Paper 116

Just Enough Database for Manufacturing Yield Analysis

Robert A. Rutledge, Linda H. Tai, Garth W. Helf
IBM Storage Systems Division, San Jose, CA

ABSTRACT

The IBM Storage Systems Division operates advanced manufacturing facilities for fabricating, assembling and testing all of the major components of IBM data storage products, from disks and magneto-resistive heads through finished disk drives and storage subsystems. The operational data from our manufacturing sites are also used for engineering analysis, yield management and problem diagnosis, within each factory and across factories throughout the division. This cross-functional data allows us to exploit the benefit of vertical integration by analyzing the relationships between hard drive performance and component measurements, and using the resulting insight to eliminate yield detractors and optimize the performance and quality of the finished product. However, the data association task is very time-consuming because the operational data are located in several different databases, and the logical relationships among databases is quite complex.

The JED (Just Enough Database) project was undertaken to optimize our use of operational data for yield analysis. The JED system consists of a Data Warehouse which combines and correlates data from several operational databases, and a graphical user interface, which automates the yield analysis process by providing engineers with preprogrammed failure mapping and yield sensitivity curves and drill-down access to source data for in-depth interactive analysis. The user interface was written in SAS/AF®, to run on the SAS system for PCs, using SAS/CONNECT® to extract data and code from the RS/6000® server. It has since been ported to run on a web browser using SAS/INTRNET®

INTRODUCTION

The disk drive industry is characterized by the rapid introduction of advanced technology that delivers ever higher capacity and performance for the end users. The industry is also highly competitive, which leads to continuous erosion of prices. In this environment there is a significant advantage to being able to ramp new products to high volume quickly and take advantage of the higher profit margins on leading-edge products. Fast ramp to volume requires fast yield learning to detect, diagnose and correct the inevitable problems that arise in new technology. The IBM Storage Products Division manufactures the major drive components and subassemblies as well as the finished disk drive. This vertical integration leads to close cooperation among the component and drive development and manufacturing teams. In particular, we can access the various operational databases to quantify the relationship of drive performance to component measurements, and use the resulting information to discover the root cause of drive problems, and optimize the specifications on in-line component measurements.

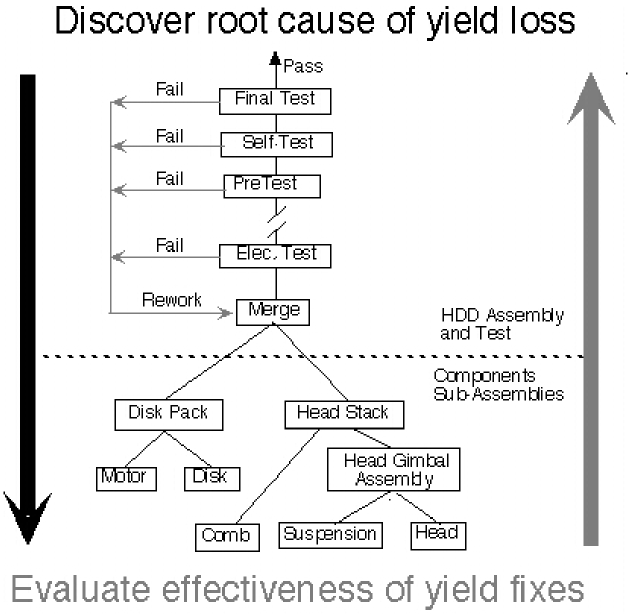


Figure 1: JED Analysis Objectives

However, the data association task is very time consuming because the operational data are located in several different databases, and the logical relationships among databases is quite complex. Drives which fail early in the assembly process due to a bad component will be reworked and then reassembled with new components. Therefore there is not a one-to-one relationship between a drive and its components, and every query must use not only serial numbers, but also the time stamps of various process steps, to insure that a particular component was in fact in the drive at the time of a specific test.

The JED (Just Enough Database) project was undertaken to optimize our use of operational data for yield control. The first step was the creation of a Data Warehouse, which brings together the operational data from our Wafer, Disk, Head Gimbal Assembly (HGA), Head Stack Assembly (HSA) and Hard Disk Drive (HDD) factories. The next step was the creation of a user interface to make the data more easily available and to automate some of the commonly used yield analysis tools.

JED DATA WAREHOUSE

The JED Data Warehouse is a DB2 database of approximately 50 Gigabytes residing on an IBM RS/6000 server. It is updated weekly by extracting the drive test results and parametric measurements for the previous week's production from the operational databases for our drive assembly plants in San Jose and Singapore. After the drive serial numbers and process time stamps are stored, we search our subassembly and component databases to find the key in-line measurements, which are then stored in JED for easy access. Before the data are stored, the drive and component data are pre-correlated so that one simple query will retrieve drive test results along with measurements taken in the component factory on the same parts. Thus the user is spared the difficulty of tracing component usage through the subassembly and drive assembly and rework processes.

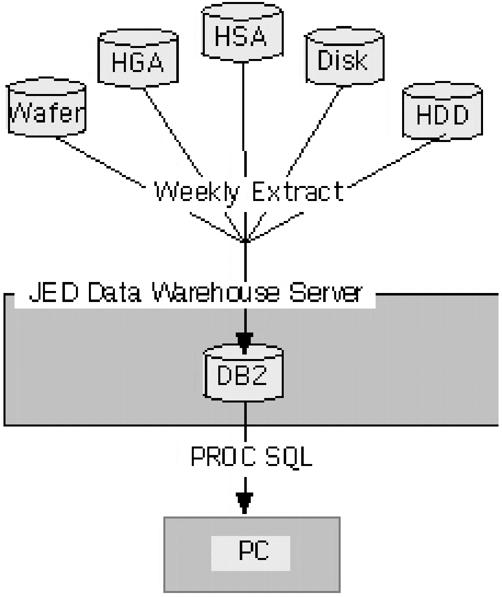


Figure 2: JED Data Warehouse

As the term "Just Enough" suggests, we do not aggregate all of the available data, but only just enough to satisfy our objective of fast and effective yield analysis. We select the drive test results that are directly related to our yield calculations, and the in-line measurements that are likely to be informative for yield and failure analysis. We further reduce the data by capturing data on only a few thousand passing and failing drives of each model each week. For a new product, this is a 100% sample, but as the product volumes increase, the sampling rate is gradually reduced to perhaps 5% of passing drives and 50% of failing drives at each test station. The reduced samples are more than enough for valid data analysis, but much more economical to update, store and analyze.

Users access the JED warehouse using the SAS system for PCs running under OS/2 or Windows 95 or NT. A typical user will run a PROC SQL with an ODBC connection to extract data from JED, analyze it, and then use PROC GPLOT or GCHART to produce charts for presentation. The JED Data Warehouse has been in operation for more than a year now, and has proven very effective in analyzing yield detractors in current and new products, and evaluating the effectiveness of proposed fixes.

However this system still left some bottlenecks in the information flow. There are many engineers who lack the SAS programming skills to get at the data, and who must therefore rely on a few experts to help out with each request. Some common analysis routines would get programmed many times as different engineers made similar requests to different SAS programmers. And a collection of static charts brought to a meeting for presentation usually led to questions like “What if we don't include the data from level X components?", which could not be answered without rerunning the analysis. We addressed all of these issues by constructing a user interface for JED using SAS/AF®.

SAS/AF USER INTERFACE FOR JED

The JED interface runs locally on the user's PC, using SAS/AF code and SAS data sets which reside on the RS/6000 JED Data Mining server. The user starts a small SAS/AF program on the PC, which establishes a SAS/CONNECT session with the server, and downloads the main SAS/AF code. This insures that the user is always accessing the latest version of the SAS/AF code. The data used by the application are stored in two tiers of SAS data sets. The first tier is produced by running weekly SAS SQL programs against the main JED data warehouse. These relatively large data sets, perhaps 500 MB per week, are then reduced to a second tier of summary data sets, which contain enough information to drive the SAS/AF interface. The reduced data sets contain only about 20 MB of information altogether for several weeks of production, which can be downloaded to the user's PC, and allow the interface to run with very fast refresh times.

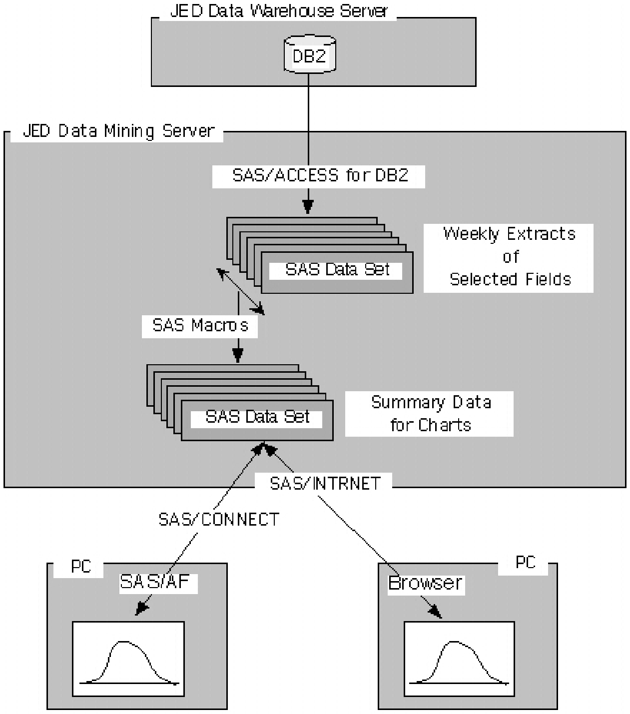


Figure 3: JED User Interfaces

The first JED frame offers the user three choices: 'Remote', 'Download' or 'Local'. If 'Remote' is chosen, the application will download data sets as required for each frame the user visits. The 'Download' option will download all data sets at once, so that thereafter all processing is done locally. The 'Local' option will run cause the application to run stand-alone on the PC, using whatever versions of code and data were most decently downloaded. This allows one to continue with analysis when the server is unavailable or inconvenient to use, e.g. while using a laptop to present data at a meeting.

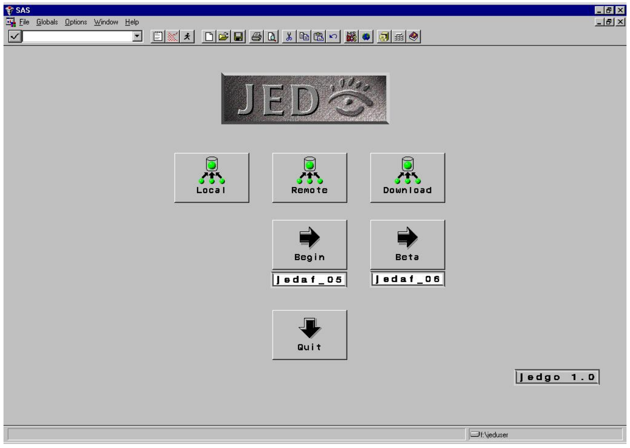
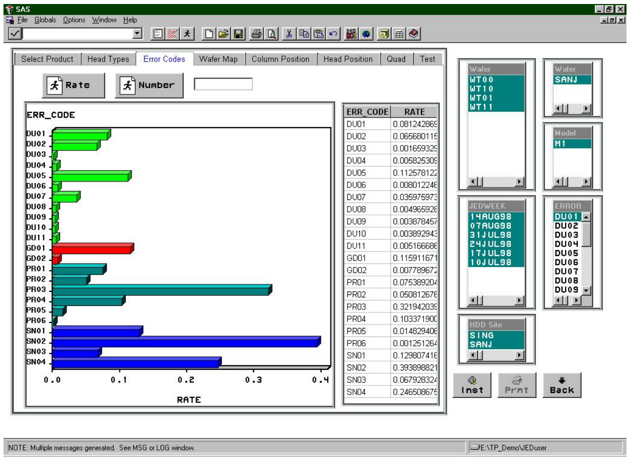


Figure 4: First Frame of SAS/AF JED Interface

The JED user interface is designed to provide an OLAP-like system for yield analysis. It provides fast answers to the most commonly asked yield analysis questions with point-and-click simplicity. Some of the more important analyses available are failure maps and yield sensitivity curves. The failure mapping functions allow the user to detect failure trends associated with the type or source of components used, the head position in the drive, or the location of heads on the parent wafer. The yield sensitivity curves show the drive failure rate as a function of in-line measurements.

Figure 5 shows a frame that is used to explore the failure rates for various error codes. The list boxes at right, which are automatically populated by SCL code, allow the user to select the component types and manufacturing sites and weeks of interest. These selections are used to reduce the summary data sets to the data of interest, and then the corresponding error rates are computed and plotted.

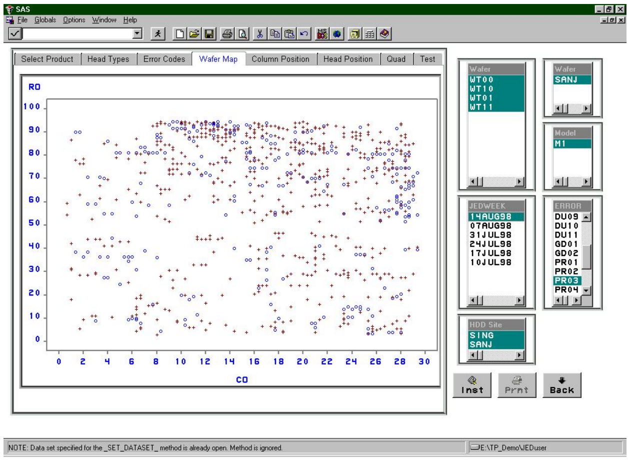


ERR_CODE	RATE
DU01	0.08124286
DU02	0.06568011
DU03	0.00165932
DU04	0.00582530
DU05	0.11257812
DU06	0.00801224
DU07	0.03597597
DU08	0.00496592
DU09	0.00387845
DU10	0.00389294
DU11	0.00516668
GD01	0.115911671
GD02	0.00778967
PR01	0.07538920
PR02	0.05081267
PR03	0.32194203
PR04	0.10337190
PR05	0.01482940
PR06	0.00125126
SN01	0.12980741
SN02	0.393898821
SN03	0.06792832
SN04	0.246508675

NOTE: Multiple messages generated. See MSG or LOG window.

Figure 5: Failure Rates by Error Code

After the main yield detractors are identified by the frame shown in Figure 5, the user can select diagnostic charts that will point toward the root cause of the problem. One important type of plot is a wafer map showing which part of the wafer the failing heads came from. Figure 6 shows an example of such a wafer map which indicates that heads from the upper right hand area of the wafer are more susceptible to failing for the particular error code plotted.



NOTE: Data set specified for the _SET_DATASET_ method is already open. Method is ignored.

Figure 6: Wafer Map for Error Mode 1

Figures 7 and 8 show the distribution of wafer location and head stack position for another specific error code. The uniformity of the wafer plot, and the bias in the head stack distribution, strongly suggest that the root cause for this problem will be found in the HDD assembly process rather than the wafer process.

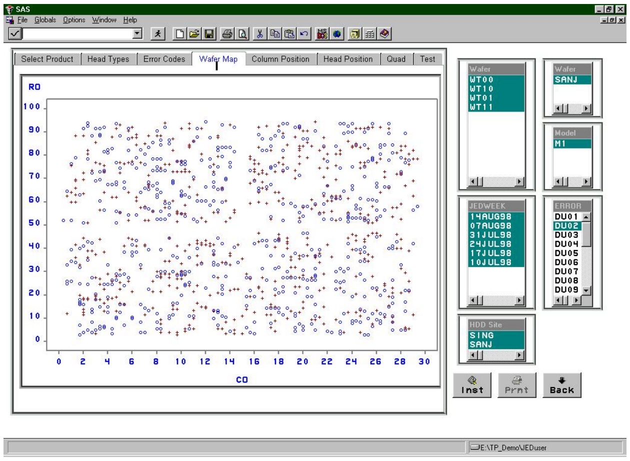


Figure 7: Wafer Map for Error Mode 2

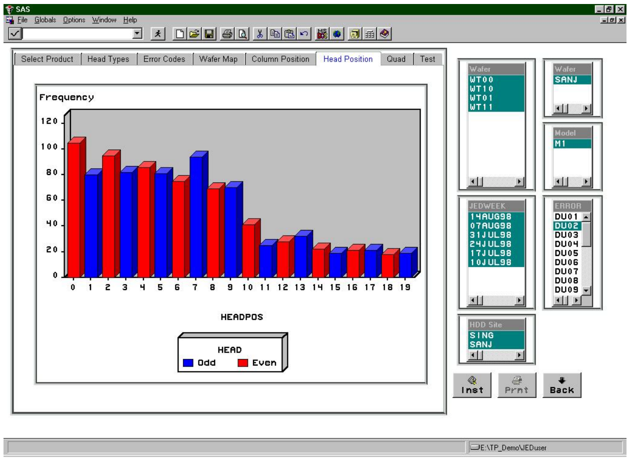


Figure 8: Head Positions for Error Mode 2

These error mapping charts provide valuable indicators of the root cause of each HDD yield detractor, and what actions we might take to eliminate them.

This interface gives the user instant access to many more charts than we could have possibly produced as static charts. A single SAS/AF frame allows the user to quickly and dynamically produce error maps depending on his selection of component type, drive model, error code, and week of production, from billions of possible combinations. Each frame that contains a chart also has a button that can be pressed to invoke SAS/INSIGHT®. This not only brings up the raw data, but also allows the user to exploit the full interactive graphic capability of SAS/INSIGHT to explore the data.

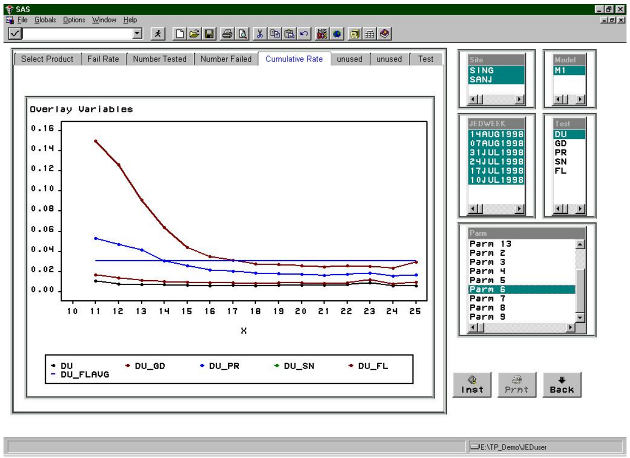


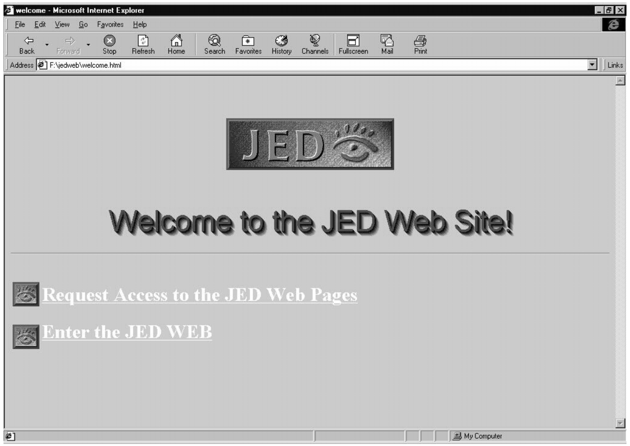
Figure 9: Yield Sensitivity Plot

Figure 9 shows an example of another powerful JED diagnostic chart. This shows the cumulative failure rate over five HDD assembly and test processes as a function of a parametric measurement on one component in the drive. The chart is assembled from a large set of tables produced by PROC MEANS, with each table containing a sensitivity curve based on a limited slice of the data, e.g. one week, one test, one site, etc. After the user selects the data region of interest from the list boxes at right, SAS macros are used to assemble the final cumulative charts from the individual tables.

From such charts we can easily see which component specifications should be adjusted to optimize the HDD yield, and how much yield improvement we can expect to see from the new specifications.

SAS/INTRNET USER INTERFACE FOR JED

The SAS/AF interface described above requires installation of the SAS system for PCs as well as the initial start-up SAS/AF application. Most of the engineers on site who are doing data analysis are already using SAS, so this is not a problem. However, we also want to make the JED analyses available to a wider community, including managers who are not SAS users, and engineers in our plants in Singapore, Japan, Ireland, Germany and Mexico. The ideal way to accomplish this was by porting the JED interface to SAS/INTRNET, which makes it easily available to anyone in IBM who has a PC, a browser, and access to the IBM Intranet.



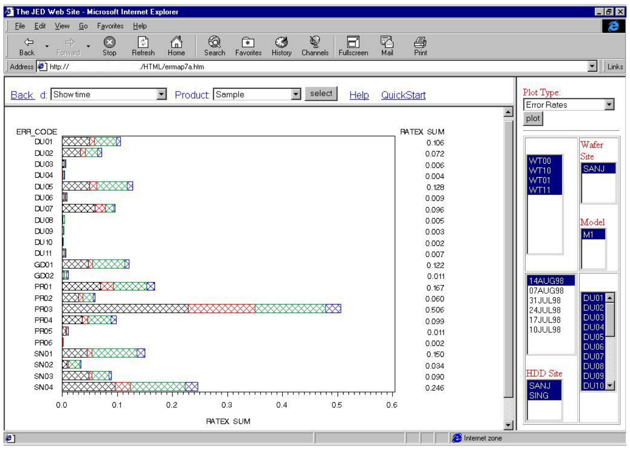
Welcome to the JED Web Site!

Request Access to the JED Web Pages

Enter the JED WEB

Figure 10: The JED Web Site

The SAS/INTRNET interface for JED was programmed to run off the same data sets created for the SAS/AF interface. (See Figure 3.) Some re-programming was required to convert SAS/AF frames to HTML pages, but the SAS macros which were written to do the data manipulation and analysis work for the SAS/AF interface could be reused for SAS/INTRNET with little or no changes. Figure 11 shows how the Error Rate plot of Figure 5 looks when translated into the SAS/INTRNET interface for JED.



d: Show time
Product: Sample
Plot Type: Error Rates

WT00
WT10
WT01
WT11

Wafer Site: SANJ

Model: M1

14AUG98
07AUG98
31JUL98
24JUL98
17JUL98
10JUL98

HDD Site: SANJ, SING

DU01
DU02
DU03
DU04
DU05
DU06
DU07
DU08
DU09
DU10

ERR_CODE	RATEX SUM
DU01	0.106
DU02	0.072
DU03	0.006
DU04	0.004
DU05	0.128
DU06	0.009
DU07	0.096
DU08	0.005
DU09	0.003
DU10	0.002
DU11	0.007
GD01	0.122
GD02	0.011
PR01	0.167
PR02	0.060
PR03	0.506
PR04	0.099
PR05	0.011
PR06	0.002
SN01	0.150
SN02	0.034
SN03	0.090
SN04	0.246

Figure 11: Error Rates in the SAS/INTRNET interface

THE NEXT STEPS

We will continue to improve the JED interface by adding new charting and analysis capabilities as we discover new ways to diagnose yield problems and evaluate yield enhancements. This effort will include the use of advanced data mining tools to explore the multi-dimensional relationships among parametric and logistic data and HDD yield. We have installed SAS Enterprise Miner® and IBM Intelligent Miner® on the Jed Data Mining server, and the preliminary results using the decision tree algorithms are very promising.

We also plan to take advantage of SAS Version 7 and SAS/WEBAF® when they are available to improve the look, feel and functionality of the JED interface. And we will investigate the possible use of SAS/MDDB®, which may be a more efficient way to accomplish much of what we are currently doing with PROC MEANS and macros.

CONCLUSION

A key ingredient in time-to-volume manufacturing is the rapid diagnosis and elimination of yield detractors. The JED data warehouse, which brings together the operational data from our component fabrication and hard drive assembly factories, is proving very effective for fast yield analysis of IBM's server class hard drive products. The JED SAS/AF and SAS/INTRNET user interfaces make the system even more effective by making the data and analysis algorithms quickly and easily available to engineers throughout the division. The use of these systems over the past year has enabled us to quickly diagnose several yield problems, propose solutions to increase HDD yield, and evaluate in advance the effect of the proposed fixes.

ACKNOWLEDGMENTS

The authors would like to express their appreciation to the many people who contributed to the success of the JED project:

The JED Steering Committee, including the authors plus: Roger Brown, Randy Edelmaier, Francisco Infante, Benny Lopez, Faye Ly, Ray Matsuoka, Pat Motheral, Gan Qiu, Rainer Richter, Rick Spanel and Ren Umeda.

The IBM Global Services Team who built and continue to maintain the JED Data Warehouse: Roger Buland, Kerry Hayes, Joe Huchel, Mike Henningsen, Bob Larocca, Barbara Okimoto, Rob Strong.

Heidi Leung, also of IBM Global Services, who built and maintains the JED Data Mining Server and the JED Web Site.

Qin Pan, with Ciber Information Services, for help with programming the JED Web interface.

John Balas of the IBM Microelectronics Division, whose DataView Application was the inspiration for the JED SAS/AF interface.

Chuck Bininger of Trilogy Consulting Corp. who did an excellent job of designing SAS/AF frames for constructing and submitting DB2 queries to the JED Data Warehouse.

Andy Norton and Jack Fuller, also of Trilogy, who spent many hours helping us to understand the confusing maze of options for implementing a user interface using SAS/AF, SAS/CONNECT, SAS/INTRNET and Java.

Art Carpenter of California Occidental Consultants for helpful discussions on SAS graphics and macros, and for raising the level of SAS competence at our site through his classes and books.

The many SAS employees who were extremely helpful in getting us up and running with SAS products in less time than we would have thought possible, including:

- Kim Fields and Greg Lloyd (SAS San Jose Office)
- Jim Georges (SAS Enterprise Miner)
- Rena Soffir and Aaron Hill (SAS/INTRNET)
- Sam Atassi and Margo Scharer (SAS/INTRNET)
- Regina Louie and Jim Luther (SAS/CONNECT)

Irene Rutledge, for design of the JED-eye logo.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Robert A. Rutledge
IBM Storage Systems Division
5600 Cottle Road
San Jose, CA 95193
Work Phone: (408) 256-0804
Fax: (408) 256-2138
Email: brutledg@us.ibm.com

Linda H. Tai
IBM Storage Systems Division
5600 Cottle Road
San Jose, CA 95193
Work Phone: (408) 256-6945
Fax: (408) 256-8987
Email: ltai@us.ibm.com

Garth W. Helf
IBM Storage Systems Division
5600 Cottle Road
San Jose, CA 95193
Work Phone: (408) 256-7514
Fax: (408) 256-2410
Email: helf@us.ibm.com