Abstract

How fast does PC/SAS run on one computer compared to another? One general comparison is the time required to run the TESTBASE installation program. This paper compares times for a variety of PCs, mainframes, and minicomputers.

PC/SAS is slow?!

My first experience with PC/SAS was on a PC/XT with a hard disk that was slightly faster than IBM's. TESTBASE, the installation test job, took 22 minutes. Worse, I watched the lines of SAS code scroll down the LOG window with agonizing slowness, realizing that my jobs would scroll just as slowly. And it gets worse: when I got my own PC, Testbase took 27 minutes to run! Clearly, all PCs are not created equal when it comes to running SAS (and SAS Institute has clearly stated that it was designed for the AT even though it runs on the PC).

What performance can be expected?

Complaining about how slowly SAS runs is not new. Users have been complaining for years that mainframe SAS is too slow. The cry has been, "It's a resource hog! We can't run it because the jobs take too long." Of course, computers get faster, and SAS Institute regularly (claims to) prove(s) SAS's performance. But, for most of us, SAS (like the other programs we run) doesn't run fast enough. So, we look for the utopian computer where SAS jobs run instantly (or at least under .5 seconds).

My curiosity aroused, I began investigating how fast SAS would run on various PC's. Academic Computing Services began to wonder which of several operating environments (at Syracuse, SAS runs under CMS, MVS, VMS, and PC/DOS) is best for SAS. Also, users began to ask, "Is going to be using PC/SAS. What PC should I buy?" There was a lot of impetus to generate performance data among machines (PC and non-PC) to answer these questions.

Data Collection

I run TESTBASE on every computer (PC and mainframe) I had access to, and started a table of these run times. Local users and members of the BITNET SAS electronic user's group contributed data. These sources provided a set of timings (benchmarks) on more than 20 PC's and several mainframes.

A Benchmarking Primer

Benchmarking is the art (science?) of comparing the relative performance of computers to each other. That task is not simple. For example, there are computers that only do arithmetic operations. Though these computers are usually very fast, they won't be very useful for word processing. Benchmarks can make any given computer look very good (or very bad), depending on which aspects of performance are measured. This says that a USEFUL benchmark should either be a typical program, or should test those features which our typical program will use most.

SAS speed mainly depends on two aspects of a computer's performance: its CPU (brain) speed and disk speed. This makes sense because SAS can be used to do a lot of calculations (say with a statistical PROC); it may read and write a lot of data; or it may do both. Suppose a given SAS program mostly does calculations and doesn't handle a lot of data. For that program, a fast CPU is important; a fast hard disk won't improve performance very much. Conversely, a very fast disk is critical to a program that reads and writes a lot of data.

Vendors provide numbers which give some useful information on these two performance measures:

-processor speed (in megahertz). A larger number is better.
-disk access time (in milliseconds). A smaller number is better.

One can generally use these numbers to predict PC/SAS's performance, but numbers do sometimes lie. For example, a "turbo" PC had one of the slowest timings. This shows why running a "real" program is key to benchmarking.

Some Principle of Benchmarking

From the previous primer, we can derive a definition and several principles (these are informal and not exhaustive, formulated for the needs of this paper).

Performance: an overall measure of how rapidly a computer does a task. In this paper, it is usually the elapsed (clock) time for a given computer to execute TESTBASE.

Principle #1: Benchmarks are useful only if they consist of tasks similar or identical to what you do.

Principle #2: Benchmarks should be repeatable. I.e., the same program run on the same (or similarly equipped) computer under the same circumstances ought to take the same amount of time (but see #3 below).

Principle #3: Many factors influence performance, and they are often difficult to identify or quantify (e.g., not all seemingly identical IBM AT's run SAS at the same speed.)

Principle #4: Benchmarking takes time!

The Benchmark

As already noted, the TESTBASE program shipped with PC/SAS is being used as the benchmark. SAS Institute does NOT ship TESTBASE for this purpose- it is designed
to check if SAS is running correctly. However, it exercises many SAS features and PROCs; it does a lot of I/O (the hard disk light stays on a lot); and the CPU gets a workout (the hard disk light isn’t always on). Furthermore, it’s standardized—there’s no question as to what it is. These characteristics fit nicely (if non-rigorously) to the above principles. Further, experience running SAS on various machines says the numbers make sense. So it seems safe to use TESTBASE while understanding its limitations.

Factors that Influence SAS Performance

Besides CPU speed and disk access time, other factors affect PC/SAS performance. The number of buffers, a disk cache program, a ramdisk, optimizing a disk, a faster disk interface, a numeric coprocessor, and other factors can make a significant difference. These areas are outside the scope of this paper, but many were covered in a paper by David and Sweetland at SUGI 12 (ref. 1). One comment: the best combination of economy, ease of use, and effectiveness is a disk cache program (may require additional memory).

Test Procedures

In order to benchmark a PC’s “raw” performance, most were tested once under standard conditions. The number of buffers was set at 20 (although I now usually recommend 40); no disk caching programs or ramdisks were used; and a co-processor was installed (although it appears to make little difference for this benchmark). The disk installed was usually appropriate to the computer, e.g., an AT had an AT class disk installed. Timing started when F10 (Submit) was pressed; it ended at the appearance of the final (red) screen. I tried (and suggested to other testers) to press the RETURN key immediately after pressing F10 in order to minimize delays from the starting (red) screen. In addition, timings were run with different factors (e.g., adding a RAMDISK), depending on the individual PC, skill of the tester, and time available.

Mainframe timing raised two compatibility issues. The starting and ending screens were removed from TESTBASE because they require the DISPLAY statement (not available in version 5). Time required to display these screens is a small percentage of the total time, and should not materially affect comparability. Of more concern is the time-sharing nature of mainframes. Run times can vary significantly depending on mainframe loading. Testers tried to do runs during “typical load” periods, but who can say what a “typical” load is? Times would usually be shorter if run during slack periods (around 3 a.m.), or longer during university finals week! CPU time is recorded, as it gives some raw indication of relative CPU speed.

Appendix 1 contains the text describing the test procedures that went to the BITNET SAS electronic user’s group.

And now the Numbers...

Following is an abbreviated table (Appendix 2 is the entire table).

<table>
<thead>
<tr>
<th>Machine</th>
<th>Speed (mhz)</th>
<th>Time</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compaq 386</td>
<td>16</td>
<td>4:40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>2:35</td>
<td>Ramdisk</td>
</tr>
<tr>
<td>IBM PC/AT</td>
<td>8</td>
<td>9:54</td>
<td></td>
</tr>
<tr>
<td>IBM PC/XT</td>
<td>4.77</td>
<td>22:30</td>
<td></td>
</tr>
<tr>
<td>IBM PS/2-30</td>
<td>10</td>
<td>13:25</td>
<td>On Loaded network</td>
</tr>
<tr>
<td>IBM PS/2-60</td>
<td>10</td>
<td>9:32</td>
<td></td>
</tr>
<tr>
<td>PC Limited 286</td>
<td>12</td>
<td>6:00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3:14</td>
<td></td>
<td>2 meg. ramdisk</td>
</tr>
<tr>
<td></td>
<td>3:45</td>
<td></td>
<td>384k cache</td>
</tr>
<tr>
<td>DEC Vax 750</td>
<td>2:08</td>
<td></td>
<td>Unloaded</td>
</tr>
<tr>
<td>IBM 3090/150(CMS)</td>
<td>:45-55</td>
<td></td>
<td>Load 65%-daytime</td>
</tr>
<tr>
<td>IBM 3341-12(MVS)</td>
<td>3:30</td>
<td></td>
<td>Daytime</td>
</tr>
</tbody>
</table>

Interpreting the Numbers

First, differences between PC’s of 10 or even 15% may not be significant. These benchmarks aren’t that precise.

Second, individual use may be more (or less) CPU (or disk) dependent than TESTBASE’s. For example, a given SAS program may do little I/O and thus be largely dependent on CPU speed (although all PC/SAS programs will be dependent to some extent on both CPU and disk speed).

Third, other factors than simple CPU or hard disk speed affect SAS performance (this has already been covered).

Fourth, the above benchmarks for PC’s were all run under 6.02. Although there has not been much time to test 6.03, first tests indicate it is slower (run on PC’s Limited 286-12). Most of the additional time seems to be from much more I/O.

Fifth, many mainframes can still run SAS much faster than the fastest PC’s (depending on loading).

Sixth, a fast 286 PC can provide up to 75% of the performance of a 386 (16 mhz) PC (at 20-30% of the cost, given current PC street prices).

Conclusions and Disclaimers

These benchmarks can help toward four goals:

1. Informed PC purchases.
3. Choosing the appropriate (PC or mainframe) environment.

4. Adding to our knowledge store (corny, but true).

While it is inadvisable to make a purchase solely based on these numbers (these numbers weren’t rigorously generated and are not guaranteed), they can be a useful part of a PC selection/upgrade plan. Also, the article “What Good is the SAS System for Personal Computers?” by J. Philip Miller and Robert P. Parks in the SUGI Conference Proceedings is required reading when evaluating PC/SAS costs and performance figures.

Credits

Many people around the world on the BITNET SAS electronic user’s group contributed benchmarks that are included in this paper. Interacting with these people has been a consistent pleasure. They weren’t asked to be rigorous, yet many ran several trials to substantiate their results. I cannot thank them enough for their willingness to contribute.

Dave Brummit of SAS Institute provided valuable (and timely) assistance on many occasions. His help is gratefully acknowledged.

Dave Boeshaar provided much early information and first supported PC/SAS at Syracuse University. Syracuse University Academic Computing Services, and especially my supervisor, Sally Webster, provided the time, resources, and incentive to write this paper. SU Computing and Network Services, especially Don Hanley, provided additional help.

Trademarks and Additional Disclaimers

All SAS products are trademarks (tm) SAS Institute, IBM PC, IBM PC/XT, IBM PC/AT trademark (tm) IBM Corporation. Other trademarks apply. Neither SAS Institute nor any of the hardware or software manufacturers provided financial support for this project. Syracuse University, Academic Computing Services, and my supervisor generously supported this project, but do not endorse any products.

Appendix 1

Following is the text regarding the test procedures sent to the BITNET SAS electronic user’s group. The grammar and construction reflects its history— it was generated in the heat of day-to-day work and not written for publication.

I’ve been running the TESTBASE SAS program on a variety of machines, including a couple of mainframes, in order to get a feel on “Is PC/SAS fast or slow (relative to various PC’s)?” I’d appreciate any comments, timings on other equipment that you might have, which I will include in my data.

Zeroth, these are informal numbers collected by myself, colleagues, and SAS-Uers (they’re colleagues, too). SAS Institute has no involvement with this effort.

First, these results can be generalized to an application only if the task mix is similar. TESTBASE puts a premium on a fast disk and fast I/O architecture (this probably makes the PCs Limited 286-12 look better than its clock speed, and may make the new PS2’s nicer SAS machines than old PCs). This is because the TESTBASE job has a lot of short data steps and procedures causing SAS to constantly shuffle its modules from disk to memory. If an application is mostly compute bound, my experience says that execution time is linearly related to clock speed.

Second, I run with BUFFERS=20 on all PC runs unless otherwise noted.

Third, some of these have 80(87) co-processors installed, but I don’t think it makes that much difference for TESTBASE. + if installed, - if not, ? if unknown.

Fourth, this is run from the display manager (except for mainframe versions) because of DISPLAY statements in two data steps.

------- Benchmark table was here -------

Test procedures start/stop points:

PC/SAS- the TESTBASE job must be run from the display manager.
1. INCLUDE the TESTBASE job into the Program Editor window.
2. Press F10 to start. **Timing starts now**.
3. A start window (red on a color monitor) will appear in 5-30 seconds— asking you to press ENTER to start the test. You can bypass this window by pressing ENTER immediately after pressing F10. If you allow the start window to sit there, it obviously adds to the timing length.
4. An ending window will appear after SAS is done (there will be about 465 statements on the log at this time). You will be asked to press ENTER again, which exits immediately to the display manager. **Timing stops**.

Since this is all a rough evaluation, if your timing is off a few seconds it doesn’t really matter. Even if you time from the start window, instead of F10, the timing won’t be affected drastically since it’s 5% or less of the total time (and the accuracy of this test probably isn’t that good anyway).

Mainframe users: since you can’t have the start and stop windows, that time before the start window just won’t be there. If I’d started these timings with mainframes in
mind, I'd probably not have included it in the PC timings.
However, I'm not going to go back, and as explained above, it really doesn't make much difference.

-CMS use Q TIME before and after to get times:
   CMS Q TIME#SAS TESTBASE#CMS Q TIME

-MVS batch: elapsed time, use XXXXXXX MINUTES

EXECUTION TIME in the
JESn JOB STATISTICS
CPU time from IEF376I message after the
dataset allocation messages

other operating systems: I'll include info from people if
you tell me what you use. I know from June there are
some problems she's looked at in giving a TSO time.

Appendix 2 - Table of all TESTBASE Timings

<table>
<thead>
<tr>
<th>Machine</th>
<th>DISK Clock</th>
<th>Time</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compaq 386</td>
<td>7f 716</td>
<td>4:40</td>
<td>&lt;SL&gt;</td>
</tr>
<tr>
<td>Compaq 386</td>
<td>25ms +16</td>
<td>4:45</td>
<td>&lt;JG&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2:35</td>
<td>All to RAMdisk (6.02 instr.)</td>
</tr>
<tr>
<td>Compaq PortIII</td>
<td>28ms -12</td>
<td>6:51</td>
<td>&lt;Rocks&gt;</td>
</tr>
<tr>
<td>Compaq Desk/286</td>
<td>28ms +12</td>
<td>5:44</td>
<td>&lt;Rocks&gt;</td>
</tr>
<tr>
<td>Compaq 386/20</td>
<td>28ms 20</td>
<td>4:13</td>
<td>&lt;Rocks&gt;</td>
</tr>
<tr>
<td>Epson EquityII</td>
<td>65ms - 9.54</td>
<td>13:30</td>
<td>Buf=4, 17+min.</td>
</tr>
<tr>
<td>Epson EquityIII</td>
<td>AN1 712</td>
<td>10:35</td>
<td>Buf=18</td>
</tr>
<tr>
<td></td>
<td>ANu 5:40</td>
<td>Buf=20</td>
<td></td>
</tr>
<tr>
<td>IBM PC/AT</td>
<td>39ms + 8</td>
<td>9:54</td>
<td>&lt;JPM&gt;</td>
</tr>
<tr>
<td>IBM PC/AT</td>
<td>39ms + 8</td>
<td>8:31</td>
<td>SASWORK in RAMDISK.</td>
</tr>
<tr>
<td>IBM PC/AT</td>
<td>? ? 8?</td>
<td>3:40</td>
<td>All SAS prog files to ramdisk</td>
</tr>
<tr>
<td>IBM PC/AT</td>
<td>39ms + 6</td>
<td>8:45</td>
<td>Batch. &lt;SM&gt;</td>
</tr>
<tr>
<td>IBM PC/AT</td>
<td>85ms + 4.77</td>
<td>22:30</td>
<td>Add SASWORK on VDISK</td>
</tr>
<tr>
<td>IBM PC/AT</td>
<td>AN1 4.77</td>
<td>26:30</td>
<td></td>
</tr>
<tr>
<td>IBM PS/2-30</td>
<td>80ms - 8</td>
<td>13:25</td>
<td>WaveMate 286 motherboard</td>
</tr>
<tr>
<td>IBM PS/2-30</td>
<td>80ms -6</td>
<td>12:32</td>
<td>Buf=40 on PS/2-30.</td>
</tr>
<tr>
<td>IBM PS/2-50</td>
<td>80ms -10</td>
<td>9:32</td>
<td>&lt;JPM&gt;</td>
</tr>
<tr>
<td>IBM PS/2-50</td>
<td>80ms +10</td>
<td>9:15</td>
<td>SASWork in RAMDISK.</td>
</tr>
<tr>
<td>IBM PS/2-60</td>
<td>30ms +10</td>
<td>5:05</td>
<td>IBMcache(384k)</td>
</tr>
<tr>
<td>IBM PS/2-80</td>
<td>40ms 716</td>
<td>3:36</td>
<td>Cache(384k); cache(64k)=4:30</td>
</tr>
<tr>
<td>IBM PS/2-80</td>
<td>15ms -20</td>
<td>5:00</td>
<td>BSDI disk &lt;Rocks&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2:13</td>
<td>Cache(1024k)</td>
</tr>
</tbody>
</table>
PC Lim 286-12 28ms +12 6:00  Buf=50, 5:50; buf=90, 5:30
+ 6 9:00
+12 5:30 SASWORK in RAMdisk
+12 7:30 384k Ramdisk w/SASXSUPV
  Longer path seems to negate RAMDISK
+12 5:00 Batch
+12 4:30 Batch. Buf=40.
+ 6 6:21 Ramdisk w/sasxsupv,sasmds,saswork
+ 6 5:53  Almost All in ramdisk.
+12 3:14 Ramdisk w/sasxsupv,sasmds, saswork
+ 6 6:37 GoldBowcache(384K); 64k-7:20;
+12 2M=6:16
+ 6 3:45 GoldBowcache(384k)
PC Lim Turbo 65ms - 8 15:06 <Rocks>
NCR PC710 28ms + 10 7:30 Buf=40, 7:14; buf=4, 9:20.
PC Source/ CompuAdd 28ms + 8 6:42 No wait states. Buf=8. <RMD>
  +12 5:14 Ramdisk w/Saswork. Buf=8
  + 6 3:40 Ramdisk w/SASwork, dummy, optio, xsupv, glm, util, xsufs, znote, zoptn, msg, fardump. Buf=8.
Tandy 1000TX 85ms ? 8 19:00 Disk is unknown hardcard.
Tandy 4000 65ms ? 7:00
  +12 5:40 IBMcache(64k) in real memory.
Telex 1240 85ms ? 8 8:40
  + 6 6:30
Zenith Z158 85ms? - 4.77 5:00
  + 8 5:00 GoldBowcache(384k)
Zenith Z241-82 56ms + 6 26:00 For Buf=4, 27:00.
 Zenith Z386 28ms ? 5:09
  +12 3:45
--------Mainframes/Minis--------
DEC VAX 750 w/Flt. pt. accel 2:08  CPU=1:28 Float pt accel, 1MB
DEC VAX 8800 12:45  CPU 11s. 1MB wrking set. <BB>
DEC VAX 8650 2:31  CPU 19n. Fairly loaded. <BB>
IBM 3081-KX (VM/HPO 4.2) 4:42  CPU 6.55. Load 12% (low).<JP>
IBM 3090/150 (VM/HPO CMS) 4:50  CPU=3.7sec. Run middle of day,
  IBM 3090/300 (MVS) 2:13  CPU use=65%. paging light
IBM 9275-50 1:35
IBM 4331-11 (VM/CMS rel 5) 3:00  Virtual CPU=74sec. Moderate-
IBM 4341-12 (MVS batch) 3:30  heavy load. <TMK>
IBM 4361-LK5 (VM/SP 4.) 3:0 sec. SAS v 82.4.
IBM 9160 (VM/CMS) 30-120s <RVP>
NAS 9160 (VM/CMS) 2:21
NAS 9160 (MVS) 2:35sec. Moderate load. <JR>
---------------------Machine-----------------
One entry is provided for each machine tested. Multiple times
for a single entry indicate multiple TESTBASE runs under different
conditions. Multiple listings for same model means more than one
sample of that model was tested.

DISK:
average access time is in milliseconds (ms)
AN=Arcnet network hardware, Novell software.
ANL=network is unloaded
? =unknown disk. ?f=factory installed unknown disk.
Clock:
CPU clock speed in megahertz (mhz).
+ before speed means coprocessor installed
- means coprocessor not installed
? means unknown if coprocessor installed
coprocessor does not seem to affect TESTBASE much

Time:
ELAPSED TIME FOR THE TESTBASE RUN IN MINUTES:SECONDS

646
Comments:

<> Initials of person who provided information
Buf=n  number of buffers installed in CONFIG.SYS.
   If not listed, buf=20.
IBMcache(nnnk)  IBM disk caching program shipped with PS/2-50.
   nnnk is size of cache.
GoldBowCache(nnnk)= Golden Bow Cache program

Disclaimer: these timings, while carefully collected, are not guaranteed
to be accurate. They may be safely used as one source of information for
running SAS/PC. They do not constitute an evaluation or endorsement of
any product.