Creating a PC-based Information System: An Evolutionary/Revolutionary Visionary Approach

Peter Parker, U.S. Dept. of Commerce, Washington D.C.
R. Keith Daly, U.S. Dept. of Commerce, Washington D.C.

ABSTRACT

This paper is intended for mainframe users who are considering changing system platforms. This paper describes why we decided to go with the PC, the considerations made, the problems encountered and, our assessment of the progress made. Finally, it touches on future enhancements that we anticipate.

The Office of Textiles and Apparel(OTEXA) of the U.S. Department of Commerce made a decision to move all databases and all computer programs from the National Institute of Health (NIH) mainframe computer time-share utility onto a PC based system. We have used mainframes for data processing since 1978. Until a few years ago, PCs couldn't store all of our data or process it quickly enough. This platform conversion involved moving billions of bytes of data and taking over functions performed transparently at the NIH Computer Center: data security, data backups, and system maintenance. The advantages of such an approach are numerous: major cost savings, software and hardware flexibility, more control over the system, and user-friendly applications.

We use the OS/2® version 6.10 of SAS® software, as well as most conventional spreadsheet and database PC software packages on a dual Pentium PC with multi-gigabyte hard drives and multiple CD-ROM units.

INTRODUCTION

The U. S. Textile Program of the Dept. of Commerce is concerned with the health of the domestic textile and apparel industry. The program monitors and establishes import agreements, assesses the domestic industry, and promotes market expansion. The evaluation of the U.S. textile and apparel markets and imports from all countries, category by category, allows for the identification of potential market disruption. Analyses of the domestic industry may determine that market disruption or the threat of disruption exists. Such a conclusion may lead to the establishment of import quotas for specific commodities and countries.

The Textile Information Management System (TIMS) began in 1978 by Congressional mandate. TIMS produces all of the Textile Program's periodic reports on time, and to have an easily accessible collection of databases for time-critical, situation-specific query, retrieval, and analysis. See Appendix III for more details on some of these reports. About fifty on-site Commerce international trade professionals and economists receive these reports on paper and microfiche. Most of these reports will become available on their PCs for viewing, via a recently implemented LAN system and object oriented software. Currently, this information is being processed on a combination of mainframes at the National Institutes of Health (NIH) and microcomputers in the Office of Textiles and Apparel (OTEXA). TIMS is implemented by two computer programmer analysts(the authors) and a computer assistant. We operate out of a studio/laboratory/workshop environment, dabbling with and researching various hardware and software technologies, writing code, pulling chips out of mother boards, building PCs, searching computer science information bases, and envisioning the future of information technology. SAS software is the major computer software used to create and maintain the import and exports databases and to produce the TIMS reports and data tables. Most of these reports, some of them numbering hundreds of pages, are generated monthly and a few weekly. Ad hoc reports ranging from a few pages to a few hundred are generated on a daily basis for OTEXA professional staff.

Q1. WHY DID WE CHOOSE THE PC PLATFORM? (EVOLUTION)

Our decision to migrate applications off a mainframe onto a PC environment was not the result of cost/benefit analysis but rather an evolutionary process. As CPUs increased in processing power and speed, we started to move applications over. After successfully attempting our most I/O and CPU intensive applications, we discovered that all of our mainframe applications would work on PCs for only a small fraction of our mainframe costs. Decisions by Commerce Officials to move our office off the NIH Computer Center by the Fall of 1996, in order to use the excess capacity of the Commerce mainframes, served to hasten our application migration. In order for this major platform change to be successful, **the following criteria must be met:**

1. **THE PLATFORM MUST HAVE ENOUGH CPU POWER** to process the large amount of data in a timely manner. The largest applications must be able to run in a few hours. While processing speed is essential, the fastest machine is not a sufficient criterion.

2. **THE PLATFORM MUST HAVE ENOUGH STORAGE SPACE** to hold the large size of the databases and the equally large sort databases used during database maintenance and processing.

3. **THE PLATFORM MUST BE ABLE TO RUN SAS SOFTWARE**(see question #2 below on why we chose SAS software)

4. The **PROCUREMENT** of these resources must be **WITHIN OUR BUDGET.**

The PC platform met all four of these criteria and clearly edged out the other platforms(other mainframes, minis, Sun workstations, etc.) for several reasons:

1. **PC EXPERTISE**-We already had expertise with PCs (over ten years). We've been working with PCs since their introduction in the early 80's and already had one PC per user in our office(about fifty), primarily for word processing, spreadsheets, and graphics. We have had experience in using most significant software and we have always kept up with many of the latest developments in hardware like RAID drive systems and CD/R write-once technology.

2. **PC POWER**-PCs have more than enough CPU power, storage and memory capacity to handle our data processing needs. We were able to run our most CPU/memory/storage-space intensive application on a 486/33 PC with 32 MB of random access memory(RAM) and one gb of storage. Our equipment has advanced greatly since this early test. Although the time required to handle data production was long, it was not prohibitively

so. Now, we can run and finish production on a PC during the workday before we start production at the NIH Computer Center, during its evening discount period.

3. **PC POTENTIAL**- We anticipate that the PC will continue to evolve substantially and quickly. Every six months bring significant improvements in CPU processing power and speed as computer chips are packed with more transistors (the new P6 chip has 5.5 million) and are boosted by faster clock speeds(200 MHz and increasing).

4. **PC COST**-PCs are cheap and easy to procure(they are much less expensive than any other alternative). They cost only a few thousand dollars, a small fraction of our mainframe expenses. When a substantial new technology comes out, we can either upgrade our PCs, dispose of them, or pass them down to our users. As our users and their software become more sophisticated, their hardware needs increase. As opposed to purchasing large computer systems, the Federal Government PC procurement process is relatively simple. Procurement rules become more complicated and much more difficult as more money is involved.

5. **PCs complement** other systems(especially LANs, Internet access and other major applications). Our office of fifty analysts and economists have now integrated with other offices in Commerce in a Local Area Network(LAN). We plan to use the LAN to enable users to electronically mail data requests to our SAS software programmers, and view SAS software generated reports on their PC monitors instead of reading hard copy.

Q2. WHY ARE WE USING SAS SOFTWARE ON OUR PC -AND WHY THE OS/2 VERSION? (EVOLUTION)

To meet our needs, a programming language must enable us to:

1. process large volumes of ASCII data(tapes from the U.S. Census Bureau and the U.S. Customs Service),

2. produce our monthly and weekly reports, and

3. quickly process ad-hoc data requests.

We can accomplish all this with SAS software.

WHY SAS SOFTWARE (as the programming language)?

1. **Many Years of Experience** in SAS Software Programming.

We have been using SAS software since the 70's when it was available only on IBM® mainframes. We have used it for statistical analysis, econometric analysis and data processing in a variety of data processing environments, from economic research for Economics professors to import analysis for government agencies. Together, the two of us have over 30 years of experience with SAS software.

2. **Easier to Use** Than Other Computer Languages.

Although SAS software was originally developed as a statistical package, we also have found it useful as a powerful programming language for maintaining large databases and producing various summarized reports. Our major concern in choosing programming languages is the ability to quickly grind out ad-hoc applications. Lower level languages like FORTRAN and COBOL are unwieldy to use, requiring much time for coding and troubleshooting. If one understands the basic concepts of computer logic and structured programming, he can learn SAS software quickly, almost like a macro-based version of these other languages. For example, instead of coding a bubble-sort with many lines of COBOL, one can type "PROC SORT" with the variables to sort by, using only two lines of code versus many. One beauty of SAS is that it can be useful on two levels. Simple but powerful data requests can be written by a non-programmer, such as an economist or statistician, with only a few lines of code, while on the other hand, SAS software is low level enough to allow an experienced programmer to devise more complex tasks.

3. Easy Transition From Mainframe Version.

We have found SAS Software code that we had written for the IBM mainframe would execute on the PC (and vice-versa) with few changes. References to external files had to be adjusted for the environment. For example, IBM job control language (JCL) from the mainframe must be replaced by relatively simple "Filename" and "Libname" statements for the PC. Other adjustments to our code were minimal and often, such as problems with length statements, were corrections to poor coding rather than mainframe to PC adjustments. We did, however, discover a need to be more efficient.

On the mainframe, we operated in a batch environment, and we submitted our major jobs to run overnight during the discount period. This 60 percent cost savings was the major reason for overnight job submission, but it also freed us from concern about speed. Running in real-time on the PC, speed becomes important. Temporary storage space, another problem of little concern on the mainframe, must also be considered. Unless one is careful to delete unneeded datasets and unneeded variables, "disk full" errors can be a problem even when seemingly enormous hard drive space is available. We appreciate the SAS Software interactive environment, including the ability to rerun parts of a program, using work databases created from previous job submissions, and extensive use of macros to feed parameters.

The choice to use the OS/2 version of SAS software was also straightforward. At the SUIGI in Hawaii a few years ago, we brought applications to the SAS Institute Demonstration Center and ran benchmarks against the different versions of SAS software. The OS/2 version ran faster than the Windows one for our types of applications, and after conferring with SAS Institute staff, we decided that the OS/2 route was the preferred one for intensive data processing. With OS/2's other advantages of multi-tasking, flat memory, and 32 bit processing, this was a logical choice. Depending on the future of OS/2 and how well Windows95 turns out to be on the PC, we may have to change operating systems. This task also should be a simple transition.

Q3. WHAT WERE THE SYSTEM DESIGN DECISIONS WE HAD TO MAKE? (REVOLUTION)

Our system is one that has evolved through time. Last year we purchased a dual Pentium PC from the Technology Advancement Group (TAG), a small company that specializes in premium PCs. We chose that company for the quality of their product, their 24 hour support, their proximity for service calls, and their trade-in upgrade policy. We hope to trade in the dual Pentium server for a quad P6 PC during the middle of this year. We have bought system boards and

memory components from them for several years and have had excellent results.

We chose the OS/2 SMP(simultaneous multiple processing) operating system for many reasons. Since we run large memory intensive applications, we needed an operating system that supports large amounts of RAM. We needed multitasking capability since the PC will also be our LAN server and we didn't want one errant application to crash the entire system. We also needed an operating system that supported SMP since we would be running a dual processor. We installed OS/2 LanServer 4.0 since this machine would also be a file server for our individual Pentium PCs running OS/2 WARP and OS/2 SAS software.

Besides designing the system setup in terms of hardware and software, intangible concerns must also be considered, particularly security and redundancy. Using the NIH computer system, a sensible level of security was easily attained through password-accessed logins and password-protected databases. We have relied on the NIH Computer Center to deal with all other aspects of security. We have begun to address our new security needs and will have a security professional check our system for any deficiencies that we may have overlooked.

Our preliminary security steps involve:

1. Placing the server in a locked room.
2. Password protecting and restricting user privileges to the LAN.
3. Backing up data on CD/R twice. (One copy is kept in the office, and the other copy is kept off-site.)
4. Writing applications which encrypt sensitive data on the LAN.

System maintenance is another important consideration. If a component breaks or wears out, we don't have the time to cut procurement requests and to wait 30 days for a new part to be procured and delivered. We have handled this problem in many ways:

1. Redundancy- We have contingency plans for running our system using a backup PC. Due to the nature of our data processing, we must be back on-line within an hour. These plans involve piecing together computer components to create a system capable of producing most data requests within a reasonable time. Although our current backup is only a 486 PC which cannot handle the load as well as a dual Pentium, it is sufficient for a short time. Total redundancy of the system is our ultimate goal.

2. RAID level 5 hard drives- If a hard drive should fail, we can install a new one into the Raid system and recreate the data from the other drives. Our RAID system is made up of five hard drives configured as one. In case of a hard drive becoming defective, we can replace that drive and reconstruct the data from the other four drives. We successfully have tested this data regeneration on at least two occasions. Unfortunately, we also have had an incident where we had to reformat all five drives and recreate the data from CDs. Maintaining current backups is still essential, no matter how sophisticated the technology may become. This mirroring of data is the main reason to use the RAID level of 5. Other levels have their advantages and disadvantages, but level 5 is the one preferred by most people. We wanted a level of mirroring that does not use too much available space or slow down data writing(since a data write would involve all five drives).

3. In-house PC hardware/software expertise- We have over ten years experience in configuring PCs, as well as plenty of tools, and a room full of spare parts. We also have computer support within the Department of Commerce. This computer support is mainly for offices who do not have their own computer professionals, but we can rely on their assistance for matters beyond our own expertise.

Q4. WHAT WERE THE PROBLEMS WE ENCOUNTERED AND WHAT DID WE DO TO RESOLVE THEM?

1. PRINTING LARGE REPORTS

PROBLEM: Currently, we print more than ten boxes of 3,500 sheets each, once a month, four or five boxes once a week, and one box daily. Obviously, PC laser printers are not designed to print this capacity and the maximum PC laser printer rate of 17 pages per minute is too slow. Even making sure that the printer has enough paper(most hold about a ream of paper, 500 sheets) would quickly become a nuisance. A single box of computer output is roughly equivalent to seven reams of paper.

SOLUTION:

A. Replace printing with user-friendly on-screen applications. We have already started to upgrade our users' monitors since they'll be viewing many of our formerly printed reports on screen. We chose 15" NEC multiSync monitors(model XE15) with ATI Graphics Xpression Video cards. 17" and 21" monitors were considered but cost much more than what we wanted to spend for only marginal increases in size. Having a monitor/video card combination with excellent resolution should be the major consideration.

B. Make our publications available on an Internet homepage. We have created a homepage which already contains some of our data and reports. We are gradually adding more data to it. Our homepage address is:

http://www.ita.doc.gov/industry/textiles/

C. Copy print files to tape and print at the NIH Computer Center. We haven't been able to implement this potential solution, because we have been unable to produce output with IBM mainframe style carriage control characters on the PC. We wonder if the SAS Institute has any suggestions in this regard.

D. Maintaining an in-house high capacity printer. Having a mainframe size laser would be too expensive, take up too much room and take up too many resources for upkeep. We have considered purchasing a large printer in league with other offices in Commerce, sharing in all expenses, but have found no interest. Other Commerce offices seem to have found other ways to print their much smaller output.

2. PRINTING SAS SOFTWARE OUTPUT

PROBLEM: This is one our major gripes with SAS software. We have spent much time and paper getting the right page breaks and margins. Also, fonts that look fine on the monitor don't always print that way.

SOLUTION: These problems may become minimal as

we move away from printed output. We're figuring out page breaks and margins that work. After conferring with the SAS Institute, we were referred to an article in the 1992 SAS Proceedings which explained how to use SAS Software to send printer control codes to laser printers. These codes override the fonts provided by the operating system and give us far greater control over our printed output than we would have with them. We can create output on our PC laser printers that very closely resemble our mainframe printed copy.

3. JOB SCHEDULING

PROBLEM: This function is a mainframe attribute not well implemented with PCs. Occasionally it would be useful to be able to execute programs at :

a. certain times of the days
b. different priorities
c. different maximum time limits

SOLUTION: As our expertise with the SAS software macro language becomes more proficient, we find that submitting SAS software jobs to be more flexible. When programs are running, we have screens appear that prompt the user for parameters to specify the type of data request. We also can set these screens to ask for time and job priority specifications.

4. MICROFICHE

PROBLEM: Currently we print microfiches overnight at the NIH Computer Center. We do not have this capability on our PC system.

SOLUTION: As we move more of our public reports to our Internet homepage and to CDs, microfiche output will become obsolete.

5. LAN PROBLEMS

PROBLEM: After installing Novell Netware®(to connect to our Departmental LAN) and IBM LanServer 4.0(to connect to our server) on our workstations we spent several very frustrating months trying to solve LAN connection problems.

SOLUTION: Changing LAN cards seems to have eliminated these problems.

CONCLUSION- WHERE DO WE GO FROM HERE (TRUE VISION)

As we have evolved into downsizing our applications from a large centralized computer system to a small file server system with individual CPUs, we have realized that this natural progression of change is actually a shift in paradigms, a true revolution in information processing. This is a major new trend that quickly should take off in many other data processing environments. However, everyone should take note of what we had to encounter and ask themselves the same questions concerning what is involved:

1. Can the PC handle your applications?
2. What software/hardware is necessary?
3. What system decisions must be made?
4. What are the anticipated problems?

As noted throughout this paper, we took all these questions under consideration and found that a PC based system running the OS/2 version of SAS software works best for our needs.

True vision requires us to anticipate the future. We expect the following developments:

1. Bigger and Better PCs and Components- With a quad-P6 running at 200MHz+, we'll be able to run everything faster. Data requests will be available sooner. With the critical data request deadlines we receive frequently (sometimes several a day), this speed will become the major determinant of the success of this system migration. As hard drives continue drastically to fall in price and to increase in capacity, we will soon have all our data(billions of bytes) on-line, using off-line storage only for data backups.

2. Visual programming software. We have created several flexible and user-friendly applications for viewing our data using Borland Delphi®, Microsoft Visual Basic® and Borland Visual dBase® software. These programming languages allow for the creation of run-time executables that we can include on a CD-ROM containing our data.

We have to admit that our SAS software experience mostly is limited to number-crunching mainframe batch processing. We have not explored any of the Institute's visual programming software capabilities and cannot comment on them. In addition, we cannot create SAS executables that we can provide with our data.

3. More sophisticated Internet homepage applications-We anticipate a major movement toward electronic publications replacing hard copy. Instead of waiting weeks for paper copies to be printed, we'll have close to instantaneous information dissemination. As well, customers will have more than just a report. They also will be able to download our databases through file transfer protocol(ftp), and be able to manipulate these databases for their own ends.

REFERENCES

Riba, S. David, "Enhancing the Printed Appearance of your SAS Reports," Proceedings of the Seventeenth Annual SAS Users Group International Conference, Cary, NC: SAS Institute Inc., 1992. 1602 pp.

ACKNOWLEDGMENTS

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. IBM and OS/2 are registered trademarks or trademarks of International Business Machine Corporation. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

Peter Parker
R. Keith Daly
U.S. Dept. of Commerce
room 3100, Main Commerce Building
14th & Constitution Ave., N.W.
Washington, D.C. 20230
(202)482-1449
Peter_Parker@ita.doc.gov
Keith_Daly@ita.doc.gov

APPENDIX I- THE OTEXA PC SYSTEM SPECIFICATION

1. Dual Pentium running at 66 MHz(planning to upgrade to quad P6 200MHz+, when available)

2. 64 megabytes of RAM(random access memory) (planning to upgrade when applications required

more or significant increases in speed)

3. ten gigabytes of storage on five SCSI drives in external enclosure.

4. ten gigabytes of storage in raid level 5

5. 4x speed seven CD-ROM tower

6. DAT tape backup unit

7. 1.2 and 1.4 MB floppy drives

8. 28.8 kps external fax/modem

9. 15" multi-sync monitor

10. 9 track tape drive

11. Zero-surge surge protector

12. APC uninterrupted power supply (ups)

APPENDIX II- OTEXA Internet Homepage-

http://www.ita.doc.gov/industry/textiles/
for comments on this homepage, send correspondence to the e-mail address-
Peter_Parker@ita.doc.gov

APPENDIX III- Major OTEXA Reports and Publications

SIZE DISTRIBUTION

1. Major Shippers-
monthly report showing textiles and apparel imports by commodity grouping and by country for various time periods for the past three years

Imports in quantity units
SIZE-650 pages COPIES-63
Imports in dollar units
SIZE-650 pages COPIES-14
Imports for special pgms
SIZE-200 pages COPIES-10

2. TQs(Total Quantity)-
monthly report like Major Shippers, but broken down by 10 digit Harmonized codes(subset of commodity grouping)

Imports in quantity and value
SIZE-24 microfiche sheets
(@48 pages per sheet,
eqiv. 1050 pages)
COPIES-50

3. Performance Report-
monthly report showing textiles and apparel agreements and imports charged to quotas.

SIZE=756 pages COPIES=45

4. Summary of Agreements-
bi-annual report showing textiles and apparel agreements by commodity grouping and country.

SIZE=210 pages COPIES=110

5. Exports Report-
quarterly report showing textiles and apparel exports by country

SIZE=216 pages COPIES=100