPY	647	1.92	31746	94.20
COMPARE	644	1.91	32390	96.11
CORR	643	1.91	33033	98.02
PLOT	346	1.03	33379	99.05
FORMAT	321	0.95	33700	100.00

Load generator SAS macro using SAS/CONNECT

```
%macro master(jobs=25,
              obs=.25,
              spacing=5,
              think_time=60,
              procs=40,
          wgt_data=.41,
          wgt_sort=.24,
          wgt_univ=.03,
         wgt_means=.03,
           wgt_glm=.02,
          wgt_freq=.03,
       wgt_compare=.02,
          … …
          wgt_format=.01);
```

Create 25 async SAS sessions

Read 25% of the data

Pause 5 secs between sessions

Pause 60 secs between SAS PROCS

Build job with 40 SAS PROCS

Randomly chose a PROC in this Proportion. Must add to 1.0

Master SAS job

```
%master(,,,);
%do i=1 %to &jobs;
 %sleep(5);
  signon session&i;
   rsubmit session&i;
     %build_job;
      %do j=1 %to &procs;
        …generate PROCs…
      %end;
    %update_results;
   endrsubmit;
  signoff session&i;
%end;
```

session&i

1. DATA step-*think*
2. PROC MEANS-*think*
3. PROC SORT-*think*
4. PROC GLM-*think*
5. DATA step-*think*

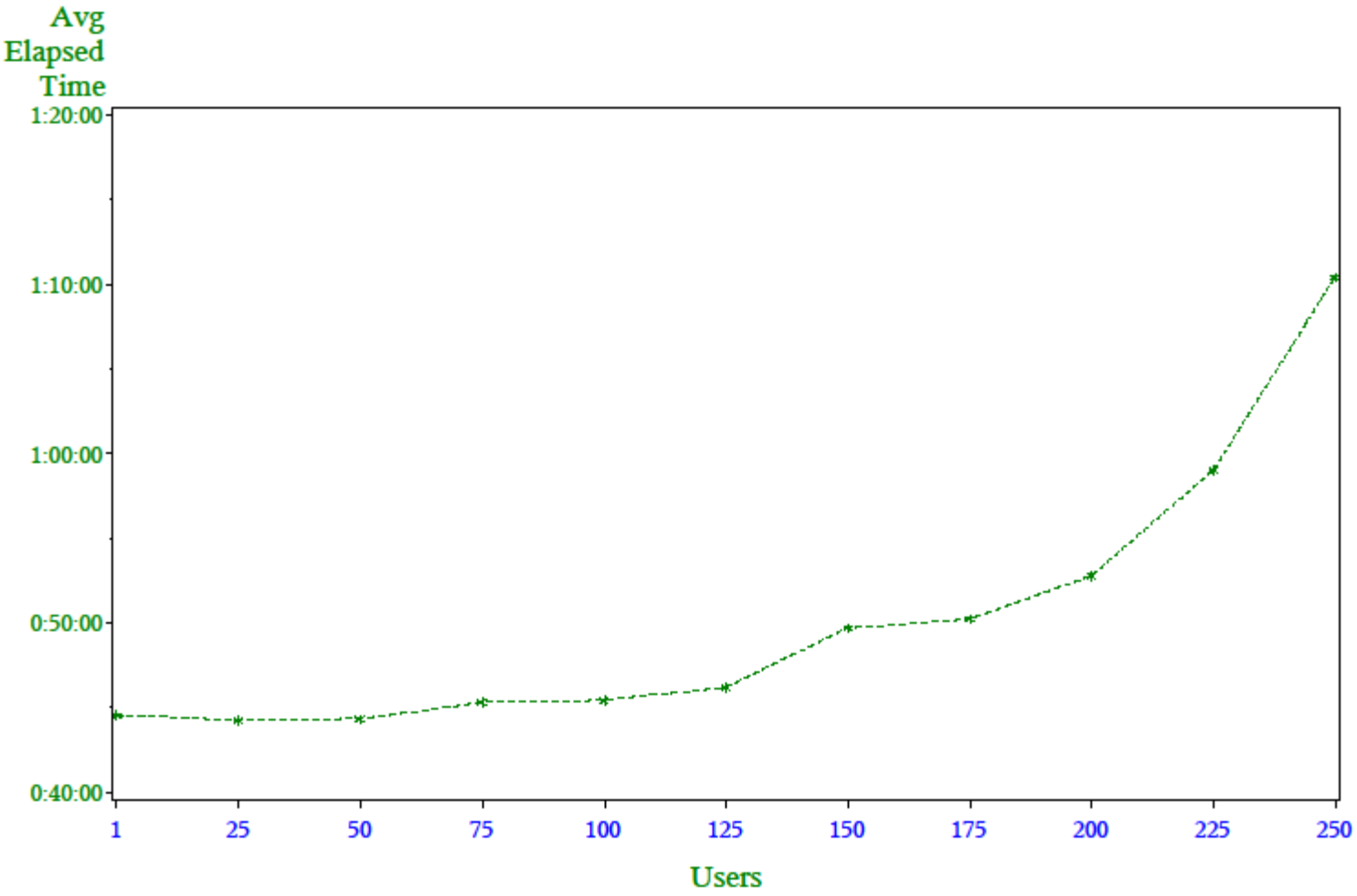
…

40. PROC SORT

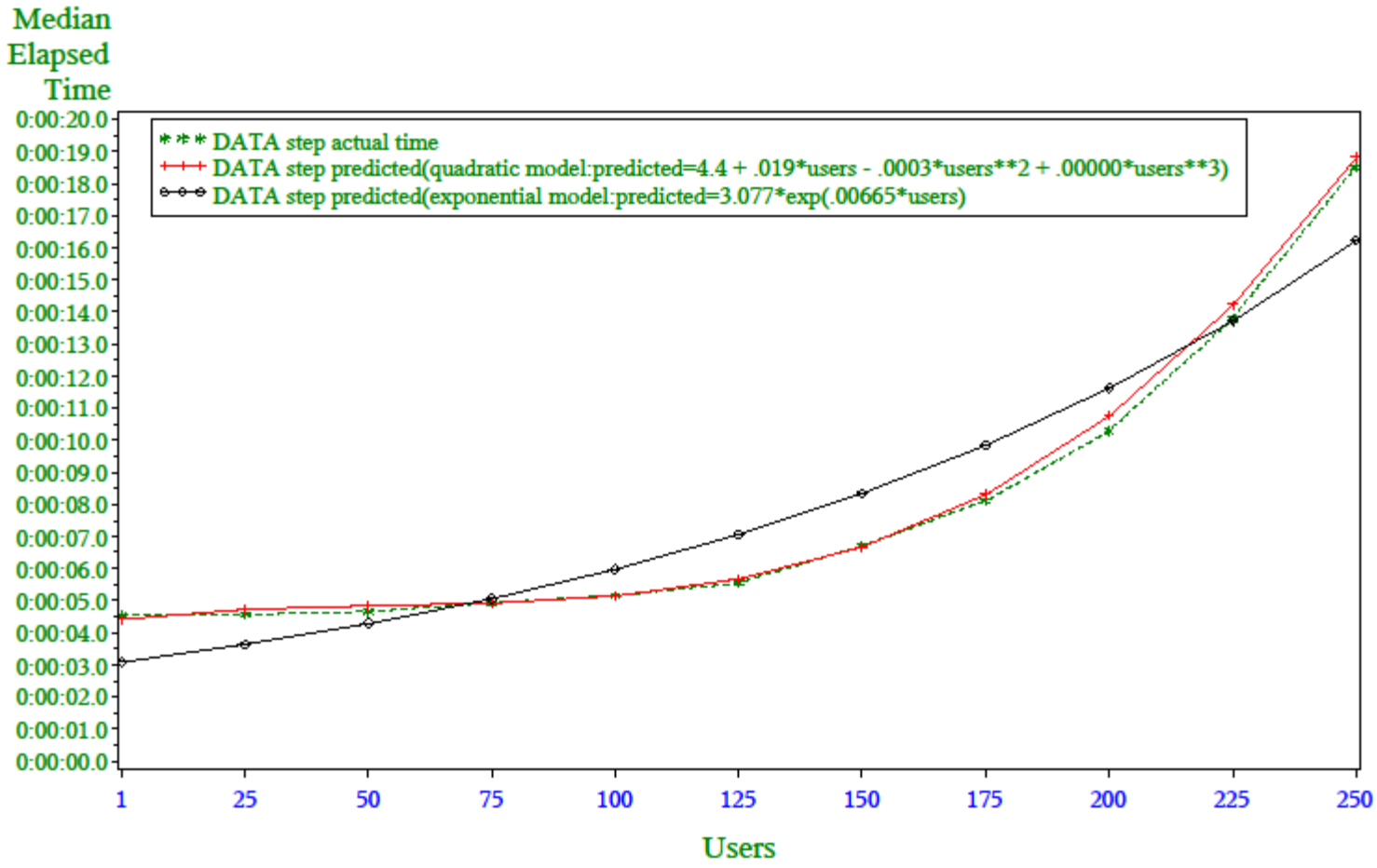
Each session Automatically Uses ARM

40 SAS PROC Session

Think Time=60



DATA Step Elapsed Time
40 SAS PROCs, Think time=60



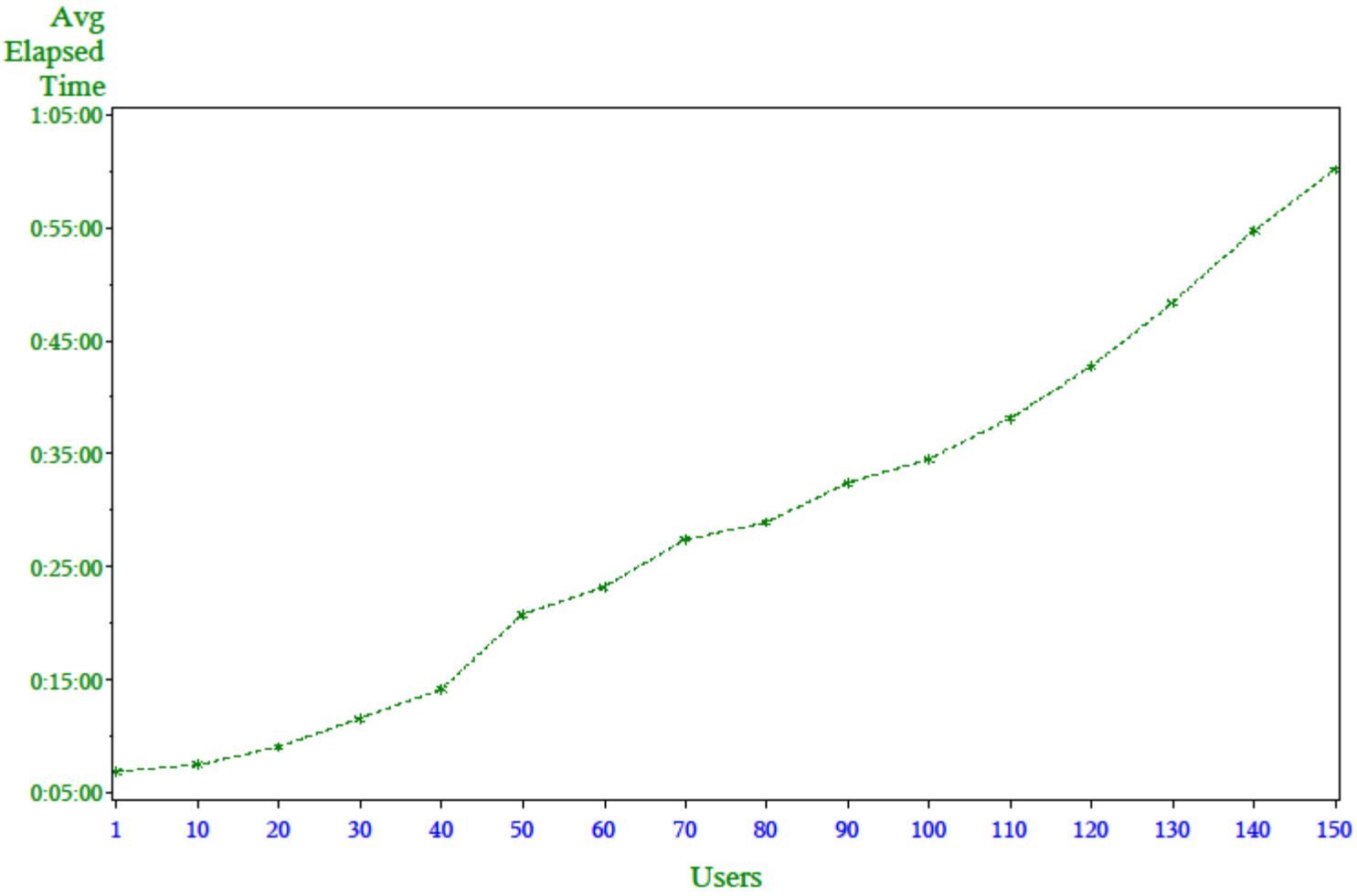
DATA step only – Duncan test

Duncan Grouping	Mean	N	jobs
A	21.0045	4251	250
B	15.8227	3784	225
C	11.0053	3392	200
D	8.5921	2896	175
E	6.9519	2588	150
E F E	5.7296	2110	125
F F	5.2388	1699	100
F F	4.9806	1290	75
F F	4.6975	864	50
F F	4.5963	407	25
F F	4.5896	21	1

No degradation in DATA step performance on the z196 up to 125 concurrent SAS sessions →

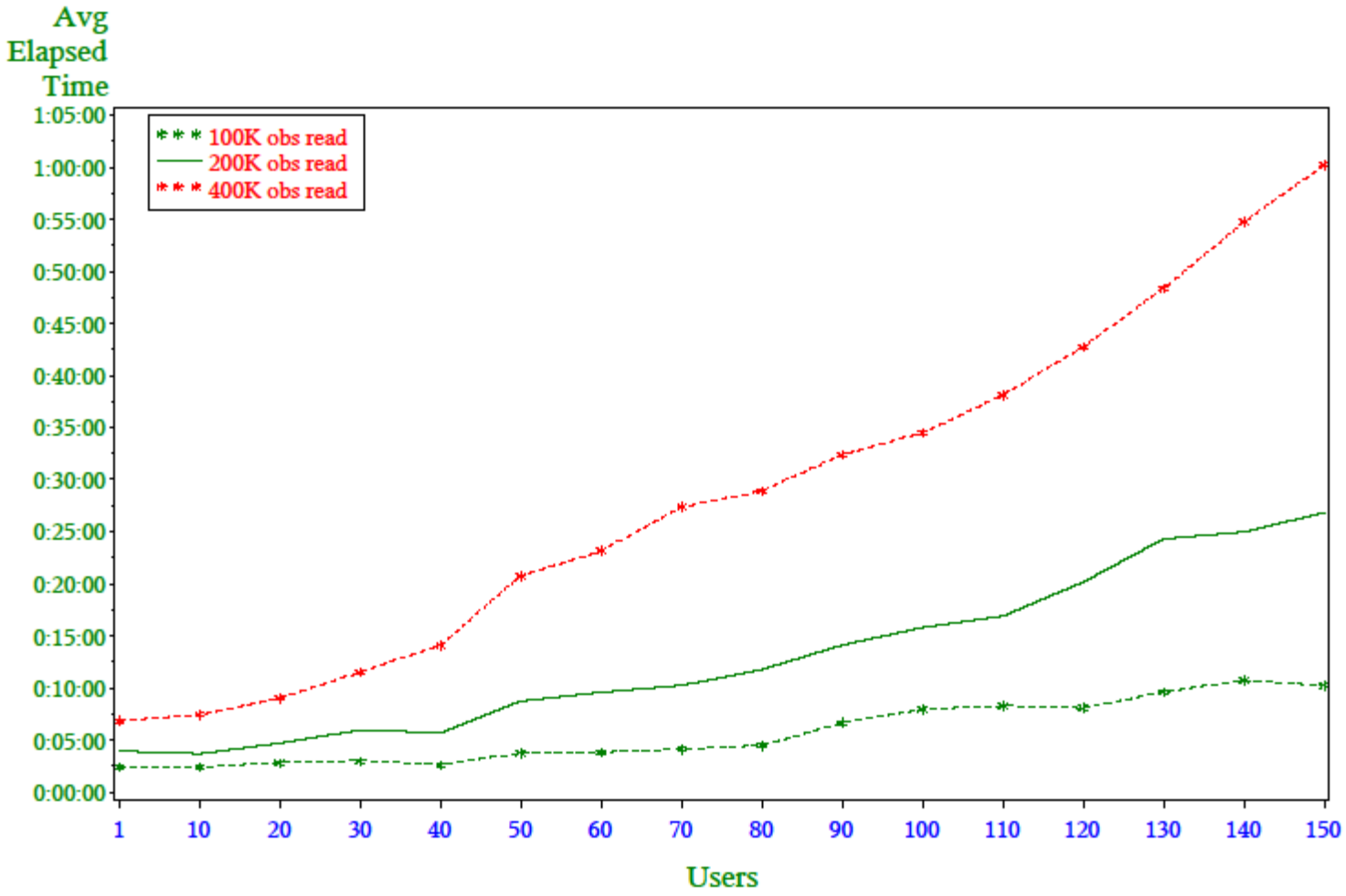
60 SAS PROC Session

Think Time=0



60 SAS PROC Session

Think Time=0



Summary

- HP Performance Center is versatile in it's ability to model BI web based apps
- DATA step elapsed times can be predicted closely with polynomial regression (slide#29)
- No DATA step degradation on the z196 for up to 125 concurrent SAS sessions (slides 29, 30)
- The z196 scales well (proportionally) across all values of concurrent SAS sessions when amount of data increases (slide#32)

- SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.
- IBM is a registered trademark of International Business Machines Corporation (IBM)
- HP is a registered trademark of Hewlett Packard Corporation (HP)
- Other brand and product names are registered trademarks or trademarks of their respective companies.

fred.forst@sas.com