

Notes from SUGI 24

Jack Hamilton, **First Health**, West Sacramento, California

What's New

It was difficult to write this paper. Last year's SUGI had a unifying theme, the great new things in SAS Version 7. SUGI 24, on the other hand, had no such theme.

The main news is that there's not really much new. Version 7 is shipping, but its installation is discouraged by SAS Institute. Version 8 is in widespread internal use at SAS Institute, and should be shipping in beta during May to sites which requested version 7. Many of the presentations at SUGI which were labeled "Version 7" actually referred to version 8, or both version 7 and version 8.

Version 8 is mostly bug fixes and incremental improvements. There are some new products, but they are not of immediate use for most users of SAS Software. Most of the action was on the GUI side; those of us who use SAS for plain old data management and reporting felt somewhat neglected.

SAS Institute appears to be repositioning itself as a "knowledge company" (a.k.a. consulting firm) rather than a software vendor. One speaker at the Futures Forum made it explicit: the old way was for customers to provide the data and the subject matter expertise, and for SAS Institute to provide the software; the new way is for SAS Institute to provide the expertise, methods, and software. Customers provide the data and the money. There was much more focus this year on commercial, as opposed to research, uses of SAS Software.

The New Buzzwords

Some of the new buzzwords, which we can expect to see in SAS Communications and presentations from SAS Institute, are "Balanced Scorecard", "Customer Relationship Management", and "Collaborative Business Intelligence".

The Balanced Scorecard

The concept comes from the book *The Balanced Scorecard : Translating Strategy into Action* by Robert S. Kaplan and David P. Norton. The Balanced Scorecard provides a new method for an organization to chart its progress. The time lag needed to see the effect of a change (2-3 years at SAS Institute, for example) causes management to spend too much time examining costs, and not enough time looking for opportunities. The Balanced Scorecard looks at four things: Internal business processes, Financials, Learning and growth (emphasizing employee development and "buy-in"), and Customer relations. "Just doing things right isn't enough – you have to do the right things."

Several SAS Software products (SAS/Intrnet, SAS/EIS, Enterprise Miner) appear to be good tools for implementing the Balanced Scorecard (which no one called by its natural acronym). This ties in well with SAS Institute's desire to become a knowledge company – these products have a steep learning curve, and implementation requires specialized knowledge not only of the business but of new business techniques. This is an area in which SAS Institute will compete with big consulting firms.

Customer Relationship Management

CRM is a marketing concept. To quote again from a SAS web page:

In customer relationship management (CRM), the focus is on retaining customers and selling them new (cross-selling) and more profitable (up-selling) products. It is generally many times more cost-effective to retain existing customers than it is to acquire new ones. The way to retain them is to delight them time and time again, by offering tailored products which exactly meet their changing wants. And to do that, you need a lot of information.¹

SAS Software is good at processing lots of information, especially with the Enterprise Miner. SAS Institute's success in this arena will depend on how well they market themselves.

Collaborative Business Intelligence

I missed the most of the presentations on this, so I'll quote from the SAS Institute web page again:

¹ <<http://www.sas.com/software/sas.com/interactive/spring98/execview/>>

Collaborative Business Intelligence combines the number-based information generated by the business intelligence process with the text-based materials that describe that information, resulting in the capture and re-use of corporate knowledge.

Collaborative Business Intelligence allows companies to unlock the power of their data stores by creating collaborative, contextual information environments. Analysts who traditionally use business-intelligence tools to look for important business factors can now save their findings into a knowledge repository, making it easily accessible to the rest of the organization over the Web. Decision makers become more engaged in the knowledge-building process as they add to the thread, and everyone can search for similar projects and re-use ideas that have worked in the past.²

This product is not aimed exclusively at commercial enterprises. It takes advantage of a number of existing SAS Software products, and there is a new publish-and-subscribe capability for automatic distribution of information. As with any product, it's too early to tell whether CBI will be a success, but it certainly has potential. The problem it addresses – how to capture and use expert knowledge – is a serious one. Unfortunately, the presentations on this topic were severely buzzworded ("contextual information environments"?), so it was hard to tell how much meat there is in the product.

Enterprise Miner

Enterprise Miner is SAS Institute's data mining tool. It was introduced 2 SUGI's ago, and has gotten major improvements in power and usability since then.

At the Opening Session, Dr. Goodnight mentioned a problem to which Enterprise Miner is currently being applied. Further detail was given during a session at (gasp) 8 am on Monday morning. Some researchers at UNC (including one of Dr. Goodnight's fellow grad students way back when) are studying the correlation between various genes and Alzheimer's Disease. They are using Enterprise Miner to help find correlations between certain base pairs and the disease. There's not enough data (yet) to draw firm conclusions, but progress is being made. Alzheimer's affects a lot of people, and this study has potential for doing a lot of good. It is also a very good showcase for the power of Enterprise Miner, as there are billions of potential interactions to test. A number of new algorithms (including at least one by Dr. Goodnight) have been included in the product, and the user interface seems much improved from previous versions.

Version 8 for Windows

The most visible addition for version 8 is the Enhanced Editor, which can replace the current program editor (you can use both if you want). It's similar to the version 7 Enterprise Guide editor, but better. Some of the features:

- The editor is context sensitive, and includes color coding for SAS, SCL, and HTML. You can also edit plain text. You can devise and save your own color schemes.
- Code is divided into sections (basically, data steps or procs) which are collapsible, and you can show section boundaries.
- Simple text autocompletion is supported as part of the unfortunately named macro facility.
- You can have several editor windows open at once.
- Multiple levels of UNDO are supported.

All of these features are also supported in the editor in Enterprise Guide (a Windows-only product). The enhanced editor is not supported for other platforms in version 8.

Some drawbacks:

- The term "macro" has been reused in the Enhanced Editor to mean something entirely different from what it means in the rest of SAS Software.
- The macro editor language is entirely new, rather than using the syntax of one of the other 9 programming languages already in use. It's also not a particularly powerful language – a macro can't make decisions based on what it's acting upon.
- Prefix commands aren't supported.

² <<http://www.sas.com/new/preleases/020599/news2.html>>

- Macro information is stored in the registry rather than a SAS catalog, making it more difficult to maintain a central library.
- In some ways it's obvious that the Enhanced Editor wasn't written by users of SAS Software.

Some other changes in SAS for Windows:

- HTML supports styles.
- Internet Explorer is no longer required unless you want to use the embedded HTML viewer (Netscape programs do not support the interface needed by SAS for Windows). You can use Netscape in a separate window.
- The Command Bar can be made longer.
- The SAS Explorer window has some new features.
- You can make FSVIEW the default viewer, replacing the V6 Viewtable.
- Help uses HTMLHelp.

I didn't spend a lot of time looking at the new version of Enterprise Guide (version 3). I described the previous version in last year's SUGI notes. Basically, it's a Windows program which lets you edit and submit SAS programs to a remote host, view and print output, and transfer data between various host and Windows formats. For some people, it will provide a replacement for SAS for Windows. It requires a version 8 server on the host.

Y2K

There was some mention of Y2K problems and solutions. SAS version 7 is pretty much Y2K compliant, but (and this was emphasized on many occasions) you must make sure that the system option YEARCUTOFF= is set correctly if you expect to encounter 2-digit years in your data.

Dave Brummel, head of Tech Support, said that he plans to be on the phone with the SAS Institute office in New Zealand on the evening of December 31. "If the phone system goes dead at midnight, we know we're in trouble." He also said that Tech Support will be open around the clock for a few days around the millennium. I wasn't sure whether he was serious or not.

In general, SAS users don't have much to worry about as long as they have stored dates as SAS date values. The most obvious possible troublemaker that's part of SAS Software, the SYSDATE system macro variable, is supplemented in V7 by SYSDATE9, which provides a Y2K-compliant data string.

OpenVMS Issues

One high-ranked person at SAS Institute told me, off the record, that OpenVMS is a dying platform, and that it will not be the object of much development work. This was already obvious from the lack of character mode support in version 7, and from the decision not to support OpenVMS as a server for Enterprise Guide and Enterprise Reporter.

I talked to one of the developers about these two issues. He said that character mode support was a nightmare, and he didn't think it would be implemented for version 8.

The lack of Enterprise Guide support is the result of a marketing decision, and it might be implemented if there is sufficient demand. It's technically feasible.

It's not clear whether SAS/ACCESS to INGRES is or will be implemented or not; I heard conflicting stories. I guess we'll have to get the product and see if it's there.

I had hoped that SAS Institute would implement a lightweight X instead of the ponderous Motif (one of the Usenet FAQ's on X says "Don't use Motif. It's a pig."), but no luck there either. SAS Software will run under OpenVMS with an X server on Windows, but it's no speed demon, and it's hard to set up correctly (it's not well documented, for one thing).

On the positive side, version 8 will have much better disk caching, eliminating the need for the RAH= and WBH= dataset options.

Distributed Processing

There are some interesting new additions in Version 7/8 for distributed processing. In these examples, I'll assume that the client is SAS for Windows, but many of these capabilities are supported for other platforms.

- Asynchronous RSUBMIT

This lets you submit SAS code to a remote host and continue working in your SAS session on the PC. You can check on the status of the remote session, and view the remote log or output locally. You can have several asynchronous sessions running at the same time. It uses SAS/CONNECT, so all server sessions do not have to be running version 8.

- CEDA (Cross Environment Data Access)

This is part of the base product, and lets you read and write foreign host datasets. It's used with FTP or NFS. Currently, updates aren't supported, nor are indexes, but it will be available for MVS with some restrictions (a change from last year).

- MultiProcessing Connect

This lets you run asynchronous or parallel processes on the same host. It uses some of the capabilities of asynchronous RSUBMIT, but on the same machine. I think it doesn't require a SAS/CONNECT license.

- Messaging Services

This is new. SAS now has mechanisms for both direct messaging and indirect messaging (store and forward). These services will be available in SCL, datasteps, and the macro language.

- Remote Objecting Services

This is an extension of Messaging Services that enables you to invoke remote methods.

- Agent Services

This provides a way to schedule SAS programs.

- Network Data Encryption

This is used by CONNECT and SHARE. You can choose between various algorithms. Anything other than the proprietary SAS algorithm requires a SAS/SECURE license.

Unfortunately, these are not open interfaces. It will not be possible to write a Visual Basic program, for example, to read and write the Messaging Services queues, or to invoke remote objects.

New Features in SUMMARY and TABULATE

There's some neat new stuff in PROC SUMMARY and PROC TABULATE, which now use a common engine. The same capabilities will be added ("someday") to PROC REPORT.

The new engine had a number of design goals:

- Provide more control over CLASSES.
- Make better use of user-written formats.
- Make it easier to print 0's in reports (automatically create rows and columns for specified values)/
- Create output datasets when possible.
- Generate quantile statistics such as MEDIAN in base procs, even for large datasets.
- Generate the same numbers as PROC UNIVARIATE and PROC GLM for weighted statistics.
- Provide independent variable weighting.
- Outperform version 6 code in most cases.

I like these changes, because they're *in the base product* and I'm likely to actually use them.

And some details:

- You can now specify multiple CLASS statements, allowing options to be specified for each level.
- **CLASSDATA=datasetname** lets you specify a dataset defining the class value combinations which must appear in the output, even if the frequency is 0. The EXCLUSIVE option lets you specify that only those combinations in CLASSDATA will be output.
- **CLASS x / PRELOADFMT** allows you to use the values in a format as a data source.
- **CLASS x / MISSING** lets you specify the MISSING option at the class level, rather than for all classes.

- **CLASS x / ORDER=** allows you to order different class variables in different ways.
- **CLASS x /MLF** turns on multilabel format processing, which allows one input variable to contribute to more than one output level.
- You can group by unformatted values as well as formatted values.
- Variables in output datasets generally inherit characteristics from variables in the input dataset. This can be controlled in various ways.

PROC SUMMARY specifics

- The **TYPES** statement allows you to specify exactly which types you want output – you're not restricted to everything or **NWAY**. For example

```
class a b c d
types a*b (c d) ( );
```

- The **WAYS** statement allows you to specify that only 2 way, 3 way, whatever-way combinations will be output.
- There are new procedure options for **PROC SUMMARY** for the new quantile measures: **MEDIAN**, **Q1**, **P10**. etc.
- **VAR y / WEIGHT=x** lets you weight variables independently.
- The **IDGROUP** statement combines and extends **ID**, **MINID**, and **MAXID** in various ways that I don't understand yet. It sounded useful – the **ID** processing in version 6 has an element of arbitrariness that I don't like.

Linux

At the Futures Forum, it was announced that they have SAS running under Linux, but have encountered problems. They would need more support from the Linux community before SAS could support SAS under Linux. Details (and potential pricing) weren't given.

ODS

The Output Delivery System is a major enhancement to SAS. It is in version 7, but there are some enhancements in version 8. Paul Kent gave a demo; he started with a basic **PROC PRINT**, added an **ODS** statement to redirect it to **HTML**, and finished by adding a **PROC GCHART** with automatic drilldowns. Admittedly, it was a demo, but the system seemed both powerful and easy to use.

RTF and **XML** engines are experimental, but they're there.

Direct writing of **PDF** files is not yet supported, and won't be for at least a year. The current method of creating **PDF** files is to write a **PostScript** file and run it through **Adobe Distiller** or the **PDF** writer. The next level of support will be for **PDFMarks** (which I think will be the most difficult part).

There is no current support for a line-printer output, which is unfortunate. One of the promises of **ODS** was that it would allow much greater control of the formatting of reports, but apparently line printers aren't important enough to justify an **ODS** engine. Your ability to write reports with **DATA _NULL_** isn't useless yet.

There are plans for a **GUI** based template editor, a style editor, and extended **DATA _NULL_** capabilities.

SQL

A number of performance enhancements are in version 7 or version 8, mostly dealing with better use of implicit passthrough queries. As much information as possible will be sent to the underlying database, which will frequently result in faster operation. More use of indexes for optimization will be done. In the longer term (version 9), there will be support for more types of **JOINS** and for more **SQL2** functions. There will be additional optimization, especially when integrity constraints or indexes are involved.

Documentation

To some extent, SAS Institute is getting out of the publishing business. Most Institute-written documentation will be delivered on the installation **CD-ROM** in **HTML** and **PDF** formats. You are free to install the documentation on your **LAN** for everyone to look at, and to print copies of the documentation. SAS Institute isn't even planning to print the **Language Reference** and **Procedures Guides** – you can print them yourself (over 2,000 pages), or have SAS Institute print them for you, but it will be a special order with on-demand printing.

The **Books By Users** series will continue.

I picked up a few manuals at SUGI that you might want to consider buying:

- *The Little SAS Book: a primer*, by Susan Slaughter, the second edition of a book which has proven popular with new users.
- *PROC TABULATE by Example*, by Lauren Haworth, a set of examples of what you can do with PROC TABULATE, including an interesting section on how to trick it into producing output it really doesn't want to produce.

I also ordered a copy of the documentation CD-ROM, but it hasn't arrived yet. It covers only version 7; the complete version 8 documentation won't be available until much nearer the production ship date.

appDev Studio

One of the more interesting niche products is appDev Studio. As I understand it, it's a development environment for Java apps which connect to a SAS Software server. In a typical project, there might be two developers: one on the server side who develops an AF application with classes and methods, and another on the client who develops a Java app which calls those methods from a client. The AF developer doesn't have to know anything about Java, and the Java developer doesn't have to know anything about SAS Software. The client calls the server over the network, and can do anything a regular AF application can do. It doesn't require a SAS license on the client, just on the server. Quite an interesting idea; it allows existing AF apps to be ported to a client server architecture, it provides more flexibility than Enterprise Reporter, and it gives more control over what the client can do than Enterprise Guide. The license fee for the development environment is apparently quite expensive.

Miscellaneous Comments

- SAS is attempting to reduce its cycle time for new products. One way they intend to accomplish that is by introducing new technologies in limited subsets. On specific example of something we can look forward to is additional data types.
- We won't get long format names in version 8, and probably not in version 9, because it requires changes to the SAS file format.
- They hope to start issuing complete install CD's, rather than base CD's plus update CD's.
- Claire Cates is looking for test data to use for performance testing. This was mentioned in more than one session, so they must really want it.
- Dr. Goodnight appeared at the mixer and in the Demo Area. I don't think I'd ever seen him out in public before.
- In the battle between "data is" and "data are", "data is" has prevailed at SAS Institute.