"Cross-Architectural Comparison of Processor Speed Using the SAS® System"
Orion V. Hanger, O.V. Hanger Consulting Services

INTRODUCTION

The current state of computer technology is encouraging the movement to a client/server type of architecture. For those who have traditionally worked in the mainframe world, the management of this downsizing, "rightsizing" or transporting of their workload to a distributed environment can pose some challenging questions in the area of capacity planning. With Version 6 of the SAS System, the performance/capacity analyst has the potential to build SAS programs that can be run on a variety of commercial computing platforms. This paper presents one case study conducted under MVS®, VM®, OS/2® and UNIX®. The code developed and used for this benchmarking analysis is presented and some of the issues to consider when conducting this type of activity are outlined. Initial analysis results of this specific benchmarking exercise are shown.

MIPS AND BENCHMARKING

In the mainframe computer world, MIPS (millions instructions per second) are the general standard for identifying the power (speed) of the processor. Knowing the MIP rating of a processor's engine(s) is useful for expressing a computer's capacity (mips per engine x number of engines) and normalizing CPU time for chargeback (the 'meaning' of a CPU second changes when going from one processor to another with a different speed). There are a variety of professional organizations and journals that publish MIP ratings for many different types of processors. With the ever increasing complexity of computer systems in recent years (multiprocessors, partitioning, vector processors, reduced instruction sets, etc) the reliability of the industry "standard" MIPS numbers has become less dependable. This is truly noticeable when making comparisons across architectures (or even across vendors within the same architecture). In one review of industry processor ratings, it was observed that a specific 486 PC was rated at 28 mips. Knowing that an existing inhouse mainframe running VM with 400+ concurrent users generally requires about 25 mainframe MIPS, it was doubted that the same workload could be supported on the 486 PC (and still have some capacity left over).

There has also been a rise in a variety of new metrics for evaluating processing power. These include TPC-A, TPC-B, Dhrystones, SPECmarks, Khomerstone, VUPs and more. These ratings address the issue of measuring processor performance in different ways and are intended to overcome the problems with MIPS. In many ways, though, they have clouded the issue even more. Each metric tries to be defined to effectively evaluate either specific types of processors or specific types of workloads on any processor. Unfortunately, new architectures and processors continue to be made available on the marketplace and there is no source that can provide all metrics for all platforms. Basically, processor speed or performance is truly a relative thing. It is relative to your workload, the performance that you have come to expect and the importance that computer systems play in your business. It is relative to the needs of each specific computing site.

Benchmarking is an activity that is conducted to evaluate the performance potential of a workload or application in one environment as compared to another computer environment (in a geological survey, a 'benchmark' is a known elevation from which unknown elevations can be determined). In it's purist (and possibly most expensive) forms, benchmarking uses a true workload and evaluates performance for a variety of resources (CPU, memory, I/O, etc) and service objectives. But, with the relative nature of performance measures, it is possible to create simple (and less costly) SAS programs that can provide CPU time (in seconds) for making cross-architectural comparisons of processor speed. In this way, a computing site can establish their own 'relative indicator of processing power' factor.

METODOLOGY AND ASSUMPTIONS

At this case study site, processor benchmarking has been a standard procedure for evaluating CPU "speed" for every processor upgrade. The procedures and programs that evolved over the 6 years that this task had been applied for the mainframe environment proved to be effective and invaluable. The suite of programs were mostly written in SAS with a few containing code that could be run under both MVS and VM. With each processor change, the programs are run before and after the change. The sums of CPU seconds across programs were compared and a relative percent change were calculated. This percent change was applied to the unit value of the "pre" processor to get the "per engine" unit value of the "post" processor. The capacity of the processor was the unit value times the number of engines available to that specific operating system. These unit values were still referred to as MIPS (all relative to the MIP rating of the first "pre" processor ever used in benchmarking at that site).

The immediate need at this site was to make some initial capacity determinations for moving a workload from the mainframe environment to a UNIX based client/server platform. In addition, an evaluation of a couple of OS/2 platforms would be performed (mostly out of curiosity). At this site, a MIP rating of 192 (per engine) had been set for the IBM 3090-500J running both MVS/ESA and VM/HPO (via PR/SM). This value was used as the 'benchmark' value for determining the relative processing power on the new computer platforms: IBM PS/2 65SX running OS/2, COMPaq Deskpro 486/33i running OS/2 and a SUN SparcServer 670 running SunOS. Each of these platforms had a Version 6 release of the SAS System installed and available (6.07 for MVS & VM, 6.08 for the SUN and 6.06 for OS/2). It was recognized that there could be significant differences in SAS System performance across release levels and that any interpretation of results must so be tempered.

For a variety of reasons, existing benchmarking programs could not be used for this case study. Three new one-step programs were written for performing this cross-architectural comparison of processor speed. The CPU time that was used for evaluating performance was obtained from the SAS log (as reported by the data step). On PC based SAS installations (such as OS/2), the reported CPU time is really elapsed time. This means that I/O time...
is also included in this value. Because of this limitation, the benchmarks were written to do only a minimum of I/O processing. No datasets were created and the only I/O done was the initial load of the SAS System, the writing to the SAS log and a few messages to the terminal screen. All benchmarking program runs were done in the non-interactive mode of the SAS System. Both the low I/O content and the singular mode of execution provided a means for controlling problem aspects of the benchmarking activity. But they also limited the extent that interpretations could be made from the results of the benchmarking activity.

**THE PROGRAMS**

The first program in this suite of routines is almost identical to one that had been historically used at the case study site. Among all of those routines, it was also one of the most reliable (it tended to best approximate the final result obtained from all runs and programs combined). It consists of 2 nested do-loops with the outer loop sending a message to the terminal and the inner loop performing several simple arithmetic expressions. This program was run 10 times with the 'to-value' in the inner do-loop being varied to elongate the CPU time that is required to complete the step.

```sas
* BENCHMARKING ROUTINE 1;
FILENAME TERM TERMINAL;
DATA _NULL_; FILE TERM;
DO J=1 TO 10;
   PUT 'Starting outer iteration number -> ' J ;
   X=0;
   DO I=1 TO 500000 ;
      X=I+X; X=X+I; X=X-I;
      X=X*I; X=X/2;
   END;
   END
   STOP;
RUN;
```

With the second benchmarking routine, a number of SAS operations and functions are performed across two arrays of variables. The size of these arrays (number of elements/variables) was varied across ten different executions of the program by changing the value of the 'numvars' macro variable via the %LET macro statement in order to elongate CPU time required by the program.

```sas
* BENCHMARKING ROUTINE 2 ;
FILENAME TERM TERMINAL;
%LET numvars=5 ;
data _null_; file term; date=mdy(12,7,52);
array x [*] x1-x&numvars;
array c [*] $ c1-c&numvars;
DO T=1 TO 20;
   PUT 'Starting outer iteration -> ' W ;
   DO K=1 TO 30;
      V=MAX(LD,X,Y,Z);
      T=SUM(LD,X,Y,Z);
      RPT=' REPEAT(W,16);
      SUB=SUBSTR(LEFT(RPT),5,8);
      DO L=1 TO 10;
         X=RND(RND(2+RND))
      END;
      A=A-Z; B=A-Z; C=A+B;
      END;
      X=X+1; X=X-I; X=X*I; X=X/2;
   END;
RUN;
```

The final program in this suite of routines is a variation of the first program. A greater variety of functions, both character and numeric, are performed. This program was run 10 times on each test platform with the 'to-value' in the 'DO 1' loop being varied to elongate CPU time across executions of the program.

```sas
* BENCHMARKING ROUTINE NUMBER 3 ;
FILENAME TERM TERMINAL;
DATA _NULL_;
DATE=MDY(1,25,82); TIME=HMS(8,0,0);
WORD= 'ABCDEFGHIJKLMNOPQRSTUVWXYZ';
DO J=1 TO 26 ;
   W=SUBSTR(WORD,J);
   PUT 'Starting outer iteration -> ' W ;
   DO K=1 TO 30;
      X=I+1+D; Y=(X*Y)/3;
      DTM=OHMS(DATE,J);
      HOUR(TIME+D), MINUTE(TIME+D), SECONDS(TIME+D);
      DOW=PUT(DATEPART(DTM), WEEKDATE3.);
      NEW=TRANSLATE(WORD,W, '');
      LEN=LENGTH(COMPRESS(NEW));
      DO L=1 TO 10;
         IF MOD(K,2)O{) THEN Y=Y+1;
         ELSEY=Y-1;
      END;
      X=X+I; X=X-I; X=X*I; X=X/2;
   END;
RUN;
```

**BENCHMARKING RESULTS**

Each of the three benchmarking programs were run on all five computing platforms as described above and the CPU seconds reported in the SAS log was noted for each run. Three datasets were created (one for each benchmark routine) to manage and analyze the collected data. One observation in these datasets contains data from all test platforms for a specific value of 'extents' (value of the varied iteration/variable count) for that related benchmarking program. The variable names for each test platform
are o2i (IBM PS/2), o2c (COMPAQ DeskPro), sos (Sun SparcServer), vmc (VM/HPO) and mve (MVS/ESA). A print of each dataset follows:

**Benchmark Test 1**

<table>
<thead>
<tr>
<th>extents</th>
<th>o2i</th>
<th>o2c</th>
<th>sos</th>
<th>vmh</th>
<th>mve</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>6.46</td>
<td>1.96</td>
<td>0.09</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>100</td>
<td>7.16</td>
<td>2.06</td>
<td>0.10</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>500</td>
<td>9.33</td>
<td>2.27</td>
<td>0.13</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>1000</td>
<td>10.97</td>
<td>2.62</td>
<td>0.17</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td>5000</td>
<td>30.21</td>
<td>4.91</td>
<td>0.40</td>
<td>0.19</td>
<td>0.20</td>
</tr>
<tr>
<td>10000</td>
<td>52.71</td>
<td>7.82</td>
<td>0.70</td>
<td>0.34</td>
<td>0.35</td>
</tr>
<tr>
<td>50000</td>
<td>239.25</td>
<td>30.87</td>
<td>3.31</td>
<td>1.52</td>
<td>1.54</td>
</tr>
<tr>
<td>100000</td>
<td>490.43</td>
<td>59.81</td>
<td>6.24</td>
<td>3.01</td>
<td>3.05</td>
</tr>
<tr>
<td>500000</td>
<td>2283.28</td>
<td>290.85</td>
<td>30.50</td>
<td>14.76</td>
<td>14.99</td>
</tr>
<tr>
<td>1000000</td>
<td>4649.53</td>
<td>579.63</td>
<td>60.80</td>
<td>29.61</td>
<td>30.00</td>
</tr>
</tbody>
</table>

**Benchmark Test 2**

<table>
<thead>
<tr>
<th>extents</th>
<th>o2i</th>
<th>o2c</th>
<th>sos</th>
<th>vmh</th>
<th>mve</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>11.82</td>
<td>3.18</td>
<td>0.29</td>
<td>0.12</td>
<td>0.13</td>
</tr>
<tr>
<td>10</td>
<td>12.72</td>
<td>3.31</td>
<td>0.31</td>
<td>0.13</td>
<td>0.14</td>
</tr>
<tr>
<td>50</td>
<td>18.55</td>
<td>4.41</td>
<td>0.48</td>
<td>0.22</td>
<td>0.23</td>
</tr>
<tr>
<td>100</td>
<td>25.78</td>
<td>5.78</td>
<td>0.65</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>500</td>
<td>83.62</td>
<td>16.87</td>
<td>1.12</td>
<td>0.72</td>
<td>0.72</td>
</tr>
<tr>
<td>1000</td>
<td>156.13</td>
<td>30.39</td>
<td>4.02</td>
<td>2.43</td>
<td>2.36</td>
</tr>
<tr>
<td>5000</td>
<td>739.28</td>
<td>140.83</td>
<td>18.39</td>
<td>11.94</td>
<td>11.72</td>
</tr>
<tr>
<td>10000</td>
<td>1477.46</td>
<td>284.21</td>
<td>37.48</td>
<td>23.90</td>
<td>23.75</td>
</tr>
<tr>
<td>15000</td>
<td>2225.55</td>
<td>431.01</td>
<td>58.16</td>
<td>35.77</td>
<td>36.41</td>
</tr>
<tr>
<td>20000</td>
<td>3012.22</td>
<td>581.90</td>
<td>77.28</td>
<td>47.91</td>
<td>48.50</td>
</tr>
</tbody>
</table>

**Benchmark Test 3**

<table>
<thead>
<tr>
<th>extents</th>
<th>o2i</th>
<th>o2c</th>
<th>sos</th>
<th>vmh</th>
<th>mve</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>44.68</td>
<td>8.22</td>
<td>0.93</td>
<td>0.31</td>
<td>0.3</td>
</tr>
<tr>
<td>10</td>
<td>77.80</td>
<td>13.32</td>
<td>1.55</td>
<td>0.50</td>
<td>0.49</td>
</tr>
<tr>
<td>15</td>
<td>110.42</td>
<td>18.41</td>
<td>2.19</td>
<td>0.70</td>
<td>0.68</td>
</tr>
<tr>
<td>20</td>
<td>143.03</td>
<td>23.46</td>
<td>2.92</td>
<td>0.89</td>
<td>0.86</td>
</tr>
<tr>
<td>25</td>
<td>175.65</td>
<td>28.37</td>
<td>3.43</td>
<td>1.08</td>
<td>1.05</td>
</tr>
<tr>
<td>50</td>
<td>338.46</td>
<td>53.86</td>
<td>6.54</td>
<td>2.04</td>
<td>1.97</td>
</tr>
<tr>
<td>75</td>
<td>501.28</td>
<td>79.19</td>
<td>9.61</td>
<td>3.01</td>
<td>2.90</td>
</tr>
<tr>
<td>100</td>
<td>664.21</td>
<td>104.53</td>
<td>12.91</td>
<td>4.00</td>
<td>3.81</td>
</tr>
<tr>
<td>200</td>
<td>1315.10</td>
<td>205.80</td>
<td>25.51</td>
<td>7.92</td>
<td>7.51</td>
</tr>
<tr>
<td>300</td>
<td>1966.39</td>
<td>307.16</td>
<td>37.94</td>
<td>11.78</td>
<td>11.19</td>
</tr>
</tbody>
</table>

The following SAS program was used to analyze the benchmarking data. Note that the program determines relative performance attained (rpa) for the PS/2 (o2irpa), COMPAQ (o2crpa) and the SUN (sosrpa) using the average of VM and MVS as the base from which to extrapolate (both are running on the same mainframe which has already been evaluated to be 19.2 MIPS per engine).

```sas
* BENCHMARKING ANALYSIS 1;
%let base=19.2;
%macrho data(sas);
data benmark(ds); set benmarks.bentest&ds;
format o2irpa o2crpa sosrpa 6.2 ;
brot=(vmh+mve)/2; drop brot;
o2irpa=(brot*&base)/o2i;
o2crpa=(brot*&base)/o2c;
sosrpa=(brot*&base)/sos;
title1 "Benchmark Test &ds";
run;
%mend datcalc;
%macrho data(1) %macrho data(2) %macrho data(3) data summary; set benmarkl(in=b1) benmark2(in=b2) benmark3(in=b3);
if b1 then routine='Benmark1'; else if b2 then routine='Benmark2'; else if b3 then routine='Benmark3';
proc means data=summary noprint;
var o2i o2c sos vmh mve ;
output out=summary
sum=o2i o2c sos vmh mve ;
class routine;
data summary; set summary;
drop_type freq "routine;format o2irpa o2crpa sosrpa 6.2 ;
brot=(vmh+mve)/2;
o2irpa=(brot*&base)/o2i;
o2crpa=(brot*&base)/o2c;
sosrpa=(brot*&base)/sos;
if routine='Total' then routine='Total';
proc sort data=summary; by routine;
proc print data=summary noobs;
title1 'Summary of Benchmark Tests';
run;

The summary data (edited from the final PROC PRINT in the above program) is as follows:

**Summary of All Benchmark Tests**

<table>
<thead>
<tr>
<th>Routine</th>
<th>o2irpa</th>
<th>o2crpa</th>
<th>sosrpa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benmark1</td>
<td>0.12</td>
<td>0.98</td>
<td>9.37</td>
</tr>
<tr>
<td>Benmark2</td>
<td>0.31</td>
<td>1.59</td>
<td>11.99</td>
</tr>
<tr>
<td>Benmark3</td>
<td>0.11</td>
<td>0.72</td>
<td>5.84</td>
</tr>
<tr>
<td>Total:</td>
<td>0.19</td>
<td>1.19</td>
<td>9.76</td>
</tr>
</tbody>
</table>

**CONCLUSION**

The transportability of programs written for Version 6 of the SAS System, provides a useful opportunity for analysts to evaluate capacity issues for moving workloads to a client/server type of environment. This methodology, assumptions and code for this benchmarking activity was reviewed. Though it is recognize that the results have limitations in their interpretation, the results that are obtained provide a good first step towards understanding a new computer systems environment and capacity.

SAS is a registered trademark of SAS Institute Inc., Cary, NC, USA. MVS/ESA, VM/HPO and OS/2 are registered trademarks of IBM Corp.

UNIX is a registered trademark of AT&T.